Electronic Colloquium on Computational Complexity, Report No. 33 (2016)

Tight bounds for communication assisted agreement distillation

Venkatesan Guruswami[*] Jaikumar Radhakrishnan[†]

Abstract

Suppose Alice holds a uniformly random string $X \in \{0,1\}^N$ and Bob holds a noisy version Y of X where each bit of X is flipped independently with probability $\varepsilon \in [0, \frac{1}{2}]$. Alice and Bob would like to extract a common random string of min-entropy at least k. In this work, we establish the communication versus success probability trade-off for this problem by giving a protocol and a matching lower bound (under the restriction that the string to be agreed upon is determined by Alice's input X). Specifically, we prove that in order for Alice and Bob to agree on a common string with probability $2^{-\gamma k}$ ($\gamma k \geqslant 1$), the optimal communication (up to $o(k)$ terms, and achievable for large N) is precisely $(C(1-\gamma) - 2\sqrt{C(1-C)\gamma})k$, where $C := 4\varepsilon(1-\varepsilon)$. In particular, the optimal communication to achieve $\Omega(1)$ agreement probability approaches $4\varepsilon(1-\varepsilon)k$.

We also consider the case when Y is the output of the binary erasure channel on X, where each bit of Y equals the corresponding bit of X with probability $1-\varepsilon$ and is otherwise erased (that is, replaced by a '?'). In this case, the communication required becomes $(\varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma})k$. In particular, the optimal communication to achieve $\Omega(1)$ agreement probability approaches εk, and with no communication the optimal agreement probability approaches $2^{-\frac{1-\sqrt{1-\varepsilon}}{1+\sqrt{1-\varepsilon}}k}$.

Our protocols are based on covering codes and extend the approach of (Bogdanov and Mossel, 2011) for the zero-communication case. Our lower bounds rely on hypercontractive inequalities. For the model of bit-flips, our argument extends the approach of (Bogdanov and Mossel, 2011) by allowing communication; for the erasure model, to the best of our knowledge the needed hypercontractivity statement was not studied before, and it was established (given our application) by (Nair and Wang 2015). We also obtain information complexity lower bounds for these tasks, and together with our protocol, they shed light on the recently popular "most informative Boolean function" conjecture of Courtade and Kumar.

1 Introduction

Suppose Alice holds a string $X = (x_1, x_2, \dots)$ of uniformly random bits and Bob holds a correlated random string $Y = (y_1, y_2, \dots)$ where the bit y_j is the bit x_j flipped (independently for each j) with probability $\varepsilon \in (0,1)$. Their goal is to communicate as little as possible and agree on a uniformly random string in $\{0,1\}^k$ (or under a relaxed requirement, sample a common string from a distribution of min-entropy at least k).

Besides being a natural task, this scenario also relates to the problem of extracting a unique ID from process variations; see [BM11], which studied the communication free version of this question,

[*] Computer Science Department, Carnegie Mellon University, Pittsburgh, USA. guruswami@cmu.edu. Some of this work was done when visiting the Simons Institute, Berkeley. Research supported in part by NSF grants CCF-0963975 and CCF-1422045.

[†] School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. Some of this work was done when visiting the Simons Institute for the Theory of Computing, Berkeley. jaikumar@tifr.res.in

ISSN 1433-8092

for further discussion of this motivation. The agreement distillation problem also naturally arises in the context of simulating communication protocols that use perfect shared randomness when the parties only share correlated randomness, and was recently studied with this motivation in [CGMS15]. The underlying information-theoretic question, on the maximum information a function of X can convey about its noisy version Y, has also received widespread interest lately, following the appealing conjecture made in [CK14] that a dictator (or canalizing) function is the most informative Boolean function (the one maximizing $I[f(X):Y]$).

Our work is a follow-up to [CGMS15, BM11] and is motivated by questions such as: How many bits of communication are needed for the agreement distillation task to succeed with high probability? At the other extreme, what is the best success probability of a strategy that involves no communication? More precisely, what is the trade-off between communication and success probability?

Note that there are two trivial protocols: one where Alice simply sends the first k bits of X to Bob (which achieves agreement probability of 1), and a zero-communication protocol where both players simply output their first k bits as the common randomness (which achieves agreement probability of $(1-\varepsilon)^k$). The former protocol does not exploit the fact that Bob holds a string Y which is correlated with X. How much can we leverage this to save on communication while at the same time ensuring good agreement probability? A simple protocol based on capacity-achieving codes for the binary symmetric channel was given in [CGMS15] with communication $(h(\varepsilon)+o(1))k$ and high agreement probability; in [CGMS15], an $\Omega(\varepsilon k)$ lower bound based on [BM11] was also observed. This established that a factor $c(\varepsilon)$ savings in communication is the best one can hope for, but left a gap even in the asymptotic growth of $c(\varepsilon)$.

1.1 Our results

We obtain tight communication complexity upper and lower bounds for the above problem, identifying the precise trade-off between communication and agreement probability (see Theorem 1.1 below, our bounds are sharp up to $o(k)$ bits). Our upper bounds are achieved by one-way communication protocols where Alice sends a single message to Bob. Our lower bounds hold for a slightly more general model where Alice's output depends only on her input X, but Bob's output may depended on his input Y and the transcript of an arbitrary two-way interaction with Alice. Below is a statement of the bounds we get.

Theorem 1.1. *Let $\gamma \in [0,1]$, $\varepsilon \in [0,1/2]$, and $k \geqslant 1$ be an integer. Consider the above setting where Alice and Bob have uniformly random strings X and Y (of sufficiently large length compared to k) that differ in each position independently with probability ε. The goal is for Alice and Bob to agree on a shared string $g_A(X)$, which only depends on Alice's input X. Define $C := 4\varepsilon(1-\varepsilon)$.*

- *(Upper bound) There is a protocol where $g_A(X)$ is uniformly distributed in $\{0,1\}^k$, and Alice sends $(C(1-\gamma)-2\sqrt{C(1-C)\gamma})k$ bits to Bob, who then succeeds in guessing $g_A(X)$ with probability at least $2^{-\gamma k - O(\log k)}$.*
- *(Matching lower bound) Suppose there is a protocol with $H_\infty[g_A(X)] \geqslant k$ where Alice and Bob exchange c bits after which Bob is able to guess $g_A(X)$ with probability $2^{-\gamma k}$. Then*

$$c \geqslant (C(1-\gamma)-2\sqrt{C(1-C)\gamma})k\ .$$

In particular, this implies that for large k, to achieve agreement probability $\Theta(1)$ the optimal communication approaches $4\varepsilon(1-\varepsilon)k$, and with zero communication the best achievable success probability approaches $2^{-\frac{\varepsilon}{1-\varepsilon}k}$.[1]

Note that in the above setup, Bob's input Y can be viewed as X that is distorted by a binary symmetric channel, $\mathsf{BSC}(\varepsilon)$, which flips each bit independently with probability ε. Inspired by this view, one can consider a similar problem for other discrete memoryless channels relating X and Y. We consider the binary erasure channel, $\mathsf{BEC}(\varepsilon)$, where each Y_j equals X_j with probability $1-\varepsilon$ and is erased (say, replaced by a '?') with probability ε, and obtain tight upper and lower bounds for this setting as well (basically the quantity $C = 4\varepsilon(1-\varepsilon)$ is replaced by ε in the bounds).

Theorem 1.2. *Let $\gamma, \varepsilon \in [0,1]$ and $k \geqslant 1$ be an integer. For the agreement distillation problem when Y is obtained by passing X through $\mathsf{BEC}(\varepsilon)$, the following hold.*

- *(Upper bound) There is a protocol where $g_A(X)$ is uniformly distributed in $\{0,1\}^k$, and Alice sends $(\varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma})k$ bits to Bob, who succeeds in guessing $g_A(X)$ with probability at least $2^{-\gamma k - O(\log k)}$.*
- *(Matching lower bound) Suppose there is a protocol where $H[g_A(X)] \geqslant k$ and Alice and Bob exchange c bits after which Bob is able to guess $g_A(X)$ with probability $2^{-\gamma k}$. Then $c \geqslant (\varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma})k$.*

In particular, it can be shown that, for large k, to achieve agreement probability $\Theta(1)$ the optimal communication approaches εk, and with zero communication the best achievable success probability approaches $2^{-\frac{(1-\sqrt{1-\varepsilon})k}{1+\sqrt{1-\varepsilon}}}$.

We also study information complexity bounds, proving the following lower bound on the information content needed in the protocol transcript.

Theorem 1.3. *Let $g_A(X)$ take values in the set $\{0,1\}^{k'}$ such that $H[g_A(X)] \geqslant k$. Suppose $\pi(X,Y)$ is the transcript of a protocol that enables Bob to guess $g_A(X)$ with probability at least $1-\delta$, for some $\delta \in [0,1]$. Then we have*

- *$H[\pi(X,Y)] \geqslant 4\varepsilon(1-\varepsilon)k - \delta k' - h(\delta)$ when Y is the output of $\mathsf{BSC}(\varepsilon)$ on X;*
- *$H[\pi(X,Y)] \geqslant \varepsilon k - \delta k' - h(\delta)$ when Y is the output of $\mathsf{BEC}(\varepsilon)$ on X.*

[Note that some term like $-\delta k'$ in the lower bounds is unavoidable. For example, $g_A(X)$ might be $0^{k'}$ with probability $1-\delta$ and a uniformly random string in $\{0,1\}^{k'}$ with probability δ. If Bob produces $0^{k'}$ always, they agree with probability $1-\delta$.]

Since the entropy $H[\pi(X,Y)]$ lower bounds the length of the transcript, the above also implies lower bounds on the communication complexity. However, the bounds are good only when $\delta \to 0$, whereas Theorems 1.1 and 1.2 apply even when the success probability $1-\delta$ is very small, and imply communication lower bounds of $(4\varepsilon(1-\varepsilon) - o(1))k$ and $(\varepsilon - o(1))k$ for any constant success probability.

The communication upper bounds from Theorems 1.1 and 1.2 of course imply protocols with the same upper bounds on entropy. In particular, when the failure probability $\delta \to 0$, the optimal entropy of the transcript of an agreement distillation protocol approaches $4\varepsilon(1-\varepsilon)k$ for $\mathsf{BSC}(\varepsilon)$ and εk for $\mathsf{BEC}(\varepsilon)$.

[1] For the problem with zero communication, lower and upper bounds in [BM11] already establish that the best probability of success is $2^{-\frac{\varepsilon}{1-\varepsilon}k}$ (see Section 1.2).

1.2 Prior and related work

The variant of agreement distillation where the goal of the two parties is to extract a single bit without any interaction was studied independently a number of times; see [Yan07] and references therein. It is known that in this case the optimal protocol is for the two parties to use the first bit. The works [MO05, MOR$^+$06] consider the problem of extracting a common random bit in the multi-party setting where m players receive noisy versions of a common random string; in this case for large m the majority function is close to being optimal in terms of maximizing the agreement probability. The problem of two parties agreeing on k random bits without any communication, when given strings X,Y correlated via $\mathsf{BSC}(\varepsilon)$, was considered by Bogdanov and Mossel [BM11]. They proved that no strategy can achieve agreement probability better than $2^{-k\varepsilon/(1-\varepsilon)}$ and also gave a protocol with agreement probability $O((k\varepsilon)^{-1/2}\cdot 2^{-k\varepsilon/(1-\varepsilon)})$ when $k \geqslant \Omega(1/\varepsilon)$.

All these results are for the model where no communication is allowed between Alice and Bob, and the goal is to maximize the agreement probability. Canonne et al. [CGMS15] considered the setting where Alice and Bob can communicate, and gave a simple scheme based on capacity-achieving codes for agreeing on k random bits with high probability when Alice sends a single message of $(h(\varepsilon)+o(1))k$ bits to Bob. They also noted an $\Omega(\varepsilon k)$ lower bound based on the agreement probability upper bound for zero communication protocols from [BM11]. Zhao and Chia [ZC11] establish that to agree with high probability on a common random variable K with *Shannon entropy* $H[K] \geqslant k$, the communication required approaches precisely $(1-\rho^2(X_1;Y_1))k$, where $\rho(A;B)$ is the Hirschfeld-Gebelein-Rényi (HGR) maximal correlation of the pair $(A;B)$ of random variables. The HGR correlation for $\mathsf{BSC}(\varepsilon)$ (resp. $\mathsf{BEC}(\varepsilon)$) equals $1-2\varepsilon$ (resp. $\sqrt{1-\varepsilon}$), so this implies the communication bounds of Theorems 1.1 and 1.2, albeit for the setting of ensuring high Shannon entropy and agreement probability tending to 1. The Shannon entropy of a random variable is lower bounded by its min-entropy, so a lower bound for distilling randomness with Shannon entropy k implies the same lower bound for min-entropy (our setting). But note that our lower bounds hold also for success probability bounded away from 1, for which we have to rely on hypercontractivity based arguments. Indeed, the main novelty in our results is the establishment of the precise trade-off between communication and probability of agreement.

Our work focuses only on the efficiency of shared randomness generation as a function of communication (and success probability). We allow the number of correlated samples $N \to \infty$ for any desired value of k, the number of shared random bits to be generated (indeed in our protocols as presented, N will be exponential in k and we did not try to optimize this trade-off). Prior work has also studied the efficiency of common randomness generation as a function of N [AC98, ZC11], specifically understanding the "CR capacity" $C(R)$ wherein $C(R)N$ bits of shared randomness can be generated (with high probability) using RN bits of communication, for a fixed $R > 0$ and growing N.[2]

Turning to our information-theoretic results, the entropy lower bound in Theorem 1.3 for $\mathsf{BSC}(\varepsilon)$ is based on the following claim. Let $X^n, Y^n \in \{0,1\}^n$ be random strings with (X_i,Y_i) being i.i.d. and related via the channel $\mathsf{BSC}(\varepsilon)$. Then, for every function $g_A : \{0,1\}^n \to \{0,1\}^k$ we have $I[g_A(X^n) : Y^n] \leqslant (1-2\varepsilon)^2 k$. This upper bound on mutual information follows from the so-called Mrs. Gerber's Lemma [WZ73]; such an upper bound was established using a limit argument in [CT14], and is attributed to Erkip [Erk96] in [CK14].

The earlier mentioned conjecture from [CK14] on the most informative Boolean function asserts that when $k = 1$, we have $I[g_A(X^n) : Y^n] \leqslant 1 - h(\varepsilon)$. If this conjecture were true for every k, then one would

[2] With zero communication, it is not possible to distill any common randomness with high probability, unless the joint distribution of X_1 and Y_1 is decomposable, which is captured by the HGR maximal correlation $\rho(X_1,Y_1)$ equaling 1 [GK72, Rom00, MM12].

have $I[g_A(X^n):Y^n] \leqslant (1-h(\varepsilon))k$ when the range of g_A is $\{0,1\}^k$. However, our communication protocol in Theorem 1.1 implies the existence of a function g_A for which

$$I[g_A(X^n):Y^n] = H[g_A(X^n)] - H[g_A(X^n) \mid Y^n] \geqslant k - (4\varepsilon(1-\varepsilon)+o(1))k = (1-2\varepsilon)^2 k - o(k) > (1-h(\varepsilon))k$$

(for $\varepsilon \in (0,1/2)$). So for functions outputting a large number k of bits, the projection onto the first k bits is *not* the most informative function. This latter result was already established in the recent work [CT14], where a function $g_A(X^n)$ based on lossy data compression (under Hamming distortion) was shown to achieve $\liminf_{n\to\infty} I[g_A(X^n):Y^n] \geqslant (1-2\varepsilon)^2 k$.

Our entropy lower bound in Theorem 1.3 for the case of $\mathsf{BEC}(\varepsilon)$ is based on the inequality $I[g_A(X^n):Y^n] \leqslant (1-\varepsilon)k$ for an arbitrary function $g_A : \{0,1\}^n \to \{0,1\}$, which we establish using Shearer's lemma. So, for the erasure channel, outputting the first k bits indeed maximizes the information about the channel output Y^n, for every $k \geqslant 1$, and in particular the dictator is the most informative function when $k=1$.

As an appealing conjecture bridging information theory and analysis of Boolean functions, the most informative function conjecture of Courtade and Kumar [CK14] has generated a lot of interest. Closely related problems were studied earlier by Erkip and Cover [EC98], and recent works addressing aspects of the Courtade-Kumar conjecture include [AGKN13a, AGKN14, CT14, OSW15, KOW15, Sam15b, Sam15a].

1.3 Techniques in brief

Our communication protocols are extensions of the Bogdanov-Mossel protocol [BM11]. Their zero communication protocol for $\mathsf{BSC}(\varepsilon)$ was based on an "affine covering code" $C \subseteq \mathbb{F}_2^n$ of size 2^k, and both Alice and Bob rounded their inputs X^n and Y^n to the closest point in C (with some explicit rule in case of ties). The probabilistic method is used to establish the existence of an affine space of $\mathbb{F}_2^n$ of dimension k such that each output is generated with the same probability 2^{-k}, and the agreement probability is high (at least $\approx 2^{-\varepsilon k/(1-\varepsilon)}$). In our scheme, we use different functions for Alice and Bob, with Bob searching for a codeword in a larger radius. This will lead to a list of candidates on Bob's side, and he will use Alice's message to pick a unique element from the list. Picking parameters carefully leads us to the protocol with the optimal trade-off between communication and agreement probability claimed in Theorem 1.1. The protocol for the erasure case in Theorem 1.2 works similarly, with the analysis handling some technicalities by conditioning on the high probability event of Y having close to εN erasures.

Turning to our lower bounds, as mentioned above, our entropy lower bounds are based on Mrs. Gerber's lemma for $\mathsf{BSC}(\varepsilon)$ and Shearer's lemma for $\mathsf{BEC}(\varepsilon)$. Our communication lower bounds rely on hypercontractive inequalities for the random variables corresponding to $\mathsf{BSC}(\varepsilon)$ and $\mathsf{BEC}(\varepsilon)$. If (X_i,Y_i) are i.i.d. copies of a correlated random variable (X,Y), and $f : X^n \to \mathbb{R}$, such a hypercontractive inequality upper bounds $\|\mathbb{E}[f(X)|Y]\|_q$ by the norm $\|f\|_p$ with $p<q$ (see Section 4.1 for the definition of these norms). The best possible relationship between p and q depends on amount of correlation between X and Y. For $\mathsf{BSC}(\varepsilon)$, it is a classical result in the analysis of Boolean functions that one can take $p = 1+(1-2\varepsilon)^2(q-1)$ [O'D14, Chap. 16]. The inequality for the erasure channel does not appear to have been studied before, and we use the bound $p = 1+(1-\varepsilon)(q-1)$, shown to be valid for $1 \leqslant q \leqslant 3$ by Nair [NW15], prompted by our application.

The lower bound for zero-communication in [BM11] was also established using hypercontractivity. The reduction to an hypercontractive inequality was more direct in their case, as the success probability can be expressed as $\mathbb{E}_{(X,Y)}[g_A(X)g_B(Y)]$ which equals an inner product $\mathbb{E}_X[g_A(X)T_{1-2\varepsilon}g_B(X)]$ for the

Bonami-Beckner noise operator $T_{1-2\varepsilon}$. When Alice is allowed to send a message to Bob, we need a bit more care in applying the hypercontractive inequality to deduce the lower bound. Also, as mentioned earlier, for the case of erasures, the requisite hypercontractive inequality seems to not have been studied before.

It is natural to wonder what the situation is for more general channels besides the BSC and the BEC. The lower bound on communication to achieve constant agreement probability, which approaches $4\varepsilon(1-\varepsilon)k$ and εk respectively for $\mathsf{BSC}(\varepsilon)$ and $\mathsf{BEC}(\varepsilon)$, arises from the limiting ratio $\frac{p-1}{q-1}$ as $q \downarrow 1$. For an arbitrary discrete channel $(X,Y) \sim p(x,y)$, this limit has been shown to equal

$$s^*(Y;X) := \sup_{r(y) \neq p(y)} \frac{D(r(x)||p(x))}{D(r(y)||p(y))} \tag{1.1}$$

where $r(x)$ denotes the x-marginal distribution of $r(x,y) = r(y)p(x|y)$ [AGKN13b]. Our methods imply a communication lower bound of $(1-s^*(Y;X))k - o(k)$ for an arbitrary channel, though we do not know if this is tight in general.

2 The model

Alice receives a random string $X = (X_1, X_2, \dots, X_N)$ and Bob receives a (correlated) string $Y = (Y_1, Y_2, \dots, Y_N)$. We will assume the length N of these strings is sufficiently large, but it will otherwise not play an important role (and will be mostly suppressed) in our arguments. Alice uses her random input string X to produce an output in $\{0,1\}^{k'}$. Then, based on the inputs, Alice and Bob interact using a two-party protocol σ to produce a transcript $\sigma(X,Y)$. Finally, Bob produces an output in $\{0,1\}^{k'}$ based on his input Y and $\sigma(X,Y)$. Their goal is to ensure that the outputs agree and have high min-entropy.

Definition 2.1. *A (k',k,η,R)-agreement distillation protocol for a pair of random variables $R=(X,Y)$ is a triple (g_A, g_B, σ), where σ is a two-party protocol and $g_A(X), g_B(Y, \sigma(X,Y)) \in \{0,1\}^{k'}$, such that*

1. $H_\infty[g_A(X)] \geqslant k$;
2. $\Pr[g_A(X) = g_B(Y, \sigma(X,Y))] \geqslant \eta$.

Let $\Pi(k',k,\eta,R)$ be the collection of all (k',k,η,R)-protocols. For $\pi \in \Pi(k',k,\eta,R)$, let $\pi(X,Y)$ denote the transcript of the underlying two-party protocol on input (X,Y). Let

$$h^R(k',k,\eta) = \min_{\pi \in \Pi(k',k,\eta,R)} H[\pi(X,Y)]; \tag{2.1}$$

$$c^R(k',k,\eta) = \min_{\pi \in \Pi(k',k,\eta,R)} \max_{x,y} |\pi(x,y)|. \tag{2.2}$$

We will consider two joint distributions of $R=(X,Y)$ in this work, where (X_i,Y_i) are independently generated as follows.

- Binary symmetric channel, $\mathsf{BSC}(N,\varepsilon)$: X_i is uniform in $\{0,1\}$, and $Y_i = X_i$ with probability $(1-\varepsilon)$ and $Y_i = 1 - X_i$ with probability ε.
- Binary erasure channel, $\mathsf{BEC}(N,\varepsilon)$: X_i is uniform in $\{0,1\}$, and $Y_i = X_i$ with probability $(1-\varepsilon)$ and $Y_i = ?$ with probability ε.

3 The entropy bounds

In this section, we show the following, which implies the lower bounds claimed in Theorem 1.3.

Theorem 3.1. *We have the following lower bounds:*

$$\begin{aligned} h^{\mathsf{BSC}(N,\varepsilon)}(k',k,\eta) &\geqslant 4\varepsilon(1-\varepsilon)k-(1-\eta)k'-h(\eta); \\ h^{\mathsf{BEC}(N,\varepsilon)}(k',k,\eta) &\geqslant \varepsilon k-(1-\eta)k'-h(\eta). \end{aligned}$$

Both parts of the theorem will be justified using the following idea. The channel limits the mutual information between Alice's output and Bob's input. Alice's message must, therefore, make up for the shortfall.

Claim 3.2. *(a) If $(X,Y)\sim\mathsf{BSC}(N,\varepsilon)$, then*

$$I[g_A(X):Y]\leqslant(1-2\varepsilon)^2 I[g_A(X):X]=(1-2\varepsilon)^2 H[g_A(X)]. \tag{3.1}$$

(b) If $(X,Y)\sim\mathsf{BEC}(N,\varepsilon)$, then

$$I[g_A(X):Y]\leqslant(1-\varepsilon)I[g_A(X):X]=(1-\varepsilon)H[g_A(X)]. \tag{3.2}$$

Proof of Theorem 3.1. First, we have

$$\begin{aligned} \mathbb{E}[|\Pi(X,Y)|] &\geqslant H[\Pi(X,Y)] \\ &\geqslant I[\Pi(X,Y):g_A(X)Y] \\ &= I[g_A(X):\Pi(X,Y)Y]-I[Y:g_A(X)]+I[Y:\Pi(X,Y)] \\ &\geqslant H[g_A(X)]-H[g_A(X)\mid\Pi(X,Y)Y]-I[Y:g_A(X)] \qquad (3.3) \\ &\geqslant H[g_A(X)]-h(\eta)-(1-\eta)k'-I[Y:g_A(X)]. \qquad (3.4) \end{aligned}$$

Our assumption implies that $H[g_A(X)]\geqslant k$. We use the claim above to bound the last term on the right.

Binary symmetric channel: From (3.4) and Claim 3.2 (a), we obtain

$$\begin{aligned} \mathbb{E}[|\Pi(X,Y)] &\geqslant (1-(1-2\varepsilon)^2)H[g_A(X)]-h(\eta)-(1-\eta)k' \\ &\geqslant 4\varepsilon(1-\varepsilon)k-(1-\eta)k'-h(\eta). \end{aligned}$$

Erasure channel: From (3.4) and Claim 3.2 (b), we obtain

$$\mathbb{E}[|\Pi(X,Y)]\geqslant(1-(1-\varepsilon))H[g_A(X)]-h(\eta)-(1-\eta)k'\geqslant\varepsilon k-(1-\eta)k'-h(\eta)\ . \qquad \square$$

Proof of Claim 3.2. (a) Recall the following consequence of Mrs. Gerber's Lemma due to Wyner and Ziv [WZ73, Corollary 4]:

> Suppose (X,W) is a pair of random variables, where X takes values in $\{0,1\}^N$ and $H[X\mid W]=Nv$. Let $Z\in\{0,1\}^N$ be sequence of N independent bits, each taking the value 1 with probability ε; let Z be independent of (X,W). Let $Y=X\oplus Z$. Then,
>
> $$H[Y\mid W]\geqslant Nh(\varepsilon*h^{-1}(v)),$$
>
> where h is the binary entropy function and $\varepsilon*v=\varepsilon(1-v)+(1-\varepsilon)v$. Note that $h(\varepsilon*h^{-1}(v))\geqslant1-(1-v)(1-2\varepsilon)^2$ (see, for example, [CT14]).

We take $W = g_A(X)$ in the above statement; then, $H[X \mid W] = H[X \mid g_A(X)] = N - H[g_A(X)]$. So, we set $v = 1 - H[g_A(X)]/N$ and conclude that $H[Y \mid g_A(X)] \geqslant N - (1-2\varepsilon)^2 H[g_A(X)]$. Thus, $I[g_A(X) : Y] = H[Y] - H[Y \mid g_A(X)] \leqslant (1-2\varepsilon)^2 H[g_A(X)]$.

(b) We first derive a version of Shearer's lemma. Let $\mathsf{sgn}(Y)$ be the erasure pattern of Y, that is, a sequence in $\{0,1\}^N$, where the 0s correspond to erasures.

$$\begin{aligned}
H[Y \mid \mathsf{sgn}(Y), g_A(X) = z] &= \mathop{\mathbb{E}}_{\sigma}[H[Y \mid \mathsf{sgn}(Y) = \sigma, g_A(X) = z]] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\sum_{i:\sigma_i=1} H[X_i \mid (X_j : j < i, \sigma_j = 1), g(X) = z]\right] \\
&\geqslant \mathop{\mathbb{E}}_{\sigma}\left[\sum_{i:\sigma_i=1} H[X_i \mid (X_j : j < i), g(X) = z]\right] \\
&= \mathop{\mathbb{E}}_{\sigma}\left[\sum_{i} \mathbf{1}\{\sigma_i = 1\} H[X_i \mid (X_j : j < i), g(X) = z]\right] \\
&= (1-\varepsilon)\sum_i H[X_i \mid (X_j : j < i), g(X) = z] \\
&= (1-\varepsilon) H[X \mid g(X) = z].
\end{aligned}$$

Taking expectations of both sides over choices of z, we obtain $H[Y \mid \mathsf{sgn}(Y) g_A(Y)] \geqslant (1-\varepsilon)H[X \mid g_A(X)]$. Then, we have

$$\begin{aligned}
H[Y \mid g_A(X)] &= H[Y\mathsf{sgn}(Y) \mid g_A(X)] \\
&= H[\mathsf{sgn}(Y)] + H[Y \mid \mathsf{sgn}(Y) g_A(X)] \\
&\geqslant h(\varepsilon)N + (1-\varepsilon)H[X \mid g_A(X)]. \qquad (3.5)
\end{aligned}$$

Thus,

$$\begin{aligned}
I[g_A(X) : Y] &= H[Y] - H[Y \mid g_A(X)] \\
&= h(\varepsilon)N + (1-\varepsilon)N - H[Y \mid g_A(X)] \\
&\leqslant (1-\varepsilon)(N - H[X \mid g_A(X)]) && \text{(using (3.5))} \\
&= (1-\varepsilon)(H[X] - H[X \mid g_A(X)]) \\
&= (1-\varepsilon) I[X : g_A(X)] = (1-\varepsilon) H[g_A(X)]. && \square
\end{aligned}$$

4 The communication lower bounds

We now turn to our lower bounds on communication, formally stated below. Note that these imply the lower bounds claimed in Theorems 1.1 and 1.2.

Theorem 4.1. *Let $\gamma, \varepsilon \in [0,1]$ and $k \geqslant 1$ be an integer.*

$$c^{\mathsf{BSC}(N,\varepsilon)}(k', k, 2^{-\gamma k}) \geqslant \left[C(1-\gamma) - 2\sqrt{C(1-C)\gamma}\right] k \qquad \textit{where } C = 4\varepsilon(1-\varepsilon); \qquad (4.1)$$

$$c^{\mathsf{BEC}(N,\varepsilon)}(k', k, 2^{-\gamma k}) \geqslant \left[\varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma}\right] k\ . \qquad (4.2)$$

The arguments for the two channels, $\mathsf{BSC}(N,\varepsilon)$ and $\mathsf{BEC}(N,\varepsilon)$, differ only in the choice of the appropriate hypercontractive inequality. We, therefore, first present the common part of the argument. Fix a protocol $\pi \in \Pi(k',k,\eta,R)$, where R is either $\mathsf{BSC}(N,\varepsilon)$ or $\mathsf{BEC}(N,\varepsilon)$. Let $\mathscr{T}$ denote the set of possible transcripts of π; let $t = |\mathscr{T}|$. We will obtain a lower bound on t.

Let $\mathscr{X},\mathscr{Y}$ denote the domains of X and Y respectively; $\mathscr{X},\mathscr{Y} = \{0,1\}^N$ for $\mathsf{BSC}(N,\varepsilon)$; $\mathscr{X} = \{0,1\}^N$ and $\mathscr{Y} = \{0,1,?\}^N$ for $\mathsf{BEC}(N,\varepsilon)$. Recall that $g_A(X)$ and $g_B(Y,\pi(X,Y))$ take values in $\mathscr{Z} = \{0,1\}^{k'}$. For $y \in \mathscr{Y}$ and $z \in \mathscr{Z}$, let

$$\beta(z|y) := \Pr[g_A(X) = z \mid Y = y] = \Pr[g_A(X) = z \wedge Y = y]/\Pr[Y = y];$$

let $\mathsf{Success}$ denote the event "$g_A(X) = g_B(Y,\pi(X,Y))$". For $y \in \mathscr{Y}$, let

$$\mathscr{Z}_y = \{g_B(y,\tau) : \tau \in \mathscr{T}\};$$

then, $t_y := |\mathscr{Z}_y| \leqslant t$. On input (x,y), if $g_A(x) \notin \mathscr{Z}_y$, then $\mathsf{Success}$ is impossible. Arrange $z \in \mathscr{Z}_y$ as $z_{y,1}, z_{y,2}, \ldots$ so that $\beta(z_{y,1}|y) \geqslant \beta(z_{y,1}|y) \geqslant \cdots \geqslant \beta(z_{y,t_y}|y)$; let $\beta_{y,i} = \beta(z_{y,i}|y)$.

Claim 4.2. *Let $\pi \in \Pi(R,k,\eta)$ be a protocol with t transcripts and let $q > 1$. Then,*

$$\Pr[\mathsf{Success}] \leqslant \mathop{\mathbb{E}}_{Y}\left[\sum_{i=1}^{t_Y} \beta_{Y,i}\right] \leqslant \left(\sum_{z} \mathop{\mathbb{E}}_{Y}[\beta(z|Y)^q]\right)^{1/q} \cdot t^{1-1/q}. \tag{4.3}$$

Proof. When Alice sends no message, Bob's best strategy on receiving y is to output the "most likely answer"; so, the probability of $\mathsf{Success}$ is at most β_{y_1}. We now generalize this principle to the case where Bob may base his decision on a transcript. We have

$$\begin{aligned}
\Pr[\mathsf{Success} \mid Y = y] &\leqslant \sum_{\tilde{z} \in \mathscr{Z}_y} \Pr[\mathsf{Success} \wedge g_B(Y,\pi(X,Y)) = \tilde{z} \mid Y = y] \\
&\leqslant \sum_{\tilde{z} \in \mathscr{Z}_y} \Pr[g_A(X) = \tilde{z} \mid Y = y] \\
&= \sum_{\tilde{z} \in \mathscr{Z}_y} \beta(\tilde{z}|y) \leqslant \sum_{i=1}^{t_y} \beta_{y,i},
\end{aligned}$$

where the last inequality holds because $\langle \beta_{y,i} : i = 1,2,\ldots,t_y \rangle$ are the top t_y values of $\beta(z|y)$. Thus,

$$\begin{aligned}
\Pr[\mathsf{Success}] &\leqslant \mathop{\mathbb{E}}_{Y}\left[\sum_{i=1}^{t_Y} \beta_{Y,i}\right] && \text{(4.4)} \\
&\leqslant \mathop{\mathbb{E}}_{Y}\left[\left(\sum_{i=1}^{t_Y} \beta_{Y,i}^q\right)^{1/q} t_Y^{1-1/q}\right] && \text{(by Hölder's inequality)} \\
&\leqslant \left(\mathop{\mathbb{E}}_{Y}\left[\sum_{i=1}^{t_Y} \beta_{Y,i}^q\right]\right)^{1/q} \cdot t_Y^{1-1/q} && \text{(by Jensen's inequality)} \quad \text{(4.5)} \\
&\leqslant \left(\sum_{z} \mathop{\mathbb{E}}_{Y}[\beta(z|Y)^q]\right)^{1/q} \cdot t^{1-1/q}. && \square
\end{aligned}$$

4.1 Hypercontractivity

For functions $\alpha : \mathscr{X} \to \mathbb{R}$ and $\beta : \mathscr{Y} \to \mathbb{R}$, let

$$\|\alpha\|_p = \mathop{\mathbb{E}}_X [|\alpha(X)|^p]^{1/p};$$
$$\|\beta\|_q = \mathop{\mathbb{E}}_Y [|\beta(Y)|^q]^{1/q}.$$

For $z \in \mathscr{X}$, let $\mathbf{1}_z$ be the indicator random variable $\mathbf{1}[g_A(X) = z]$ and $\beta_z : \mathscr{Y} \to \mathbb{R}$ be defined by $\beta_z(y) = \beta(z|y) = \mathbb{E}[\mathbf{1}_x(X) \mid Y = y]$. Then,

$$\left(\mathop{\mathbb{E}}_Y [\beta(z|Y)^q]\right)^{1/q} = \|\beta_z\|_q = \|\mathbb{E}[\mathbf{1}_z(X) \mid Y]\|_q.$$

Using this, we may rearrange inequality (4.3) and obtain

$$t \geqslant \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[\sum_z \|\mathbb{E}[\mathbf{1}_z(X) \mid Y]\|_q^q\right]^{-1/(q-1)}. \tag{4.6}$$

Now assume that we have a pair (p,q), $1 \leqslant p < q$, such that for all functions $f : \mathscr{X} \to \mathbb{R}$,

$$\mathbb{E}[f(X) \mid Y]\|_q \leqslant \|f\|_p. \tag{4.7}$$

Later we will choose an appropriate pair (p,q) depending on the channel. Using (4.7) with the function $\mathbf{1}_z$, we obtain

$$\begin{aligned}
t &\geqslant \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[\sum_z \|\mathbf{1}_z\|_p^q\right]^{-1/(q-1)} \\
&= \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[\sum_z \mathsf{Pr}[g_A(X) = z]^{q/p}\right]^{-1/(q-1)} \\
&\geqslant \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[\sum_z \mathsf{Pr}[g_A(X) = z]\,\mathsf{Pr}[g_A(X) = z]^{(q-p)/p}\right]^{-1/(q-1)} \\
&\geqslant \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[2^{-k(q-p)/p} \sum_z \mathsf{Pr}[g_A(X) = z]\right]^{-1/(q-1)} \qquad (\text{since } H_\infty[g_A(X)] \geqslant k) \\
&\geqslant \mathsf{Pr}[\mathsf{Success}]^{q/(q-1)} \left[2^{k(q-p)/p}\right]^{1/(q-1)}.
\end{aligned}$$

The above argument was general, and applicable for any channel where we can find an appropriate pair (p,q) so that (4.7) holds. We now specialize the argument to $\mathsf{BSC}(N,\varepsilon)$ and $\mathsf{BEC}(N,\varepsilon)$.

Binary symmetric channel: In this case, we set $q = 1+\delta$ and $p = 1+(1-2\varepsilon)^2\delta$ [O'D14, Chap. 16]. Then,

$$t \geqslant \mathsf{Pr}[\mathsf{Success}]^{(1+\delta)/\delta} \cdot 2^{4\varepsilon(1-\varepsilon)k/(1+(1-2\varepsilon)^2\delta)}. \tag{4.8}$$

Binary erasure channel: In this case, for $q = 1+\delta$, we can take $p = 1+(1-\varepsilon)\delta$ [NW15], and deduce

$$t \geqslant \mathsf{Pr}[\mathsf{Success}]^{(1+\delta)/\delta} \cdot 2^{\varepsilon k/(1+(1-\varepsilon)\delta)}. \tag{4.9}$$

4.2 The trade-off

Let us fix the success probability at $\eta = 2^{-\gamma k}$ and try to choose δ above so that we obtain the best lower bound on t from (4.8) and (4.9).

Binary symmetric channel: Plugging in $\Pr[\mathsf{Success}] = 2^{-\gamma k}$ into (4.8), we conclude that

$$t \geqslant 2^{r_\gamma^{\mathsf{BSC}(N,\varepsilon)}(\delta)k},$$

where

$$r_\gamma^{\mathsf{BSC}(N,\varepsilon)}(\delta) := \frac{C}{1+(1-C)\delta} - \frac{\gamma}{\delta} - \gamma,$$

and $C = 4\varepsilon(1-\varepsilon)$. We need to choose δ so that $r_\gamma(\delta)$ is maximum. Setting the derivative to zero gives us the optimum choice δ_γ^* for which

$$r_\gamma^{\mathsf{BSC}(N,\varepsilon)}(\delta_\gamma^*) = C(1-\gamma) - 2\sqrt{C(1-C)\gamma} \quad .$$

This justifies our lower bound (4.1) for $\mathsf{BSC}(N,\varepsilon)$.

Note that at $\gamma = 0$ (success probability constant), this quantity is $4\varepsilon(1-\varepsilon)$. As γ increases, $r_\gamma(\delta_\gamma^*)$ decreases monotonically, and becomes 0 when $\gamma = \varepsilon/(1-\varepsilon)$, at which point we may only conclude that $t \geqslant 1$ (which is consistent with the results of Bogdanov and Mossel [BM11] for zero communication).

Erasure channel: The calculations are identical. We obtain

$$t \geqslant 2^{r_\gamma^{\mathsf{BEC}(N,\varepsilon)}(\delta)k},$$

where

$$r_\gamma^{\mathsf{BEC}(N,\varepsilon)}(\delta) := \frac{\varepsilon}{1+(1-\varepsilon)\delta} - \frac{\gamma}{\delta} - \gamma.$$

Fixing γ, we find the optimum value δ_γ^* for δ, such that

$$r_\gamma^{\mathsf{BEC}(N,\varepsilon)}(\delta_\gamma^*) = \varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma}.$$

This justifies our lower bound (4.2) for $\mathsf{BEC}(N,\varepsilon)$. When $\gamma = (1-\sqrt{1-\varepsilon})/(1+\sqrt{1-\varepsilon})$, we obtain $r_\gamma^{\mathsf{BEC}(N,\varepsilon)}(\delta_\gamma^*) = 0$; in the next section we will show that there is indeed a zero communication protocol of $\mathsf{BEC}(N,\varepsilon)$ that succeeds with probability close to $2^{-(1-\sqrt{1-\varepsilon})k/(1+\sqrt{1-\varepsilon})}$.

5 Communication protocols

Our protocols are similar to the protocol of Bogdanov and Mossel [BM11]. We first recall their protocol. Let $Z = \{0,1\}^k$. Alice and Bob use an affine subspace of $\mathbb{F}_2^n$ (where $\mathbb{F}_2 = \{0,1\}$ is the field with two elements) with 2^k vectors $\mathbf{v} = (v_z : z \in \{0,1\}^k)$. We will assume that this subspace is constructed at random, by the following process: pick k linearly independent vectors $w_1, w_2, \ldots, w_k$ uniformly at random and another random vector $w_0 \in \{0,1\}^N$; then set

$$v_z = w_0 + \sum_{i=1}^{k} z_i w_i.$$

Note, in particular, that if $z, z' \in \{0,1\}^k$ and $z \neq z'$, then $(v_z, v_{z'})$ ranges uniformly over $\{0,1\}^N \times \{0,1\}^N$.

On receiving $X \in \{0,1\}^n$, Alice's output $g_A(X)$ will be the $z \in Z$ for which v_z is closest to X. To break ties, the following rule is used. Fix a total ordering $\preceq$ on $\{0,1\}^N$ such that if the Hamming weight of x is less than the Hamming weight of x', then $x \preceq x'$. Then, $g_A(x) = z$ for which $x + v_z$ is the smallest with respect to $\preceq$. For this function, Bogdanov and Mossel [BM11] show the following.

Lemma 5.1. *For all $z \in \{0,1\}^k$, we have* $\Pr_X[f(X) = z] = 2^{-k}$.

In the original protocol of Bogdanov and Mossel, Bob uses the same function as Alice to produce his output. We extend the above protocol, allowing Alice to send a short message to Bob. Fix a function $\chi : Z \to \{0,1\}^{ck}$ such that $|\chi^{-1}(\alpha)| = 2^{(1-c)k}$ for all $\alpha \in \{0,1\}^{ck}$. Alice's message to Bob is then $m = \chi(g_A(X))$. On receiving the message m, Bob's output is $z \in \chi^{-1}(m)$ for which v_z agrees most with Y (breaking ties arbitrarily).

It will be convenient to state our proofs using $\{+1,-1\}$ instead of $\{0,1\}$; so we assume that the vectors v_z and the random string X take values in $\{+1,-1\}^N \subseteq \mathbb{R}^N$. If the channel is $\mathsf{BSC}(\varepsilon)$, then we will assume that $Y \in \{+1,-1\}^N$; if the channel is $\mathsf{BEC}(\varepsilon)$, then we will assume that $Y \in \{+1,-1,0\}^N$, where 0 corresponds to erasures. Also, we will assume that $\varepsilon \neq 0$, for Alice and Bob have identical strings and they can just out the first k bits.

5.1 Agreement distillation protocol for $\mathsf{BSC}(\varepsilon)$

We fix $\gamma > 0$, and describe a protocol with low communication that achieves success probability $2^{-\gamma k - o(k)}$. We will do the computation assuming that the affine space of vectors $\mathbf{v}$ is chosen at random. The overall success probability then is averaged over the random choices of the affine subspace. Clearly, there is a choice of an affine subspace where the success probability is at least this average.

Fix $z \in Z$. Note that the quantity $X \cdot v_z = \sum_i X[i] v_z[i]$ is then a sum of N independent random variables taking values in $\{+1,-1\}$, such that $\mathbb{E}[X \cdot v_z] = 0$ and $\mathbf{var}[X \cdot v_z] = N$. To estimate the probabilities, we will assume that N is large and use the normal approximation. Let

$$\varphi(r) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{r^2}{2}\right);$$
$$\Phi^c(r) = \int_r^\infty \varphi(x)\mathrm{d}x.$$

Theorem 5.2 (Berry-Esseen theorem [Fel71, Sec. XVI.5, Theorem 2]). *Let $S = \xi_1 + \xi_2 + \cdots + \xi_N$, where the ξ_i are independent random variables. Suppose $\mu_i = \mathbb{E}[\xi_i]$, $\sigma_i^2 = \mathbf{var}[\xi_i]$ and $\tau_i = \mathbb{E}[|\xi_i - \mu_i|^3]$. Let $\mu = \mathbb{E}[S] = \sum_i \mu_i$ and $\sigma^2 = \mathbf{var}[S] = \sum_i \sigma_i^2$ and $\tau = \sum_i \tau_i$. Then,*

$$|\Pr[S \geqslant \mu + r\sigma] - \Phi^c(r)| \leqslant \frac{6\tau}{\sigma^3}. \tag{5.1}$$

In all our applications $\sigma^2 = \Theta(N)$ (the constant depends on ε) and $\tau \leqslant N$; thus, the right hand side is $O(1/\sqrt{N})$, where the implicit constant depends only on ε and is positive if $\varepsilon \in (0,1)$ is positive. In particular, using standard estimates for $\Phi^c(r)$ (see, for example, [Coo09]), we conclude that for all $r > 0$ and all large enough N

$$\frac{r^2}{r^2+1}\varphi(r) < \Pr[S \geqslant \mu + r\sigma] < \varphi(r) \tag{5.2}$$

Thus, for all large N, one has

$$\varphi(r)\left(1-O(r^{-2})\right) < \Pr[X \cdot v_z \geqslant r\sqrt{N}] < \varphi(r)\ .$$

Note the following behavior of φ when its argument is scaled.

$$\varphi(\alpha r) = \frac{1}{\alpha(\sqrt{2\pi}r)^{1-\alpha^2}}\varphi(r)^{\alpha^2}. \tag{5.3}$$

Let

$$\eta = 2\varepsilon - \sqrt{4\varepsilon(1-\varepsilon)\gamma}.$$

For $z \in \{0,1\}^k$, let

$$A_z := \left\{x \in \{+1,-1\}^n : v_z \cdot x \geqslant r\sqrt{N}\right\};$$
$$B_z := \left\{x \in \{+1,-1\}^n : v_z \cdot x \geqslant (1-\eta)r\sqrt{N}\right\}.$$

Fix $r = \Theta(\sqrt{k})$, such that for all large enough N

$$2^{-(k+1)} \leqslant \mu(A_z) \leqslant 2^{-(k+1)}\left(1+O(1/k)\right)\ . \tag{5.4}$$

Consider the following events for $z \in \{0,1\}^k$.

$$\mathscr{E}_1(z) := (X,Y) \in A_z \times B_z;$$
$$\mathscr{E}_2(z) := \forall z' \neq z : X \notin A_{z'};$$
$$\mathscr{E}_3(z) := \forall z' \neq z\,(\chi(z) = \chi(z')) : Y \notin B_{z'}.$$

$\mathscr{E}_1(z)$ and $\mathscr{E}_2(z)$ ensure that Alice outputs z; $\mathscr{E}_1(z)$ and $\mathscr{E}_3(z)$ ensure that Bob outputs z; thus, if all three events hold, then Alice and Bob both output the string z. Thus,

$$\Pr[\mathsf{Success}] \geqslant \sum_z \Pr[\mathscr{E}_1(z)]\left(1 - \Pr[\overline{\mathscr{E}_2(z)} \mid \mathscr{E}_1(z)] - \Pr[\overline{\mathscr{E}_3(z)} \mid \mathscr{E}_1(z)]\right). \tag{5.5}$$

We will estimate the probabilities appearing on the right separately.

First, we have

$$\Pr[\mathscr{E}_1(z)] = \Pr[X \in A_z] \cdot \Pr[Y \in B_z \mid X \in A_z]. \tag{5.6}$$

For our choice of r (see (5.4)), $\Pr[X \in A_z] \geqslant 2^{-(k+1)}$. To compute the second factor, fix $\mathbf{v}$ (the affine space of 2^k vectors) and $x \in A_z$; say $x \cdot v_z = r'\sqrt{N}$ for some $r' \geqslant r$. Now, $Y \cdot v_z$ is the sum of N independent random variables taking values in $\{+1,-1\}$, such that $\mathbb{E}[Y \cdot v_z] = (1-2\varepsilon)r'\sqrt{N}$ and $\mathbf{var}[Y \cdot v_z] = 4\varepsilon(1-\varepsilon)N$. Thus,

$$\begin{aligned}
\Pr[Y \cdot v_z \geqslant (1-\eta)r\sqrt{N} \mid X = x] &\geqslant \varphi\left(\frac{(1-\eta)r-(1-2\varepsilon)r'}{\sqrt{4\varepsilon(1-\varepsilon)}}\right)(1-O(1/k)) \\
&\geqslant \varphi\left(\frac{(1-\eta)r-(1-2\varepsilon)r}{\sqrt{4\varepsilon(1-\varepsilon)}}\right)(1-O(1/k)) \quad \text{(since } \varphi(r) \text{ is decreasing)} \\
&= \varphi\left(\frac{(2\varepsilon-\eta)r}{\sqrt{4\varepsilon(1-\varepsilon)}}\right)(1-O(1/k)) \\
&= \varphi\left(\sqrt{\gamma}r\right)(1-O(1/k)) \\
&\geqslant \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}}2^{-\gamma(k+1)}(1-O(1/k)).
\end{aligned} \tag{5.7}$$

Using these in (5.6), we obtain

$$\Pr[\mathscr{E}_1(z)] = \Pr[X \in A_z] \cdot \Pr[Y \in B_z \mid X \in A_z] \geqslant \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}} 2^{-(\gamma+1)(k+1)}(1 - O(1/k)). \tag{5.8}$$

Recall that if $z \neq z'$, then as $\mathbf{v}$ varies, v_z and $v_{z'}$ vary uniformly over all pairs of distinct vectors in $\{+1,-1\}^N$. It follows that

$$\begin{aligned} \Pr[\overline{\mathscr{E}_2(z)} \mid \mathscr{E}_1(z)] &\leqslant \sum_{z' : z' \neq z} \Pr[0^N \in A_{z'}] \\ &\leqslant (2^k - 1) \cdot 2^{-(k+1)}(1 + O(k^{-1})) \\ &\leqslant \frac{1}{2}(1 + O(k^{-1})). \end{aligned} \tag{5.9}$$

Similarly, we have

$$\begin{aligned} \Pr[\overline{\mathscr{E}_3(z)} \mid \mathscr{E}_1(z)] &\leqslant \sum_{z' : \chi(z) = \chi(z'), z' \neq z} \Pr[Y \in B_{z'}] \\ &\leqslant 2^{(1-c)k} \varphi((1-\eta)r) \\ &\leqslant 2^{(1-c)k} \frac{1}{(1-\eta)(\sqrt{2\pi}r)^{\eta(2-\eta)}} 2^{-(1-\eta)^2(k+1)}. \qquad \text{(see (5.3) above)} \end{aligned} \tag{5.10}$$

Thus, if $c \geqslant 1 - (1-\eta)^2 = C(1-\gamma) - 2\sqrt{C(1-C)\gamma}$ where $C = 4\varepsilon(1-\varepsilon)$, then this quantity is at most $\frac{1}{4}$ (say) for all large k. It follows from (5.5), (5.8), (5.9) and (5.10) that

$$\begin{aligned} \Pr_{\mathbf{v},X,Y}[\mathsf{Success}] &\geqslant \sum_z \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}} 2^{-(\gamma+1)(k+1)} \left(1 - \frac{1}{2} - \frac{1}{4}\right)(1 - O(k^{-1})) \\ &= 2^{-\gamma k - O(\log \gamma k)} \\ &= 2^{-\gamma k(1+o(1))}. \end{aligned}$$

Thus, there exists a choice of the subspace $\mathbf{v}$ such that Alice and Bob succeed with probability at least $2^{-\gamma k(1+o(1))}$.

Constant probability of success: The above argument, was carried out with $\gamma > 0$ a constant, so that it yielded agreement with probability $2^{-\gamma k(1+o(1))}$. We may, in fact, set $\gamma = 1/r^2 = \Theta(1/k)$ in the above argument, and conclude that with communication $ck \approx (C(1-\gamma) - 2\sqrt{C(1-C)\gamma})k = 4\varepsilon(1-\varepsilon)k - \Theta(\sqrt{k})$, we obtain $\Pr_{\mathbf{v},X,Y}[\mathsf{Success}] = \Omega(1)$.

5.2 Agreement distillation protocol for $\mathsf{BEC}(\varepsilon)$

The calculations are similar to the one we used above. We fix r and A_z as before. However, this time we set $\eta = \varepsilon - \sqrt{\varepsilon(1-\varepsilon)\gamma}$ and let

$$B_z := \left\{ x \in \{+1,-1\}^n : v_z \cdot x \geqslant (1-\eta)r\sqrt{N} \right\}.$$

We define events $\mathscr{E}_1(z)$, $\mathscr{E}_2(z)$ and $\mathscr{E}_3(z)$ as before, and observe that

$$\Pr[\mathsf{Success}] \geqslant \sum_z \Pr[\mathscr{E}_1(z)](1 - \Pr[\overline{\mathscr{E}_2(z)} \mid \mathscr{E}_1(z)] - \Pr[\overline{\mathscr{E}_3(z)} \mid \mathscr{E}_1(z)]). \tag{5.11}$$

continues to hold. To estimate the first factor, we expand it as before and obtain

$$\Pr[\mathscr{E}_1(z)] = \Pr[X \in A_z] \cdot \Pr[Y \in B_z \mid X \in A_z].$$

$\Pr[X \in A_z] \geqslant 2^{-(k+1)}$. As before, for each fixed $x \in A_z$ (such that $x \cdot v_z = r'\sqrt{N}$, $r' \geqslant r$), we view $Y \cdot v_z$ as a sum of N independent random variables, each taking values in either $\{0,+1\}$ or $\{0,-1\}$; in particular, $\mathbb{E}[Y \cdot v_z] = (1-\varepsilon)r'\sqrt{N}$ and $\mathbf{var}[Y \cdot v_z] = \varepsilon(1-\varepsilon)N$. Thus,

$$\begin{aligned}\Pr[Y \cdot v_z \geqslant (1-\eta)r \mid X = x] &\geqslant \varphi\left(\frac{(1-\eta)r-(1-\varepsilon)r'}{\sqrt{\varepsilon(1-\varepsilon)}}\right)(1-O(k^{-1})) \\ &\geqslant \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}} 2^{-\gamma(k+1)}(1-O(k^{-1})).\end{aligned}$$

using calculations identical to those leading to (5.7). We finally have the following lower bound for the first factor of (5.11).

$$\begin{aligned}\Pr[\mathscr{E}_1(z)] &= \Pr[X \in A_z] \cdot \Pr[Y \in B_z \mid X \in A_z] \\ &\geqslant \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}} 2^{-(\gamma+1)(k+1)}(1-O(k^{-1})). \qquad (5.12)\end{aligned}$$

Calculations that lead to

$$\Pr[\overline{\mathscr{E}_2(z)} \mid \mathscr{E}_1(z)] \leqslant \frac{1}{2}(1+O(k^{-1})) \qquad (5.13)$$

remain the same.

Finally, we consider $\mathscr{E}_3(z)$. First, we observe that since N is large, we may assume that with probability tending to 1, the number of ones in Y is $(1-\varepsilon)N \pm N^{3/4}$ (say), even when conditioning on $\mathscr{E}_1$. Now, the pair $(v_z, v_{z'})$ is uniformly distributed over all possible pairs of distinct vectors. So, we will fix v_z, and assume that $v_{z'}$ is uniformly distributed in $\{+1,-1\}^N$ (that it cannot be v_z can be overlooked). Fix y with say $\ell = (\varepsilon + N^{-1/4})N = \varepsilon' N$ zeroes. Then, $Y \cdot v_{z'}$ is the sum of $(1-\varepsilon')N$ independent random variables, each taking values uniformly in $\{+1,-1\}$. In particular, $\mathbb{E}[Y \cdot v_{z'}] = 0$ and $\mathbf{var}[Y \cdot v_{z'}] = (1-\varepsilon')N$. Then,

$$\begin{aligned}\Pr[\overline{\mathscr{E}_3(z)} \mid \mathscr{E}_1(z)] &\leqslant \sum_{z':\chi(z)=\chi(z'),z'\neq z} \Pr[Y \in B_{z'}] \\ &\leqslant 2^{(1-c)k}\varphi\left(\frac{(1-\eta)}{\sqrt{1-\varepsilon'}}r\right)\left(1+O(k^{-1})\right) \\ &\leqslant 2^{(1-c)k}\frac{1}{(1-\eta)(\sqrt{2\pi}r)^{1-(1-\eta)^2/(1-\varepsilon')}}2^{-(1-\eta)^2(k+1)/(1-\varepsilon')}\left(1+O(k^{-1})\right). \quad \text{(see (5.3) above)}\end{aligned}$$
$$(5.14)$$

Thus, if $c \geqslant 1-(1-\eta)^2/(1-\varepsilon') = \varepsilon(1-\gamma) - 2\sqrt{\varepsilon(1-\varepsilon)\gamma}$, then this quantity is at most $\frac{1}{4}$ (say) for all large k. It follows from (5.11), (5.12), (5.13) and (5.14) that

$$\begin{aligned}\Pr_{\mathbf{v},X,Y}[\mathsf{Success}] &\geqslant \sum_z \frac{1}{\sqrt{\gamma}(\sqrt{2\pi}r)^{1-\gamma}} 2^{-(\gamma+1)(k+1)}\left(1-\frac{1}{2}-\frac{1}{4}\right)(1-O(k^{-1})) \\ &= 2^{-\gamma k - O(\log \gamma k)}.\end{aligned}$$

We may, as before, fix a choice of $\mathbf{v}$ such that Alice and Bob succeed with probability at least $2^{-\gamma k(1+o(1))}$.

Constant probability of success: Again, we may set $\gamma = 1/r^2 = \Theta(1/k)$ in the above argument, and conclude that with communication $\varepsilon k - \Theta(\sqrt{k})$, we obtain $\Pr_{\mathbf{v},X,Y}[\mathsf{Success}] = \Omega(1)$.

6 Open problems

Our work raises a number of intriguing open questions, such as:

- Is there a protocol for general channels whose communication, for agreeing on a k-bit random string with constant probability, approaches $s^*(Y;X)k$? Here $s^*(Y;X)$ is the channel parameter defined in (1.1).
- We considered protocols where the shared random string was a function $g_A(X)$ of Alice's input X. What can we achieved by a general multi-round communication protocol, where the shared random string can depend on both X and Y? Can we do better than the lower bounds we established, or do the lower bounds continue to hold in this (seemingly) more powerful model?
- The setup for $\mathsf{BEC}(\varepsilon)$ is not symmetric between Alice and Bob. What can be done if Alice and Bob switch roles, and the shared randomness should be a function of Y? What are the possible trade-offs in the symmetric setup where X and Y are the independent outputs of $\mathsf{BEC}(\varepsilon)$ on a common random string $Z \in \{0,1\}^N$?

Acknowledgments

We thank the Simons Institute for the Theory of Computing, Berkeley, where most of this work was done. We are grateful to T.S. Jayram and Chandra Nair for several illuminating discussions related to this work.

References

[AC98] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography - part II: CR capacity. *IEEE Transactions on Information Theory*, 44(1):225–240, 1998.

[AGKN13a] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. On hypercontractivity and the mutual information between boolean functions. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 13–19, 2013.

[AGKN13b] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover. *CoRR*, abs/1304.6133, 2013.

[AGKN14] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. On hypercontractivity and a data processing inequality. In *2014 IEEE International Symposium on Information Theory*, pages 3022–3026, 2014.

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *IEEE Transactions on Information Theory*, 57(10):6351–6355, 2011.

[CGMS15] Clément Louis Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 257–262, 2015.

[CK14] Thomas A. Courtade and Gowtham R. Kumar. Which Boolean functions maximize mutual information on noisy inputs? *IEEE Transactions on Information Theory*, 60(8):4515–4525, 2014.

[Coo09] John D Cook. *Upper and lower bounds for the normal distribution function*, 2009. http://www.johndcook.com/normalbounds.pdf.

[CT14] Venkat Chandar and Aslan Tchamkerten. Most informative quantization functions. In *Proceedings of ITA Workshop*, 2014. Available at http://perso.telecom-paristech.fr/ tchamker/CTAT.pdf.

[EC98] Elza Erkip and Thomas M. Cover. The efficiency of investment information. *IEEE Transactions on Information Theory*, 44(3):1026–1040, 1998.

[Erk96] Elza Erkip. *The efficiency of information in investment*. PhD thesis, Stanford University, 1996.

[Fel71] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley and Sons, 1971.

[GK72] Péter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):119–162, 1972.

[KOW15] Guy Kindler, Ryan O'Donnell, and David Witmer. Remarks on the most informative function conjecture at fixed mean. *CoRR*, abs/1506.03167, 2015.

[MM12] Konstantin Makarychev and Yury Makarychev. Chain independence and common information. *IEEE Transactions on Information Theory*, 58(8):5279–5286, 2012.

[MO05] Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Struct. Algorithms*, 26(4):418–436, 2005.

[MOR$^+$06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse bonamibeckner inequality. *Israel J. Math.*, 154:299336, 2006.

[NW15] Chandra Nair and Yannan Wang. Evaluating hypercontractivity parameters using information measures. Manuscript, 2015.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[OSW15] Or Ordentlich, Ofer Shayevitz, and Omri Weinstein. Dictatorship is the most informative balanced function at the extremes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:84, 2015.

[Rom00] Andrei Romashchenko. Pairs of words with nonmaterializable mutual information. *Problems of Information Transmission*, 36(1):1–18, 2000.

[Sam15a] Alex Samorodnitsky. The ”most informative boolean function” conjecture holds for high noise. *CoRR*, abs/1510.08656, 2015.

[Sam15b] Alex Samorodnitsky. On the entropy of a noisy function. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:129, 2015.

[WZ73] Aaron D. Wyner and Jacob Ziv. A theorem on the entropy of certain binary sequences and applications: Part 1. *IEEE Transactions on Information Theory*, 19(6):769–772, 1973.

[Yan07] Ke Yang. On the (im)possibility of non-interactive correlation distillation. *Theor. Comput. Sci.*, 382(2):157–166, 2007.

[ZC11] Lei Zhao and Yeow-Khiang Chia. The efficiency of common randomness generation. In *Proceedings of 49th Annual Allerton Conference on Communication, Control, and Computing*, pages 944–950, 2011.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de