Round Complexity of Common Randomness Generation: The Amortized Setting

Noah Golowich* Madhu Sudan†

Abstract

In this work we study the effect of rounds of interaction on the common randomness generation (CRG) problem. In the CRG problem, two parties, Alice and Bob, receive samples X_i and Y_i, respectively, where (X_i, Y_i) are drawn jointly from a *source distribution* μ. The two parties wish to agree on a common random key consisting of many bits of randomness, by exchanging messages that depend on each party's respective input and the previous messages. In this work we study the *amortized* version of the problem, i.e., the number of bits of communication needed per random bit output by Alice and Bob, in the limit as the number of bits generated tends to infinity. The amortized version of the CRG problem has been extensively studied in the information theory literature, though very little was known about the effect of interaction on this problem. Recently Bafna et al. (SODA 2019) considered the *non-amortized* version of the problem (so here the goal of the interaction is to generate a fixed number of random bits): they gave a family of sources $\mu_{r,n}$ parameterized by $r, n \in \mathbb{N}$, such that with $r+2$ rounds of communication one can generate n bits of common randomness with this source with $O(r \log n)$ communication, whereas with roughly $r/2$ rounds the communication complexity is $\Omega(n/\operatorname{poly}\log n)$. Note in particular that their source is designed with the target number of bits in mind and hence the result does not apply to the amortized setting.

In this work we strengthen the work of Bafna et al. in two ways: First we show that the results extend to the classical amortized setting. We also reduce the gap between the round complexity in the upper and lower bounds to an additive constant. Specifically we show that for every pair $r, n \in \mathbb{N}$ the (amortized) communication complexity to generate $\Omega(n)$ bits of common randomness from the source $\mu_{r,n}$ using $r+2$ rounds of communication is $O(r \log n)$ whereas the amortized communication required to generate the same amount of randomness from r rounds is $\Omega(\sqrt{n})$. Our techniques exploit known connections between information complexity and CRG, and the main novelty is our ability to analyze the information complexity of protocols getting inputs from the source $\mu_{r,n}$.

1 Introduction

In this paper we study the problem of *common randomness generation* (CRG) and the companion problem of *secret key generation* (SKG). In each of these problems, there are two parties Alice and Bob, who are given several samples of correlated randomness: Alice is given random variables $X_1, X_2, \ldots$, and Bob is given random variables $Y_1, Y_2, \ldots$, where the pairs (X_i, Y_i) are distributed i.i.d. according to some distribution μ. In the CRG problem, the goal of Alice and Bob is to agree, with high probability, on some shared key K of high entropy by communicating as little as possible. In the SKG problem, they have the additional secrecy requirement that an eavesdropper Eve who observes their transcript of communication cannot determine much information on K.

The problems of CRG and SKG were introduced independently by Maurer [Mau91, Mau92, Mau93] and by Ahlswede and Csiszár [AC93, AC98]. An important motivation for their work was from cryptography, where the posession of a shared secret key allows parties to securely transmit information using a private-key cryptosystem. Rather than generating private keys based on computational hardness assumptions, as in [DH76, RSA78], these works sug-

*Massachusetts Institute of Technology, EECS, nzg@mit.edu. Currently supported by an MIT Akamai Fellowship, a Fannie & John Hertz Foundation Fellowship, and an NSF Graduate Fellowship. This work was performed while the author was a student at Harvard University.

†Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, MA 02138, USA. madhu@cs.harvard.edu. Work supported in part by a Simons Investigator Award and NSF Award CCF 1715187.

gested the study of secret key generation from an information-theoretic viewpoint, under information-theoretic assumptions such as access to a correlated source. Subsequently techniques similar to those developed in [Mau93, AC93], such as privacy amplification, have been used in work on quantum key agreement [BBB+92, HAD+95]. Shared common randomness, and the generation thereof, has also found additional applications in identification capacity [AD89b, AD89a], communication complexity [CGMS17, GKS16, GS17, BGI14], locality-sensitive hashing, [GJ18] and coding theory [BBT60, CN91].

The initial introduction of CRG and SKG by Maurer, Ahlswede, and Csiszár was in the *amortized setting*, which has since been studied in many works (such as [CN00, CN04, ZC11, Tya13, LCV15, Liu16, LCV17, Ye05, GA10a, GA10b]). In this setting, given a source of correlation μ, the goal is to characterize the "achievability region", i.e., those pairs (C, L) of non-negative real numbers, such that if Alice and Bob receive N i.i.d. copies of the inputs $(X, Y) \sim \mu$, by communicating roughly $C \cdot N$ bits, they can generate nearly $L \cdot N$ bits of common randomness (or secret key) with probability approaching 1 as $N \to \infty$; a formal definition is presented in Definitions 2.1 and 2.2.

In the theoretical computer science community the *non-amortized setting* of CRG has also been extensively studied. In this setting, Alice and Bob still receive some number N of i.i.d. samples from the source μ, but the communication and key length do not have to grow linearly with the number of samples, and the probability of agreeing on a key need not approach 1. This problem was first studied in its zero-communication variant, where it is also known as *non-interactive correlation distillation*, and in the setting where Alice and Bob wish only to agree on a single bit, by Gacs and Körner [GK73] and Witsenhausen [Wit75], as well as later works [MO05, MOR+06, Yan07]. Bogdanov and Mossel [BM11] and Chan et al. [CMN14] study the version where Alice and Bob wish to agree on many bits, again in the zero-communication setting. Finally, several more recent works [CGMS17, GR16, GJ18] have studied the non-amortized version of CRG where communication is allowed. These latter works generally study relatively simple sources, such as the bivariate Gaussian source (BGS) and the binary symmetric source (BSS).

1.1 Overview of main results: does interaction help?

Despite the large amount of work on CRG and SKG in the last several decades, until recently, very little was known about the role of interaction in these problems. While initial work in the area [AC93, AC98] studied only 1-round and 2-round protocols, recent works [LCV15, Liu16, LCV17] have generalized those initial results to multi-round protocols; however, until our work, it was not known in the amortized setting if increasing the number of rounds of some r-round protocol can actually allow the parties to communicate less (and generate random keys of the same length).

This question of whether Alice and Bob can reduce the communication cost of their protocol at the expense of increasing the number of rounds is central to our work. Curiously, for the amortized setting, the answer to this question is negative in several cases: for instance, when (X, Y) is distributed according to the binary symmetric source (BSS) or the bivariate Gaussian source (BGS), Liu et al. [LCV17] and Tyagi [Tya13] showed that increasing the number of rounds does not help to reduce communication cost. In terms of separation results, Tyagi [Tya13] presented a source on a ternary alphabet for which a 1-round protocol has smaller communication cost than any 2-round protocol by a constant factor, and this is the only known round-based separation in the amortized setting.

Orlitsky [Orl90, Orl91] studied a slightly different version of CRG in which the key K is required to be equal to Alice's input X; thus the problem becomes that of Bob learning Alice's input. Orlitsky showed (in the non-amortized case) that 2-round protocols can require exponentially less communication than 1-round protocols. However, for any $r > 2$, he showed that r-round protocols can save on communication cost over 2-round protocols by at most a factor of 4. This version of the problem was also studied in the amortized case by Ma and Ishwar [MI08], who showed that interaction does not help at all; in fact, the 1-round protocol that achieves minimum communication cost is simply given by Slepian-Wolf coding [CT12].

The most relevant work is that of Bafna et al. [BGGS19], who showed the following in the non-amortized setting (see [BGGS19, Theorems 1.1 & 1.2]): for any fixed r, there is sufficiently large n such that for some source $\mu = \mu_{r,n}$, we have:

(1) Alice and Bob can generate secret keys of length n with $r + 2$ rounds of communication and $O(\log n)$ communication cost;

(2) When restricted to $\lfloor (r+1)/2 \rfloor$ rounds, any protocol which generates common random keys of length n must have communication cost $n/\log^{\omega(1)} n$.

Notice that the above result is not tight in the dependence on the number of rounds; a tight result (up to polylogarithmic factors) would state that any $(r+1)$-round protocol must have communication cost $n/\log^{\omega(1)} n$. More significantly, the result does not establish any separation in the amortized setting, which is the main target of this paper. (In particular, the lower bound for $\lfloor (r+1)/2 \rfloor$-round communication in item (2) above does not rule out the possibility that if Alice and Bob receive N i.i.d. copies of the source μ, there is some $\lfloor (r+1)/2 \rfloor$-round protocol for generating $O(N \cdot n)$ bits of randomness but with communication a lot less than $O(N \cdot n)$, i.e., as small as $O(N \log n)$, or even smaller.)

Our results Our main result is an extension of the above results of Bafna et al. [BGGS19] to the amortized setting. Along the way we also get a nearly tight dependence on the number of rounds (losing a quadratic factor in communication cost and a single additional round of communication). In particular, we show:

- For the source $\mu = \mu_{r,n}$ mentioned above, any protocol with at most r rounds and which generates common random keys of length n must have communication cost at least $\sqrt{n}/\log^{\omega(1)} n$. (See Theorem 2.2 for a formal statement.)
- Moreover, an identical rounds-communication tradeoff holds for the amortized case. (See Theorem 2.3.)

We emphasize that the second result above gives the first rounds-communication tradeoff for the amortized case (apart from the constant-factor separation between 1-round and 2-round protocols given by Tyagi [Tya13]).

Technical Challenge At a very high level the source in [BGGS19] is built around the concept of "pointer-chasing problems" that are well-known to lead to separations in round-complexity [NW93, DGS84, PS82]. The main contribution in their work is to show how the hardness of pointer chasing (or a variation they consider) translates to the hardness of generating common randomness in their source.

Getting an amortized lower bound turns out to be significantly more challenging. For one thing we can no longer build a source that is crafted around a targeted length of the common random string. Indeed this ability allows Bafna et al. [BGGS19] to focus on the case where the two players get a single copy of the randomness $(X, Y) \sim \mu$, and the core of their negative result is showing that $r/2$ rounds of communication are insufficient to generate any non-trivial randomness from this single copy ("non-trivial" meaning more than the number of bits communicated). In the amortized case such results are not possible: if there is a protocol with small communication and many rounds getting some amount of randomness, then we can simulate the protocol with large communication in two rounds, and then (here using the ability to amortize) we can scale back the communuication and generate proportionately less, but non-trivial amounts of randomness. Thus no matter how small the amortized communication budget is, it is always possible to get some non-trivial amounts of randomness. So our lower bounds really need to address a "direct product" version of the pointer chasing question.

Indeed, the idea of our proof is to "reduce to the non-amortized case" by using similar types of techniques that have been applied to show direct sum and direct product results for the communication complexity of functions [CSWY01, JRS03, HJMR07, BBCR13, JPY12, BR11, BRWY13]. However, the task of CRG is "more flexible" than that of computing a function as there is no prescribed output for given inputs, so implementing this reduction is non-trivial. Roughly, our results have to analyze notions such as the internal and external information cost of *all* bounded round protocols (and show that these are close) whereas most of the previous use in communication complexity lower bounds only needed to work with protocols that computed a specific function. We go into further details on this in Section 2.6 after we get more specific about the sources we consider and the kind of results we seek.

Organization of this paper In Section 2 we formally introduce the problems of CRG and SKG (in both the non-amortized and amortized settings) and state our main results. Section 3 presents the proof of our main results in the non-amortized setting, and Section 4 presents the proof of our main results in the amortized setting. Section 5 collects some open questions motivated by our work.

2 Background and Overview of Main Results

2.1 Notation We first describe some of the basic notational conventions we use throughout the paper.

We use capital script font, such as $\mathcal{S}, \mathcal{X}, \mathcal{Y}$, to denote sets, and capital letters, such as X, Y, Z, to denote random variables. We typically use the letters μ, ν, D to denote distributions. $\mathcal{S}_n$ denotes the set of all permutations on $[n]$.

Basic probability If $\mathcal{E} \subset \mathcal{X}$ is some event, then we will write $\mathbf{1}[X \in \mathcal{E}]$ to denote the random variable that is 1 if $X \in \mathcal{E}$, and 0 otherwise. We will slightly abuse notation, e.g., if $(X, Y) \sim \nu$ then $\mathbf{1}[X = Y]$ is 1 when $X = Y$ and 0 otherwise. If $f : \mathcal{X} \to \mathbb{R}$, then $\mathbb{E}_\mu[f(X)]$ denotes the expectation of $f(X)$ when X is distributed according to μ. For $\mathcal{E} \subset \mathcal{X}$, $\mathbb{P}_\mu[\mathcal{E}] := \mathbb{E}_\mu[\mathbf{1}[X \in \mathcal{E}]]$ is the probability that $X \in \mathcal{E}$ when $X \sim \mu$. We will omit the subscript μ if the distribution is obvious. This notation extends naturally to conditional expectations.

Total variation distance & KL divergence For random variables X, Y distributed according to μ, ν, respectively, on a finite set $\mathcal{X}$, $\Delta(\mu, \nu) := \frac{1}{2}\sum_{x \in \mathcal{X}} |\mathbb{P}_\mu[X = x] - \mathbb{P}_\nu[Y = x]|$ denotes the *total variational distance* between X and Y. For distributions μ and ν supported on a set $\mathcal{X}$, the *KL divergence* between μ, ν, denoted $\mathrm{KL}(\mu||\nu)$, is given by, for $X \sim \mu, Y \sim \nu$, $\mathrm{KL}(\mu||\nu) := \sum_{x \in \mathcal{X}} \mathbb{P}[X = x] \cdot \log\left(\frac{\mathbb{P}[X=x]}{\mathbb{P}[Y=x]}\right)$. We will often abuse notation when denoting KL divergences or total variation distances: for $X \sim \mu, Y \sim \nu$ supported on a set $\mathcal{X}$, we will write $\Delta(X, Y) = \Delta(\mu, \nu)$ and $\mathrm{KL}(X||Y) = \mathrm{KL}(\mu||\nu)$.

Information theory If $X \sim \mu$, then the *entropy* of X is given by $H_\mu(X) = H(X) = \mathbb{E}_{x \sim \mu}[\log(1/\mathbb{P}_\mu[X = x])]$. Now suppose (X, Y) are random variables with $X \in \mathcal{X}, Y \in \mathcal{Y}$ jointly distributed according to some distribution ν. Letting X_y denote the random variable distributed as X, conditioned on $Y = y$, then $H(X|Y = y) := H(X_y)$. Then the *conditional entropy* $H_\nu(X|Y) = H(X|Y)$ is given by $H(X|Y) =: \mathbb{E}_{y \sim \nu}[H(X|Y = y)]$. The *mutual information* is given by $I_\nu(X; Y) = I(X; Y) := H(X) - H(X|Y)$; it is well-known that $I(X; Y) = H(Y) - H(Y|X)$. If (X, Y, Z) are jointly distributed according to some distribution, then the *conditional mutual information* $I(X; Y|Z)$ is given by $I(X; Y|Z) := H(X|Z) - H(X|Y, Z)$.

Multiple random variables For random variables $(X, Y) \sim \mu$ distributed jointly, we will often use $XY \in \mathcal{X} \times \mathcal{Y}$ to denote the pair. The *marginals* $X \sim \mu_X, Y \sim \mu_Y$ are the distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively, given by $\mathbb{P}_{X \sim \mu_X}[X = x] := \mathbb{P}_{XY \sim \mu}[X = x]$, and similarly for μ_Y. Then $X \otimes Y \in \mathcal{X} \times \mathcal{Y}$ denotes the random variable distributed according to the product of the marginals $\mu_X \otimes \mu_Y$. For a sequence of random variables $X_1, X_2, \ldots, X_i, \ldots$, for any $j \geq 1$, we let X^j denote the tuple $(X_1, \ldots, X_j)$, and for $1 \leq j \leq j'$, let $X_j^{j'}$ denote the tuple $(X_j, X_{j+1}, \ldots, X_{j'})$. Two common usages of this notation are as follows: (1) for $N \in \mathbb{N}$, and a distribution $Z \sim \mu$, the random variable distributed according to N i.i.d. copies of μ is denoted as $Z^N = (Z_1, \ldots, Z_N) \sim \mu^{\otimes N}$; (2) if $\Pi_1, \ldots, \Pi_t$ denote the first t messages in a communication protocol (see Section 2.2), then $\Pi^t = (\Pi_1, \Pi_2, \ldots, \Pi_t)$.

2.2 Communication protocols

We follow the standard setup of interactive communication protocols [Yao79], and mostly follow the notational conventions of [BBCR13, BR11]. There are finite sets $\mathcal{X}, \mathcal{Y}$, and parties Alice and Bob, who receive inputs $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, respectively. Depending on the setting, Alice and Bob may additionally have access to private coins $R_{\mathtt{A}}, R_{\mathtt{B}}$, respectively, and public coins $R_{\mathtt{Pub}}$. Formally, $R_{\mathtt{A}}, R_{\mathtt{B}}, R_{\mathtt{Pub}}$ may be interpreted as infinite strings of independently and uniformly distributed random bits.

An *interactive r-round protocol* Π consists of a sequence of r *messages*, $\Pi_1, \ldots, \Pi_r \in \{0,1\}^*$ that Alice and Bob alternatively send to each other, with Alice sending the first message Π_1. The messages $\Pi_1, \ldots, \Pi_r$ are also referred to as the *rounds* of the protocol, and each message is a deterministic function of the previous messages, one party's input, and any randomness (public and/or private) available to that party. For $1 \leq t \leq r$ with t odd, we will write Alice's message Π_t as $\Pi_t = \Pi_t(X, R_{\mathtt{A}}, R_{\mathtt{Pub}}, \Pi^{t-1})$ if the protocol can use public and private coins (with obvious modifications if public and/or private coins are not available), and for t even, Bob's message Π_t as $\Pi_t = \Pi_t(Y, R_{\mathtt{B}}, R_{\mathtt{Pub}}, \Pi^{t-1})$.[1] The *communication cost* of Π, denoted by $\mathrm{CC}(\Pi)$, is the maximum of $\sum_{t=1}^r |\Pi_t|$, taken over all inputs $X \in \mathcal{X}, Y \in \mathcal{Y}$, and all settings of the random coins $R_{\mathtt{A}}, R_{\mathtt{B}}, R_{\mathtt{Pub}}$ (if applicable). The tuple consisting of all the messages, i.e., $\Pi^r = (\Pi_1, \ldots, \Pi_r)$, is referred to as the *transcript* of the protocol Π.

2.3 Rate regions for amortized CRG & SKG

Recall that in amortized CRG, Alice and Bob receive some large number N of copies (X, Y) from the

[1] It is required that for each t and each instantiation of Π^{t-1}, the set of possible values of Π_t (over all possible instantiaions of $X, Y, R_{\mathtt{A}}, R_{\mathtt{B}}, R_{\mathtt{Pub}}$) must be prefix-free. This technical detail, which is introduced so that each party knows when to "start speaking" when the other finishes, will not be important for us.

source, are allowed to communicate some number of bits that grows linearly with N, and must agree upon a key whose entropy grows linearly with N with probability tending to 1 as $N \to \infty$. The word "amortized" refers to the fact that the communication and key entropy both grow linearly with N. The parties may use private but not public coins (as with access to public randomness, there would be no need to generate a shared random string). Definition 2.1 below follows the exposition of Liu et al. [LCV17].

DEFINITION 2.1. (AMORTIZED CRG) *The tuple (C, L) is r-achievable for CRG for a source distribution $(X, Y) \sim \nu$ if for every $N \in \mathbb{N}$, there is some ϵ_N with $\epsilon_N \to 0$ as $N \to \infty$, a key set $\mathcal{K}_N$, and a private-coin protocol $\Pi = \Pi(N)$ that takes as input $(X^N, Y^N) \sim \nu^{\otimes N}$, such that if $\Pi(N)_t \in \{0,1\}^*$ denotes the message sent in the t-th round of $\Pi(N)$, $1 \le t \le r$, and $K_A = K_A(N), K_B = K_B(N) \in \mathcal{K}_N$ denote the output keys of Alice and Bob for the protocol $\Pi(N)$, then:*

1. $\limsup_{N\to\infty} \frac{1}{N} \cdot \mathrm{CC}(\Pi(N)) \le C$.
2. $\liminf_{N\to\infty} \frac{1}{N} \log |\mathcal{K}_N| \ge L$.
3. *Letting K_N be the random variable that is uniformly distributed on $\mathcal{K}_N$, then*

$$\Delta((K_A(N)K_B(N)), (K_N K_N)) \le \epsilon_N.$$

In particular, there exists a coupling of $K_A(N)K_B(N)$ with $K_N K_N$ such that $\mathbb{P}[K_A(N) = K_B(N) = K_N] \ge 1 - \epsilon_N \to 1$ as $N \to \infty$. (To be clear, $K_N K_N$ denotes the tuple (K_N, K_N) which is distributed uniformly on the set $\{(k,k) : k \in \mathcal{K}_N\}$.)

We denote the subset of pairs $(C, L) \subset \mathbb{R}^2_{\ge 0}$ that are r-achievable from the source $(X, Y) \sim \nu$ by $\mathcal{T}_r(X, Y)$; this set $\mathcal{T}_r(X, Y)$ is known as the achievable rate region for r-round CRG (or simply rate region, with r and the task of CRG implicit) for the source μ.

To interpret Definition 2.1, notice that C denotes the communication of the protocols $\Pi = \Pi(N)$, whereas L (approximately) gives the entropy of the key produced.

Corresponding to Definition 2.1 for CRG we have the following Definition 2.2 for SKG in the amortized setting:

DEFINITION 2.2. (AMORTIZED SKG) *The tuple (C, L) is r-achievable for SKG for a distribution ν if there is some choice of a sequence $\epsilon_N \to 0$ such that the following holds: for each $N \in \mathbb{N}$ there is some choice of private coin protocol[2] $\Pi = \Pi(N)$ such that, first, items 1 and 2 of Definition 2.1 are satisfied for these $\epsilon_N, \Pi(N), N$, and, second,*

(2.1)
$$\Delta(K_A(N)K_B(N)\Pi(N)^r, K_N K_N \otimes \Pi(N)^r) \le \epsilon_N.$$

As in Definition 2.1, K_N denotes the random variable that is uniform on on $\mathcal{K}_N$; notice that (2.1) above implies item 3 of Definition 2.1.

We denote the set of pairs (C, L) that are r-achievable for SKG from ν by $\mathcal{S}_r(X, Y)$.

It is clear from the definition that r-achievability for SKG is a stronger requirement than r-achievability for CRG; that is, for every source $(X, Y) \sim \nu$, we have $\mathcal{S}_r(X, Y) \subset \mathcal{T}_r(X, Y)$. It is also well-known [LCV17, Han03] that both $\mathcal{T}_r(X, Y)$ and $\mathcal{S}_r(X, Y)$ are closed subsets of $\mathbb{R}^2$.

2.4 Non-Amortized Setting The non-amortized setting is similar to the amortized setting, in that Alice and Bob receive arbitrarily many i.i.d. samples of $(X, Y) \sim \mu$, except the entropy of their key and their communication no longer grow linearly with the number of samples. Rather, the keys lie in some fixed set $\mathcal{K}$, and the goal is to use as little communication (and rounds) as possible to generate a single key uniformly distributed in $\mathcal{K}$. Moreover, whereas the agreement probability $1 - \epsilon_N$ in the amortized case was assumed to approach 1 asymptotically, in the non-amortized case, it is often of interest to study settings in which the parties may disagree with some probability that is bounded away from 0. In fact, this probability of disagreement may be arbitrarily close to 1. The non-amortized setting has recently received much attention among the theoretical computer science community [BM11, CGMS17, GR16, GJ18, BGGS19], where it is also known as the *agreement distillation problem*.

In the below definition we assume that $(X, Y) \sim \nu$ and ν is supported on a set $\mathcal{X} \times \mathcal{Y}$.

DEFINITION 2.3. (NON-AMORTIZED CRG) *For $r, C \in \mathbb{N}$, and $L, \epsilon \in \mathbb{R}_{\ge 0}$, we say that the tuple (C, L, ϵ) is r-achievable from the source ν (for CRG) if there is some $N \in \mathbb{N}$ and an r-round protocol Π with private randomness that takes as input $(X^N, Y^N) \sim \nu^{\otimes N}$, such that at the end of Π, Alice and Bob output keys $K_A, K_B \in \mathcal{K}$ given*

[2] As for CRG, the protocol Π cannot use public coins.

by deterministic functions $K_A = K_A(X^N, R_A, \Pi^r)$, $K_B = K_B(Y^N, R_B, \Pi^r)$, *such that:*

1. $\mathrm{CC}(\Pi) \le C$.
2. $|\mathcal{K}| \ge 2^L$.
3. *There is a random variable* K *uniformly distributed on* $\mathcal{K}$ *such that* $\mathbb{P}_\nu[K = K_A = K_B] \ge 1 - \epsilon$.

As in the amortized case, for tuples (C, L, ϵ), observe that C denotes communication and L denotes entropy.

Definition 2.3 differs slightly from the definition of achievable rates for non-amortized CRG in [BM11, CGMS17, GR16, GJ18, BGGS19], which do not limit the size of the key space $\mathcal{K}$, but rather require a lower bound on the min-entropy of each of K_A, K_B. In the full version of this paper [GS19, Appendix A], we present this latter definition and show that it is essentially equivalent to Definition 2.3.

As in the amortized setting, in the non-amortized setting secret key generation is the same as common randomness generation except the key is additionally required to be "almost independent" from the transcript of the protocol:

DEFINITION 2.4. (NON-AMORTIZED SKG) *For* $r, C \in \mathbb{N}$ *and* $L \in \mathbb{R}_{\ge 0}$, $\epsilon, \delta \in [0, 1)$, *we say that the tuple* (C, L, ϵ, δ) *is* r*-achievable from the source* ν *(for SKG) if the tuple* (C, L, ϵ) *is* r*-achievable for CRG from the source* ν, *and if there exists a protocol* $\Pi = (\Pi^1, \ldots, \Pi^r)$ *achieving the tuple such that*

$$I(\Pi^r; K_A K_B) \le \delta. \tag{2.2}$$

Notice that condition (2.2) is quite strong: it implies, for instance, that $\Delta(\Pi^r K_A K_B, \Pi^r \otimes K_A K_B) \le \sqrt{\delta/2}$, by Pinsker's inequality.

2.5 Main Results: Analogue of Pointer-Chasing Separations for CRG & SKG In this section we present our main results. We first state formally the main result of [BGGS19] discussed in Section 1.1, which establishes an exponential separation in communication cost between $\lfloor (r+1)/2 \rfloor$-round protocols and $(r+2)$-round protocols in the non-amortized setting:

THEOREM 2.1. (THMS. 1.1 & 1.2 OF [BGGS19]) *For each* $r \in \mathbb{N}, \epsilon \in [0, 1)$, *there exists* $\eta > 0$, $\beta < \infty$, $n_0 \in \mathbb{N}$ *such that for any* $n \ge n_0$ *and any* $\ell \in \mathbb{N}$, *there is a source* $\mu_{r,n,\ell}$ *such that, in the non-amortized setting:*

(1) The tuple $((r+2)\lceil \log n \rceil, \ell, 0, 0)$ *is* $(r+2)$*-achievable for SKG from* $\mu_{r,n,\ell}$ *(and thus* $((r+2)\lceil \log n \rceil, \ell, 0)$ *is* $(r+2)$*-achievable for CRG).*

(2) For any $L \in \mathbb{N}$ *and* $C \le \min\{\eta L - \beta, n/\log^\beta n\}$, *the tuple* (C, L, ϵ) *is not* $\lfloor (r+1)/2 \rfloor$*-achievable for CRG (and thus the tuple* (C, L, ϵ, δ) *is not* $\lfloor (r+1)/2 \rfloor$*-achievable for all* $\delta \ge 0$*).*

The interpretation of the parameters n, ℓ in Theorem 2.1 is described in detail in Definition 3.1 of the source $\mu_{r,n,\ell}$. We remark that the proof of item (1) of the theorem is immediate once this definition is made, and so the main content of Theorem 2.1 is in the second item (i.e., the lower bound).

To aid understanding of Theorem 2.1, fix any $r \in \mathbb{N}, \epsilon \in [0, 1)$, and consider parameters $\ell = n \to \infty$; the length of Alice's and Bob's inputs under $\mu_{r,n,n}$ are $O(n^2)$. The theorem gives that with only $O(\log n)$ communication, n bits of entropy can be generated in $r+2$ rounds, but if we have only roughly half as many rounds (i.e., $\lfloor (r+1)/2 \rfloor$ rounds) then generating n bits of entropy takes at least $n/\operatorname{poly}\log n$ communication, which is exponentially larger than $\log n$. It follows that for some r' with $\lfloor (r+1)/2 \rfloor \le r' < r+2$, the ratio in communication cost between the best r'-round protocol and the best $(r'+1)$-round protocol is at least $n^{1/(1+\lceil (r+1)/2 \rceil)}/\log^{\omega(1)} n$. Our first main result improves this ratio to $n^{1/4}/\log^{\omega(1)} n$ and moreover shows that such an r' lies in $\{r, r+1\}$:

THEOREM 2.2. (TIGHTER ROUND; NON-AMORTIZED) *For each* $r \in \mathbb{N}, \epsilon \in [0, 1)$, *there exists* $\eta > 0$, $\beta < \infty$, $n_0 \in \mathbb{N}$ *such that for any* $n \ge n_0$ *and any* $\ell \in \mathbb{N}$, *the source* $\mu_{r,n,\ell}$ *of Theorem 2.1 satisfies:*

(1) The tuple $((r+2)\lceil \log n \rceil, \ell, 0, 0)$ *is* $(r+2)$*-achievable for SKG from* $\mu_{r,n,\ell}$ *(and thus* $((r+2)\lceil \log n \rceil, \ell, 0)$ *is* r*-achievable for CRG).*

(2) For any $L \in \mathbb{N}$, $C \le \min\{\eta L - \beta, \sqrt{n}/\log^\beta n\}$, *the tuple* (C, L, ϵ) *is not* r*-achievable for CRG from* $\mu_{r,n,\ell}$ *(and thus for any* $\delta \ge 0$, *the tuple* (C, L, ϵ, δ) *is not* r*-achievable for SKG).*

Our second main result provides an exact analogue of Theorems 2.1 and 2.2 for the amortized setting:

THEOREM 2.3. (AMORTIZED SETTING) *For each* $r \in \mathbb{N}, \gamma \in (0, 1)$, *there is a constant* $c_0 > 0$ *such that for* $n \ge c_0$, *the source* $\mu_{r,n,\ell}$ *of Theorem 2.1 satisfies:*

(1) The tuple $((r+2)\lceil \log n \rceil, \ell)$ *is* $(r+2)$*-achievable for SKG (and thus CRG) from* $\mu_{r,n,\ell}$.

(2) *Set $\ell = n$. For any $C, L \in \mathbb{R}$ with $C \le n/\log^{c_0} n$ and $L > \gamma\ell = \gamma n$, the tuple (C, L) is not $\lfloor (r+1)/2 \rfloor$-achievable for CRG (and thus for SKG) from $\mu_{r,n,n}$ for $n \ge c_0$.*

(3) *Again set $\ell = n$. For any $C, L \in \mathbb{R}$ with $C \le \sqrt{n}/\log^{c_0} n$ and $L > \gamma n$, the tuple (C, L) is not r-achievable for CRG (and thus for SKG) from $\mu_{r,n,n}$ for $n \ge c_0$.*

Notice that parts (2) and (3) of Theorem 2.3 only provide a lower bound on the communication rate C for protocols when the entropy rate L is at least a constant factor times n. The problem of determining such a result for L that grow sublinearly with n, or even those L that do not grow at all (such as $L = o_n(1)$) remains open. Such a problem boils down to showing a rounds-communication tradeoff for the r-round *common random bits per interaction bit* (CBIB) of the source $\mu_{r,n,\ell}$, or equivalently, for the r-round *strong data processing constant* (SDPC) [LCV17]; see Problem 5.1. As we discuss in Section 4.4, this problem seems to be quite difficult as a proof of it would immediately imply Theorem 2.3.

2.6 Discussion and overview of proof of Theorems 2.2 & 2.3

The source $\mu_{r,n,\ell}$ referred to in Theorems 2.1, 2.2 and 2.3 is a variant of the well-known *pointer chasing distribution* from communication complexity [NW93, DGS84, PS82]. This distribution was introduced to show a similar type of rounds/communication tradeoff as in the above theorems, except for the task of *computing functions* rather than generating a shared string.

Alice's and Bob's inputs from $\mu_{r,n,\ell}$ are given as follows: for an integer n and odd r, Alice receives permutations indexed by odd integers $\Sigma_1, \Sigma_3, \ldots, \Sigma_r : [n] \to [n]$, and Bob receives permutations indexed by even integers $\Sigma_2, \Sigma_4, \ldots, \Sigma_{r-1} : [n] \to [n]$, as well as an integer $I_0 \in [n]$. Let $J_0 = \Sigma_r(\Sigma_{r-1}(\cdots \Sigma_1(I_0))) \in [n]$. Alice and Bob also receive strings $A_1, \ldots, A_n \in \{0,1\}^\ell$ and $B_1, \ldots, B_n \in \{0,1\}^\ell$, respectively, which are distributed uniformly at random conditioned on $A_{J_0} = B_{J_0}$. If Alice and Bob have $r+2$ rounds, then the following protocol generates secret keys distributed uniformly on $\{0,1\}^\ell$: Alice sends Bob I_0, who responds with $\Sigma_1(I_0)$, Alice responds with $\Sigma_2(\Sigma_1(I_0))$, and so on, until both parties possess J_0, at which point they can output $A_{J_0} = B_{J_0}$.

To prove that Alice and Bob cannot generate shared common random strings with high entropy and communication $n/\log^{\omega(1)} n$ (item (2) of Theorem 2.1), the following approach was used: Bafna et al. [BGGS19] first reduced the problem to showing that Alice and Bob cannot succeed with high probability on a distributional version of the following communication problem: Alice receives permutations $\Sigma_1, \Sigma_3, \ldots, \Sigma_r : [n] \to [n]$, an Bob receives permutations $\Sigma_2, \Sigma_4, \ldots, \Sigma_{r-1} : [n] \to [n]$ and indices $I_0, J_0 \in [n]$. Their task is to determine if $\Sigma_r(\Sigma_{r-1} \cdots \Sigma_1(I_0)) = J_0$. This problem, called *pointer verification*, has a protocol with $(r+5)/2$ rounds and communication $O(\log n)$, given by Alice and Bob chasing the pointers forwards and backwards simultaneously. Bafna et al. [BGGS19] showed however that there is no protocol with $(r+3)/2$ rounds and communication $n/\log^{\omega(1)} n$, and this led to item (2) of Theorem 2.1. We are able to prove Theorem 2.2 by employing a reduction from the CRG/SKG problem to the pointer verification problem with indices in $[n^2]$ (as opposed to in $[n]$) and with $2r$ permutations (as opposed to r permutations).

The proof of Theorem 2.3 (in particular, of the lower bounds (2) and (3) in the theorem, as (1) is immediate) is somewhat more involved. The overall goal is to reduce to the non-amortized case (Theorems 2.1 and 2.2), and to do this, three main ingredients are needed. The first ingredient is a characterization of the achievable rate region $\mathcal{T}_r(X,Y)$ for CRG in terms of the *internal information cost* and *external information cost* [BBCR13] of private-coin communication protocols, which has been referred to many times in the literature (e.g., [STW19, GJ18]). This characterization shows that that if for $L \le \ell$, the pair (C, L) is r-achievable for CRG from $\mu_{r,n,\ell}$ (i.e., belongs to $\mathcal{T}_r(X,Y)$), then there is an r-round private-coin protocol Π with inputs $(X,Y) \sim \mu_{r,n,\ell}$ with internal information cost at most C and external information cost at least L (see Corollary 4.1).

The second ingredient of the proof is a result of Jain et al. [JPY12] (which is implicit in the earlier work of Braverman and Rao [BR11]) stating that for any r-round protocol $\hat{\Pi}$ with internal information cost I, there exists an r-round protocol $\hat{\Pi}'$ that simulates $\hat{\Pi}$ up to some accuracy loss ϵ and has communication cost at most $\frac{I+O(r)}{\epsilon}$ (see Theorem 4.3). Applying this to $\hat{\Pi} = \Pi$, one might hope to show that Π leads to a protocol with communication $O(C)$ and key length $\Omega(L)$ for non-amortized CRG. However, the error ϵ introduced in the information-to-communication compression result of Jain et al. [JPY12] makes this conclusion nontrivial, which necessitates the third ingredient: a delicate argument that makes use of

the specific structure of $\mu_{r,n,\ell}$ is needed to complete the reduction (see Lemmas 4.1 and 4.2).

3 Proof of Theorem 2.2; non-amortized setting

In this section we prove Theorem 2.2. To prove this theorem we need to introduce the pointer chasing source of [BGGS19], and also recall the notion of "indistinguishability" of two distributions to low-round low-communication protocols. We then state our main technical theorem (Theorem 3.1) about the indistinguishability of the pointer chasing source from an "independent source" (where Alice and Bob get inputs that are independent of each other). Section 3.1 is devoted to the proof of Theorem 3.1.

We begin by formally defining the *pointer-chasing source* $\mu_{r,n,\ell}$ that the theorem uses to achieve the rounds-communication tradeoff.

DEFINITION 3.1. ([BGGS19], DEFINITION 2.1)
For positive integers r, n and ℓ, the support of $\mu = \mu_{r,n,\ell}$ is $(\mathcal{S}_n^{\lceil r/2 \rceil} \times \{0,1\}^{n\ell}) \times ([n] \times \mathcal{S}_n^{\lfloor r/2 \rfloor} \times \{0,1\}^{n\ell})$. Denoting $X = (\Sigma_1, \Sigma_3, \ldots, \Sigma_{2\lceil r/2 \rceil - 1}, A_1, \ldots, A_n)$ and $Y = (I, \Sigma_2, \Sigma_4, \ldots, \Sigma_{2\lfloor r/2 \rfloor}, B_1, \ldots, B_n)$, a sample $(X,Y) \sim \mu$ is drawn as follows:

- *$I \in [n]$ and $\Sigma_1, \ldots, \Sigma_r \in \mathcal{S}_n$ are sampled uniformly and independently.*
- *Let $J = \Sigma_r(\Sigma_{r-1}(\cdots \Sigma_1(I) \cdots)) \in [n]$.*
- *$A_J = B_J \in \{0,1\}^\ell$ is sampled uniformly and independently of I and Σ's.*
- *For every $k \neq J$, $A_k \in \{0,1\}^\ell$ and $B_k \in \{0,1\}^\ell$ are sampled uniformly and independently.*

We use the following notational convention for samples $(X,Y) \sim \mu_{r,n,\ell}$. We write $I_0 := I$, and for $1 \leq t \leq r$, $I_t := \Sigma_t(I_{t-1})$. Similarly, we write $J_0 := J$, and for $1 \leq t \leq r$, $J_t = \Sigma_t^{-1}(J_{t-1})$. Over the distribution $\mu_{r,n,\ell}$, we thus have $I_t = J_{r-t}$ for $0 \leq t \leq r$ with probability 1.

We establish the following basic property of the pointer chasing source $\mu_{r,n,\ell}$ for future reference:

LEMMA 3.1. *When $(X,Y) \sim \mu_{r,n,\ell}$, $I(X;Y) = \ell$.*

We defer the proof of Lemma 3.1 to the full version of this article [GS19].

It is immediate from the definition of $\mu_{r,n,\ell}$ that part (1) (i.e., the upper bound) of Theorem 2.2 holds: in particular, the parties "chase the pointers", i.e., alternatively send I_t, $0 \leq t \leq r$, and finally output $A_{I_r} = B_{I_r}$ as their keys. The main content of Theorems 2.1 and 2.2 is then in part (2) (i.e., the lower bound) of each; its proof, for both theorems, proceeds via arguments about indistinguishbility of inputs to protocols, which we will now define. For $r, C \in \mathbb{R}_+$, we say that a communication protocol Π is an (r,C) *protocol* if Π has at most $\lfloor r \rfloor$ rounds and communication cost at most $\lfloor C \rfloor$.

DEFINITION 3.2. ([BGGS19], DEFINITION 3.1)
Let $0 \leq \epsilon \leq 1$. Two distributions μ_1, μ_2 on pairs (X,Y) are ϵ-indistinguishable to a protocol Π if the distribution of the transcript Π^r when $(X,Y) \sim \mu_1$ has total variation distance at most ϵ from the distribution of Π^r when $(X,Y) \sim \mu_2$.

Two distributions μ_1, μ_2 are (ϵ, r, C)-indistinguishable if they are ϵ-indistinguishable to every (r,C) protocol. The distributions μ_1, μ_2 are (ϵ, r, C)-distinguishable if they are not (ϵ, r, C)-indistinguishable. If Π is a protocol such that the total variation distance of the transcript between inputs $(X,Y) \sim \mu_1$ and inputs $(X,Y) \sim \mu_2$ is at least ϵ, then we say that Π distinguishes between μ_1 and μ_2 with advantage ϵ.

Proposition 3.1 reduces the problem of showing that certain tuples (C, L) are not achievable for CRG from $\mu_{r,n,\ell}$ to that of showing indistinguishability of $\mu_{r,n,\ell}$ from the product of its marginals $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$.

PROPOSITION 3.1. ([BGGS19], PROP. 3.3 & 3.4)
There are positive constants η, ξ such that the following holds. Suppose $\rho, C, L \in \mathbb{N}$ and $0 < \gamma < 1$. Suppose that $C < \eta L - 3/2 \cdot \log 1/\gamma - \xi$ and that the tuple $(C, L, 1-\gamma)$ is ρ-achievable for CRG from the source $\mu_{r,n,\ell}$. Then there is some $N \in \mathbb{N}$ such that $\mu_{r,n,N\ell}$ and $(\mu_{r,n,N\ell})_X \otimes (\mu_{r,n,N\ell})_Y$ are $(\gamma/10, C + \xi \log 1/\gamma, \rho + 1)$-distinguishable.

Our main theorem for this section is the following indistiguishability result for $\mu = \mu_{r,n,\ell}$ versus $\mu_X \otimes \mu_Y$. In contrast to the analogous result in [BGGS19, Lemma 4.5], our result shows indistinguishability for protocols with $r+1$ rounds albeit with a smaller communication budget.

THEOREM 3.1. *For every $\epsilon > 0$ and $r \in \mathbb{N}$ there exists β, n_0 such that for every $n \geq n_0$ and ℓ, the distributions $\mu = \mu_{r,n,\ell}$ and $\mu_X \otimes \mu_Y$ are $(\epsilon, r+1, \sqrt{n}/\log^\beta n)$-indistinguishable.*

The proof of Theorem 2.2 is immediate from Proposition 3.1 and Theorem 3.1; we refer the readers

to the full version of this paper [GS19] for the details. To complete the proof of Theorem 2.2 it therefore suffices to prove Theorem 3.1. We do so in the following subsection.

3.1 Disjointness and Proof of Theorem 3.1

Next we work towards the proof of Theorem 3.1; the proof parallels that of a corresponding result of Bafna et al., which shows that the distributions $\mu = \mu_{r,n,\ell}$ and $\mu_X \otimes \mu_Y$ are $(\epsilon, \lfloor (r+3)/2 \rfloor, n/\log^{\beta} n)$-indistinguishable (see [BGGS19, Lemma 4.5]). A central ingredient in the proof of [BGGS19] is a "pointer verification problem" (see Definition 3.3 below) and an indistinguishability result they show for this problem (see Theorem 3.2). We use the same notion and indistiguishability result, with the main difference being that we are able to reduce a "$2r$"-round pointer verification problem to our problem whereas the proof in [BGGS19] could only reduce an r-round pointer verification problem to the same. This factor of 2 leads to the gain in this section. We remark here that a somewhat similar problem to pointer verification was considered in [GO13], though with quite different applications.[3]

The proof proceeds by eliminating each of two possible strategies Alice and Bob can use to distinguish $\mu_{r,n,\ell}$ and $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$: first, they can try to follow the chain of pointers, compute I_r, and check if $A_{I_r} = B_{I_r}$ (which is true with probability 1 under $\mu_{r,n,\ell}$ but only with probability $1/2^{\ell}$ under $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$). Computing I_r, however, with fewer than $r+2$ rounds requires communication $\Omega(n)$ by standard results for the pointer chasing problem [NW93]. Alternatively, Alice and Bob can ignore the chain of pointers and try to determine if there is *any* i such that $A_i = B_i$ (under the product distribution the probability that such an i exists is at most $n/2^{\ell} \ll 1$). As observed in [BGGS19], determining the existence of such an i is no easier than solving the set disjointness problem [Raz92], which requires communcation $\Omega(n)$. However, combining the pointer chasing and set disjointness lower bounds takes some care, and ultimately leads to the fact that we are only able to lower-bound the communication cost of r-round (as opposed to $(r+1)$-round) protocols, and get a bound of $\tilde{\Omega}(\sqrt{n})$ (as opposed to $\tilde{\Omega}(n)$). We begin by recalling the $\Omega(n)$ lower bound on the distributional communication complexity of disjointness with respect to a particular distribution:

We will use the following result that establishes the hardness of disjointness for a distribution under which the parties' sets are either disjoint or have intersection size $\sqrt{n}$.

LEMMA 3.2. *For every $\epsilon > 0$ there exists $\delta > 0$ such that for all n the following holds. Let $\mathrm{Disj}^{\mathrm{Y}}_{n,\sqrt{n}}$ (respectively, $\mathrm{Disj}^{\mathrm{N}}_{n,\sqrt{n}}$) denote the uniform distribution on pairs (U,V) with $U,V \subseteq [n]$ and $|U| = |V| = n/4$ such that $|U \cap V| = \lfloor \sqrt{n} \rfloor$ (respectively, $|U \cap V| = 0$). Then if Alice gets U and Bob gets V as inputs, $\mathrm{Disj}^{\mathrm{Y}}_{n,\sqrt{n}}$ and $\mathrm{Disj}^{\mathrm{N}}_{n,\sqrt{n}}$ are $(\epsilon, \delta\sqrt{n}, \delta\sqrt{n})$-indistinguishable to Alice and Bob.*

The proof of Lemma 3.2 uses standard reductions to the the result of [Raz92] and is deferred to the full version of the paper [GS19].

Next we state the second main ingredient in the proof of Theorem 3.1, which is a hardness result for the *pointer verification* problem introduced in [BGGS19, Definition 4.1]. The inputs to pointer verification are similar to those of the standard pointer chasing problem, except that Bob receives as input a final pointer J_0 in addition to the initial pointer I_0, and the goal is to determine if $\Sigma_r \circ \cdots \circ \Sigma_1(I_0) = J_0$:

DEFINITION 3.3. ([BGGS19], DEFINITION 4.1) *Let $r, n \in \mathbb{N}$ with r odd. Then the distributions $D^{\mathrm{Y}}_{\mathrm{PV}} = D^{\mathrm{Y}}_{\mathrm{PV}}(r,n)$ and $D^{\mathrm{N}}_{\mathrm{PV}} = D^{\mathrm{N}}_{\mathrm{PV}}(r,n)$ are both supported on $((\mathcal{S}_n^{\lceil r/2 \rceil}) \times ([n]^2 \times S_n^{\lfloor r/2 \rfloor})$, and are defined as follows:*

- *$D^{\mathrm{N}}_{\mathrm{PV}}$ is the uniform distribution on $((\mathcal{S}_n^{\lceil r/2 \rceil}) \times ([n]^2 \times S_n^{\lfloor r/2 \rfloor})$.*
- *$(X,Y) \sim D^{\mathrm{Y}}_{\mathrm{PV}}$, with $X = (\Sigma_1, \Sigma_3, \ldots, \Sigma_r), Y = (I_0, J_0, \Sigma_2, \Sigma_4, \ldots, \Sigma_{r-1})$ is sampled by letting $\Sigma_1, \Sigma_2, \ldots, \Sigma_r$ be independent and uniform over $\mathcal{S}_n$, letting $I_0 \in [n]$ be uniform and independent of the Σ_t, and setting $J_0 = \Sigma_r \circ \cdots \circ \Sigma_1(I_0)$.*

Notice that with $(r+5)/2$ rounds of communication, by communicating at most $1+(r+1)\lceil \log n \rceil$ bits, Alice and Bob can distinguish between $D^{\mathrm{Y}}_{\mathrm{PV}}(r,n)$ and

[3] In the pointer verification variant of [GO13], there are $2r$ players, each holding a single function $f : [n] \to [n]$, and the players are given a total of $r-1$ rounds to communicate. Due to our application of showing lower bounds on CRG, we need to use lower bounds on pointer verification for particular distributions in which Alice's and Bob's inputs are not independent (unlike the work of [GO13], which considered a distribution in which the parties' inputs are independent). There are additional minor differences between our work and theirs, such as the fact that we work with permutations $\Sigma : [n] \to [n]$ whereas [GO13] work with functions $f : [n] \to [n]$.

$D^{\mathrm{N}}_{\mathrm{PV}}(r,n)$ with advantage $1-1/n$. In particular, Alice sends Bob an arbitrary bit in the first round, Bob sends I_0, J_0 in the second round, Alice responds with $I_1 = \Sigma_1(I_0)$ and $J_1 = \Sigma_r^{-1}(J_0)$, Bob responds with I_2 and J_2, and so on. After $(r+3)/2$ rounds either Alice or Bob will know both $I_{(r-1)/2}$ and $J_{(r-1)/2}$, and this person sends $\mathbf{1}[\Sigma_{(r+1)/2}(I_{(r-1)/2}) = J_{(r-1)/2}]$ (which is 1 with probability 1 under $D^{\mathrm{Y}}_{\mathrm{PV}}$ and only with probability $1/n$ under $D^{\mathrm{N}}_{\mathrm{PV}}$) as the final bit.

Theorem 3.2 states that if Alice and Bob are only allowed 1 fewer round, then they must communicate exponentially more bits to distinguish $D^{\mathrm{Y}}_{\mathrm{PV}}$ and $D^{\mathrm{N}}_{\mathrm{PV}}$:

THEOREM 3.2. ([BGGS19], THEOREM 4.2) *For every $\epsilon > 0$ and odd r there exists β, n_0 such for every $n \geq n_0$, $D^{\mathrm{Y}}_{\mathrm{PV}}(r,n)$ and $D^{\mathrm{N}}_{\mathrm{PV}}(r,n)$ are $(\epsilon, (r+3)/2, n/\log^{\beta} n)$-indistinguishable.*

Using Theorem 3.2 and Lemma 3.2, we now prove Theorem 3.1. We do so using a sequence of hybrid distributions: the first distribution in this sequence is $\mu_{r,n,\ell}$ and the last distribution in this sequence is $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$. We will show that any two distributions in this sequence are nearly indistinguishable to $(r+1, n/\log^{\beta} n)$-protocols, which implies by the triangle inequality that the same holds for $\mu_{r,n,\ell}$ and $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$.

Proof. (of Theorem 3.1) We first introduce a hybrid distribution which we denote by $\hat{\mu}$ (or $\hat{\mu}_{r,n,\ell}$ when we want to emphasize dependence on r, n, ℓ); $\hat{\mu}$ is a distribution supported on $(\mathcal{S}_n^{\lceil r/2 \rceil} \times (\{0,1\}^{\ell})^n \times (\mathcal{S}_n^{\lfloor r/2 \rfloor} \times [n] \times (\{0,1\}^{\ell})^n)$. We denote a sample from $\hat{\mu}$ by (X,Y), with

$$X = (\Sigma_1, \Sigma_3, \ldots, \Sigma_{2\lceil r/2 \rceil - 1}, A_1, \ldots, A_n),$$
$$Y = (I_0, \Sigma_2, \Sigma_4, \ldots, \Sigma_{2\lfloor r/2 \rfloor}, B_1, \ldots, B_n),$$

which is distributed as follows:

- $I_0 \in [n]$ and $\Sigma_1, \ldots, \Sigma_r \in \mathcal{S}_n$ are sampled uniformly and independently. Let $I_r = \Sigma_r \circ \cdots \circ \Sigma_1(I_0)$.
- Let $P \subset [n]$ be a uniformly random subset of size $\lfloor \sqrt{n} \rfloor$, conditioned on the event that it contains I_r.
- For every $j \in P$, $A_j = B_j \in \{0,1\}^L$ is sampled uniformly and independently of I_0, Σ's, and P.
- For every $j \notin P$, $A_j, B_j \in \{0,1\}^L$ are sampled uniformly and independently (and independently of all Σ's, I_0, and P).

CLAIM 3.1. *For every $\epsilon > 0$, there exists $\delta > 0$ such that the distributions $\mu_{r,n,\ell}$ and $\hat{\mu}_{r,n,\ell}$ are $(\epsilon, \delta\sqrt{n}, \delta\sqrt{n})$-indistinguishable.*

The proof of Claim 3.1 proceeds by showing that any protocol Π with $\mathrm{CC}(\Pi) \leq C$ distinguishing $\mu = \mu_{r,n,\ell}$ and $\hat{\mu} = \hat{\mu}_{r,n,\ell}$ with advantage ϵ can be converted into a protocol Π' with $\mathrm{CC}(\Pi') \leq C$ and which distinguishes $\mathrm{Disj}^{\mathrm{Y}}_{n,\sqrt{n}}$ and $\mathrm{Disj}^{\mathrm{N}}_{n,\sqrt{n}}$, and applying Lemma 3.2. It is quite similar to that of [BGGS19, Lemma 4.5], and is deferred to the full version of the paper [GS19].

Next, notice that the two distributions $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$ and $(\hat{\mu}_{r,n,\ell})_X \otimes (\hat{\mu}_{r,n,\ell})_Y$ are identical. Thus by Claim 3.1 and the triangle inequality for total variation distance, Theorem 3.1 follows from the following claim:

CLAIM 3.2. *For every $\epsilon > 0$ and $r \in \mathbb{N}$ there exists β, n_0 such that for every $n \geq n_0$ and ℓ, the distributions $\hat{\mu} = \hat{\mu}_{r,n,\ell}$ and $\hat{\mu}_X \otimes \hat{\mu}_Y$ are $(2\epsilon, r+1, \sqrt{n}/\log^{\beta} n)$-indistinguishable.*

To prove Claim 3.2, we introduce another hybrid distribution $\mu^{\mathrm{mid}} = \mu^{\mathrm{mid}}_{r,n,\ell}$, which is the same as $\hat{\mu}_{r,n,\ell}$, except the distribution of the uniformly random subset $P \subset [n]$ with $|P| = \lfloor \sqrt{n} \rfloor$ is not conditioned on the event that it contains I_r (i.e. it is drawn uniformly at random from the set of all $\lfloor \sqrt{n} \rfloor$-element sets, independent of $I_0, \Sigma_1, \ldots, \Sigma_r$). Thus, with probability at least $1 - 1/\sqrt{n}$, $I_r \notin P$ under μ^{mid}. Now Claim 3.2 follows directly from the triangle inequality and Claims 3.3 and 3.4 below.

CLAIM 3.3. *For every $\epsilon > 0$ and $r \in \mathbb{N}$ there exists $\beta, n_0 \in \mathbb{R}_+$ such that for all integers $n \geq n_0$ and ℓ, the distributions $\hat{\mu}_{r,n,\ell}$ and $\mu^{\mathrm{mid}}_{r,n,\ell}$ are $(\epsilon, r+1, \sqrt{n}/\log^{\beta} n)$-indistinguishable.*

CLAIM 3.4. *For every $\epsilon > 0$, there exists $\delta > 0$ such that $\mu^{\mathrm{mid}}_{r,n,\ell}$ and $(\hat{\mu}_{r,n,\ell})_X \otimes (\hat{\mu}_{r,n,\ell})_Y$ are $(\epsilon, \delta\sqrt{n}, \delta\sqrt{n})$-indistinguishable for all $n \in \mathbb{N}$.*

We prove Claim 3.3 below, and defer the proof of Claim 3.4 to the full version.

Proof. (of Claim 3.3) The proof of Claim 3.3 proceeds by using Theorem 3.2. In particular, we will show how Alice and Bob can distinguish between samples from $D^{\mathrm{Y}}_{\mathrm{PV}}(2r-1, n)$ and $D^{\mathrm{N}}_{\mathrm{PV}}(2r-1, n)$ by using a protocol that can distinguish between $\hat{\mu}_{r,n,\ell}$ and $\mu^{\mathrm{mid}}_{r,n,\ell}$.

We assume for simplicity that n is a perfect square (see [GS19] for the case that this is not so). Fix

r, n, ℓ, and suppose that Π is a ρ-round protocol ($\rho \in \mathbb{N}$) with communication at most C that distinguishes between $\hat{\mu}_{r,n^2,\ell}$ from $\mu^{\text{mid}}_{r,n^2,\ell}$ with advantage ϵ. (Notice that we are replacing n with n^2 in the notation.)

We now construct a protocol Π' with the same number of rounds and communication as Π and which distinguishes between $D^{\text{Y}}_{\text{PV}}(2r-1,n)$ and $D^{\text{N}}_{\text{PV}}(2r-1,n)$ with advantage at least ϵ. Suppose Alice and Bob are given inputs $X = (\Sigma_1, \Sigma_3, \dots, \Sigma_{2r-1})$ and $Y = (I_0, J_0, \Sigma_2, \Sigma_4, \dots, \Sigma_{2r-2})$, respectively, which are distributed according to $D^{\text{Y}}_{\text{PV}}(2r-1,n)$ or $D^{\text{N}}_{\text{PV}}(2r-1,n)$. Next, for $1 \le t \le r-1$, let $\Sigma'_t = \Sigma_t$, and for $r+2 \le t \le 2r$, let $\Sigma'_t = \Sigma_{t-1}$. Finally let $\Sigma'_r, \Sigma'_{r+1} \in \mathcal{S}_n$ be uniformly random conditioned on $\Sigma'_{r+1} \circ \Sigma'_r = \Sigma_r$. Notice that each Σ'_t, $1 \le t \le 2r$ may be computed by either Alice or Bob. Next, interpret $[n^2] \simeq [n] \times [n]$, so that any pair $\sigma, \tau \in \mathcal{S}_n$ of permutations on $[n]$ determines a permutation on $[n^2]$, which we denote by $\sigma||\tau$, so that $(\sigma||\tau)((i,j)) = (\sigma(i), \tau(j))$. (Note that the vast majority of permutations on $[n^2]$ cannot be obtained in this manner, however.) The protocol Π' proceeds by taking the chain of permutations $\Sigma'_1, \dots, \Sigma'_{2r} \in \mathcal{S}_n$, "folding it in half" to construct a chain of permutations $\hat{\Sigma}_1, \dots, \hat{\Sigma}_{2\lfloor (r+1)/2 \rfloor} \in \mathcal{S}_{n^2}$, and then running Π on this "folded" chain of permutations. Formally, Π' is given as follows:

1. Alice and Bob use their common randomness to generate uniformly random permutations $\tau_0, \tau_1, \dots, \tau_r \in S_{n^2}$ and uniformly random strings $A_1, \dots, A_{n^2-n}, B_1, \dots, B_{n^2-n}, C_1, \dots, C_n \in \{0,1\}^\ell$.
2. Bob computes $\hat{I}_0 := \tau_0((I_0, J_0)) \in [n] \times [n] \simeq [n^2]$.
3. For $t = 1, 3, \dots, 2\lfloor (r+1)/2 \rfloor$, Alice computes $\hat{\Sigma}_t := \tau_t \circ (\Sigma'_t || (\Sigma'_{2r+1-t})^{-1}) \circ \tau_{t-1}^{-1} \in S_{n^2}$.
4. For $t = 2, 4, \dots, 2\lfloor r/2 \rfloor$, Bob computes $\hat{\Sigma}_t := \tau_t \circ (\Sigma'_t || (\Sigma'_{2r+1-t})^{-1}) \circ \tau_{t-1}^{-1} \in S_{n^2}$.
5. For $1 \le i \le n$, Alice and Bob set $\hat{A}_{\tau_r((i,i))} = \hat{B}_{\tau_r((i,i))} = C_i$.
6. For the $n^2 - n$ pairs $(i,j) \in [n] \times [n]$ with $i \ne j$, Alice sets $\hat{A}_{(i,j)}$ to be equal to one of the A_k, $1 \le k \le n^2 - n$ so that each A_k is used once. Bob does the same with $\hat{B}_{(i,j)}$ with respect to the B_k.
7. Alice and Bob now run the protocol Π on the inputs $\hat{X} := (\hat{\Sigma}_1, \hat{\Sigma}_3, \dots, \hat{\Sigma}_{2\lfloor (r+1)/2 \rfloor}, \hat{A}_1, \dots, \hat{A}_{n^2})$ and $\hat{Y} := (\hat{I}_0, \hat{\Sigma}_2, \hat{\Sigma}_4, \dots, \hat{\Sigma}_{2\lfloor r/2 \rfloor}, \hat{B}_1, \dots, \hat{B}_{n^2})$.

Certainly the communication cost and number of rounds of Π' are both the same as the communication cost and number of rounds, respectively, of Π.

It is not hard to see that (1) if $(X,Y) \sim D^{\text{Y}}_{\text{PV}}(2r-1,n)$, then $(\hat{X},\hat{Y}) \sim \hat{\mu}_{r,n^2,\ell}$, and (2) if $(X,Y) \sim D^{\text{N}}_{\text{PV}}(2r-1,n)$, then $(\hat{X},\hat{Y}) \sim \mu^{\text{mid}}_{r,n^2,\ell}$. Details can be found in [GS19].

Thus the distribution of the transcript of Π' (excluding the additional public randomness used by Π' in the simulation above) when run on D^{Y}_{PV} (respectively, D^{N}_{PV}) is the same as the distribution of the transcript of Π when run on $\hat{\mu}_{r,n^2,\ell}$ (respectively, $\mu^{\text{mid}}_{r,n^2,\ell}$). It then follows from Theorem 3.2 and the fact that $((2r-1)+3)/2 = r+1$ that for every $\epsilon > 0$, there exists $\beta, n_0 \in \mathbb{R}_+$ such that for all $\ell \in \mathbb{N}$ and perfect squares $n \ge n_0$, the distributions $\hat{\mu}_{r,n,\ell}$ and $\mu^{\text{mid}}_{r,n,\ell}$ are $(\epsilon, r+1, \sqrt{n}/\log^\beta n)$-indistinguishable. □

We have now verified Claims 3.3, 3.4, which establishes Claim 3.2, which completes the proof of Theorem 3.1, and thus of Theorem 2.2. □

4 Proof of Theorem 2.3; amortized setting

In this section we work towards the proof of Theorem 2.3; recall that part (1) is immediate, so the main work is in proving parts (2) and (3). As discussed in Section 2.6, there are 3 main steps in the proof, which proceeds by initially assuming that the tuple (C,L) is r-achievable for appropriate values of C, L and eventually deriving a contradiction. The first step is to establish a *single-letter characterization*[4] of the achievable rate region $\mathcal{T}_r(X,Y)$ for amortized CRG, which we explain in Section 4.1. This single-letter characterization will show that if the tuple (C,L) is r-achievable for CRG from any source ν, then there is an r-round protocol with internal information cost at most C and external information cost at least L. In Section 4.2, we show how to convert this protocol into

[4] The term "single-letter characterization" is used relatively loosely in the literature. Following [CK81], for any $k \in \mathbb{N}$ and a closed subset $\mathcal{S} \subset \mathbb{R}^k$, we call a characterization of $\mathcal{S}$ a *single-letter characterization* if it implies, for any $\eta > 0$, the existence of an algorithm that decides whether a point $x \in \mathbb{R}^k$ is of Euclidean distance at most η to $\mathcal{S}$. Moreover, this algorithm must run in time at most $T_{\mathcal{S}}(\eta)$, for some function $T_{\mathcal{S}} : \mathbb{R}_+ \to \mathbb{N}$. This is related, for instance, to ideas on the computability of subsets of $\mathbb{R}^k$ considered in [Bra05].

a nearly equivalent protocol whose *communication cost* is at most C (recall that in general, $\mathrm{CC}(\Pi) \geq \mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi)$, so upper bounding communication cost is more difficult). Finally, in Section 4.3 we show how to use the fact that the external information cost is at least L to obtain a protocol that can distinguish between the pointer-chasing distribution $\mu_{r,n,\ell}$ and the product of the marginals $(\mu_{r,n,\ell})_X \otimes (\mu_{r,n,\ell})_Y$. At this point we will obtain a contradiction for appropriate values of C, L by Theorems 3.1 and 4.4, which were the key ingredients in the proof for the corresponding lower bounds in the non-amortized setting (i.e., item (2) of Theorems 2.1 and 2.2).

4.1 Single-letter characterization of $\mathcal{T}_r(X,Y)$

It follows immediately from Definitions 2.1 and 2.2 that the r-round rate region for amortized CRG and SKG is completely characterized by, for each communication rate C, the maximum real number L, known as the *capacity*, such that (C, L) is r-achievable for CRG or SKG:

DEFINITION 4.1. (CR & SK CAPACITY) *Suppose a source $(X,Y) \sim \nu$ is fixed. Then for $r \in \mathbb{N}, C \in \mathbb{R}_+$, define the CR capacity with communication C to be*

$$\mathscr{C}_r^{\mathtt{am\text{-}cr}}(C) := \sup_{(C,L)\in\mathcal{T}_r(X,Y)} L,$$

and the SK capacity with communication C to be

$$\mathscr{C}_r^{\mathtt{am\text{-}sk}}(C) := \sup_{(C,L)\in\mathcal{S}_r(X,Y)} L.$$

The single-letter characterization of $\mathcal{T}_r(X,Y)$ relies on the concepts of *internal information cost* and *external information cost* of a protocol Π [BBCR13, BR11, BRWY13, BGPW13, Bra12]. The external information cost of a (multiple-round) protocol Π describes how much information Π reveals about the inputs X, Y to an external observer who only sees the transcript of the protocol, while the internal information cost describes how much information Alice and Bob reveal to *each other* about their own inputs:

DEFINITION 4.2. (INFORMATION COSTS) *Given any communication protocol Π with a maximum of r rounds, public randomness $R_{\mathtt{Pub}}$, and a distribution $(X,Y) \sim \nu$ of inputs, the external information cost $\mathrm{IC}^{\mathrm{ext}}_{\nu}(\Pi)$ is given by:*

$$\mathrm{IC}^{\mathrm{ext}}_{\nu}(\Pi) := I(\Pi^r, R_{\mathtt{Pub}}; X, Y).$$

If Π does not use public randomness, then $\mathrm{IC}^{\mathrm{ext}}_{\nu}(\Pi) := I(\Pi^r; X, Y)$.

The internal information cost $\mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi)$ is given by

$$\mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi) := I(\Pi^r, R_{\mathtt{Pub}}; X|Y) + I(\Pi^r, R_{\mathtt{Pub}}; Y|X).$$

If Π does not use public randomness, then $\mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi) := I(\Pi^r; X|Y) + I(\Pi^r; Y|X)$.

REMARK 4.1. *It is well-known that for any distribution ν, and any protocol Π, $\mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi) \leq \mathrm{IC}^{\mathrm{ext}}_{\nu}(\Pi) \leq \mathrm{CC}(\Pi)$.*

An original motivation behind the introduction of internal and external information costs was to understand the possibility of proving *direct sum* results for communication complexity [CSWY01, JRS03, HJMR07, BBCR13]. In light of the connection with direct sum results, the fact that internal and external information costs appear in characterizations for amortized CRG and SKG is not surprising. In particular, the amortized CRG and SKG problems can be viewed as the task of solving N independent instances of CRG or SKG from a source ν, with an additional requirement that each of Alice's N output strings must agree with each of Bob's N output strings *simultaneously* with high probability.

An additional ingredient in the single-letter chararcterization of $\mathcal{T}_r(X,Y)$ is the *minimum r-round interaction for maximum key rate* (i.e., the *r-round MIMK*). Ahlswede and Csiszár showed in their seminal work [AC93] that the maximum key rate L that Alice and Bob can generate from a source $(X,Y) \sim \nu$, without restricting communication, is $I_\nu(X;Y)$. In other words, we have: $\sup_{C\geq 0} \mathscr{C}_r^{\mathtt{am\text{-}sk}}(C) = I(X;Y)$. The r-round MIMK describes the minimum amount of communication needed to obtain this key rate of $I(X;Y)$:

DEFINITION 4.3. *If $(X,Y) \sim \nu$ is a source and $r \geq 1$, Then the r-round MIMK is defined as*

$$\mathscr{I}_r(X;Y) = \inf_{C\geq 0:\mathscr{C}_r^{\mathtt{am\text{-}sk}}(C)=I(X;Y)} \{C\}.$$

Tyagi [Tya13] proved the following single-letter characterization of the r-round MIMK $\mathscr{I}_r(X;Y)$:

THEOREM 4.1. ([TYA13], THEOREM 4) *For a source $(X,Y) \sim \nu$, the r-round MIMK is the infimum of all $C \geq 0$ such that there exists an r-round private-coin protocol Π such that $\mathrm{IC}^{\mathrm{int}}_{\nu}(\Pi) \leq C$ and $\mathrm{IC}^{\mathrm{ext}}_{\nu}(\Pi) \geq C + I(X;Y)$.*

Using Theorem 4.1, we finally can state the single-letter characterization of achievable rates for

r-round CRG. It is stated most precisely in [STW19], but similar results are shown in [LCV17, GJ18, Liu16, Ye05, GA10a, GA10b].

THEOREM 4.2. ([STW19], THEOREM III.2) *For $C \geq 0$, define*

(4.3)
$$\tilde{\mathscr{C}}_r^{\text{am-cr}}(C) := \begin{cases} \sup_{\Pi}\{\mathrm{IC}_\nu^{\mathrm{ext}}(\Pi)\} & : \quad C \leq \mathscr{I}_r(X;Y) \\ I(X;Y)+C & : \quad C > \mathscr{I}_r(X;Y), \end{cases}$$

where the supremum is over all r-round private-coin protocols $\Pi = (\Pi_1, \ldots, \Pi_r)$ with $\mathrm{IC}_\nu^{\mathrm{int}}(\Pi) \leq C$.

Then for a source $(X,Y) \sim \nu$, the r-round CR capacity is given by

$$\mathscr{C}_r^{\text{am-cr}}(C) = \tilde{\mathscr{C}}_r^{\text{am-cr}}(C). \tag{4.4}$$

For the purpose of proving Theorem 2.3, we will only need the inequality $\mathscr{C}_r^{\text{am-cr}}(C) \leq \tilde{\mathscr{C}}_r^{\text{am-cr}}(C)$ (which is often called the *converse direction* of the equality in Theorem 4.2). As a full proof of Theorem 4.2 (and in particular, of this inequality) does not appear to have been collected in the literature, we provide one in the full version of this paper [GS19, Section 5]. The following is an immediate consequence of this inequality:

COROLLARY 4.1. *For a source $(X,Y) \sim \nu$, for each tuple $(C,L) \in \mathcal{T}_r(X,Y)$ with $L < I(X;Y)$, there is some protocol $\Pi = (\Pi_1, \ldots, \Pi_r)$ such that $\mathrm{IC}_\nu^{\mathrm{int}}(\Pi) \leq C$ and $\mathrm{IC}_\nu^{\mathrm{ext}}(\Pi) \geq L$.*

See [GS19] for a proof.

4.2 Using the compression of internal information to communication A crucial technical ingredient in doing so is the use of an "compression of internal information cost to communication" result for bounded round protocols, saying that for any protocol with a fixed number r of rounds and internal information cost I, there is another protocol with the same number r of rounds and communication cost not much larger than I. As we discussed in Section 2, these types of theorems were originally proved in order to establish direct sum and direct product results for communication complexity. Our use of these compression results may be interpreted as a roughly analogous approach for the setting of amortized CRG and SKG, which can be thought of as the "direct sum version of non-amortized CRG and SKG".

THEOREM 4.3. (LEMMA 3.4, [JPY12]) *Suppose that $(X,Y) \sim \nu$ are inputs to an r-round communication protocol Π with public randomness R_{Pub} (and which may use private coins as well). Then for every $\epsilon > 0$, there is a public coin protocol L with r rounds and communication at most $\frac{\mathrm{IC}_\nu^{\mathrm{int}}(\Pi)+5r}{\epsilon} + O(r\log(1/\epsilon))$ such that at the end of the protocol the parties possess random variables $((\hat{\Pi}_A)_1, \ldots, (\hat{\Pi}_A)_r)$, $((\hat{\Pi}_B)_1, \ldots, (\hat{\Pi}_B)_r)$, each representing a transcript for Π, which satisfy*

$$\begin{aligned} \Delta((R_{Pub}, X, Y, (\hat{\Pi}_A)_1, \ldots, (\hat{\Pi}_A)_r), & \\ (R_{Pub}, X, Y, \Pi_1, \ldots, \Pi_r)) &\leq 6\epsilon r \\ \Delta((R_{Pub}, X, Y, (\hat{\Pi}_B)_1, \ldots, (\hat{\Pi}_B)_r), & \\ (R_{Pub}, X, Y, \Pi_1, \ldots, \Pi_r)) &\leq 6\epsilon r \\ \mathbb{P}[((\hat{\Pi}_B)_1, \ldots, (\hat{\Pi}_B)_r) \neq ((\hat{\Pi}_A)_1, \ldots, (\hat{\Pi}_A)_r)] &\leq 6\epsilon r. \end{aligned}$$

Our first lemma, Lemma 4.1, uses Theorem 4.3 to show that for any protocol Π which satisfies $\mathrm{IC}_\mu^{\mathrm{ext}}(\Pi) \gg \mathrm{IC}_\mu^{\mathrm{int}}(\Pi)$ for the source $\mu = \mu_{r,n,L}$, then there exists another protocol Π with communication cost not much greater than $\mathrm{IC}_\mu^{\mathrm{int}}(\Pi)$ and which satisfies some additional properties (which arise from $\mathrm{IC}_\mu^{\mathrm{ext}}(\Pi)$ being large):

LEMMA 4.1. *Fix any $r, n, \ell \in \mathbb{N}$, and let $\mu = \mu_{r,n,\ell}$. Suppose $\rho \in \mathbb{N}$ and $C, L \in \mathbb{R}_+$. Suppose Π is a ρ-round protocol with $\mathrm{IC}_\mu^{\mathrm{ext}}(\Pi) = L$ and $\mathrm{IC}_\mu^{\mathrm{int}}(\Pi) = C$ and public randomness R_{Pub} (and which may use private randomness as well). Then for every $\epsilon > 0$ there is some ρ-round protocol Π' with inputs $(X,Y) \sim \mu$, public randomness R_{Pub}, with communication at most $\frac{C+5\rho}{\epsilon} + O(\rho \log 1/\epsilon)$ and which outputs keys K'_A, K'_B, such that*

1. $\mathbb{P}_\mu[K'_A = K'_B] \geq 1 - 6\epsilon\rho$.
2. *When inputs (X,Y) are drawn from μ, $I(K'_A; B_{I_r}) = I(K'_A; A_{I_r}) \geq L - (C + 1 + 2\log n + 36\epsilon\rho\ell)$.*
3. *When inputs (X,Y) are drawn from $\mu_X \otimes \mu_Y$,*

$$\begin{aligned} & I_{\mu_X \otimes \mu_Y}(K'_A, R_{Pub}, (\Pi')^\rho; B_1, \ldots, B_n) \\ & \leq \frac{C+5\rho}{\epsilon} + O(\rho \log 1/\epsilon) \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} & I_{\mu_X \otimes \mu_Y}(K'_B, R_{Pub}, (\Pi')^\rho; A_1, \ldots, A_n) \\ & \leq \frac{C+5\rho}{\epsilon} + O(\rho \log 1/\epsilon). \end{aligned} \tag{4.6}$$

Proof. Let Π' be the protocol given by Theorem 4.3 for the protocol Π and the given ϵ. Then the communication of Π' is at most $\frac{C+5\rho}{\epsilon} + O(\rho \log 1/\epsilon))$. At

the end of Π', Alice and Bob possess random variables $((\hat{\Pi}_{\mathtt{A}})_1, \ldots, (\hat{\Pi}_{\mathtt{A}})_\rho)$, $((\hat{\Pi}_{\mathtt{B}})_1, \ldots, (\hat{\Pi}_{\mathtt{B}})_\rho)$, respectively, such that, when $(X,Y) \sim \mu$,

$$\begin{aligned}&\Delta((R_{\mathtt{Pub}}, X, Y, (\hat{\Pi}_{\mathtt{A}})_1, \ldots, (\hat{\Pi}_{\mathtt{A}})_\rho),\\ &\qquad (R_{\mathtt{Pub}}, X, Y, \Pi_1, \ldots, \Pi_\rho)) \le 6\epsilon\rho. \end{aligned} \tag{4.7}$$

(Notice that $\hat{\Pi}^\rho_{\mathtt{A}} = ((\hat{\Pi}_{\mathtt{A}})_1, \ldots, (\hat{\Pi}_{\mathtt{A}})_\rho)$ and $\hat{\Pi}^\rho_{\mathtt{B}}$ are different from the transcript $(\Pi')^\rho = (\Pi'_1, \ldots, \Pi'_\rho)$ of Π'.) Now Alice sets $K'_{\mathtt{A}} = \hat{\Pi}^\rho_{\mathtt{A}}$ and Bob sets $K'_{\mathtt{B}} = \hat{\Pi}^\rho_{\mathtt{B}}$, which immediately establishes item (1) of the lemma (by Theorem 4.3).

To establish point (2), we will first argue that it holds for Π; in particular we show that when $(X,Y) \sim \mu$,

$$H(B_{I_r} | \Pi^\rho) \le \ell + C - L + 2\log n. \tag{4.8}$$

(Since $H(B_{i_r}) = \ell$ it will follow from (4.8) that $I_\mu(\Pi^\rho; B_{I_r}) \ge L - C - 2\log n$, though we will not use this directly.) To see this, first notice that

$$\begin{aligned} I(X; Y | \Pi^\rho) &= I(Y; X, \Pi^\rho) - I(\Pi^\rho; Y)\\ &= I(X;Y) + I(\Pi^\rho; Y | X) +\\ &\quad I(\Pi^\rho; X | Y) - I(\Pi^\rho; X, Y)\\ &= I(X;Y) + \mathrm{IC}^{\mathrm{int}}_\mu(\Pi) - \mathrm{IC}^{\mathrm{ext}}_\mu(\Pi) \qquad (4.9)\\ &\le \ell + C - L. \end{aligned}$$

Recalling the notation $I_r = \Sigma_r \circ \cdots \circ \Sigma_1(I_0)$, we observe by [GS19, Lemma 6.2] and the data processing inequality that

$$\begin{aligned} I(X; Y | \Pi^\rho) &\ge I(X; Y | \Pi^\rho, I_r) - \log n\\ &\ge I(A_{I_r}; B_{I_r} | \Pi^\rho, I_r) - \log n\\ &\ge I(A_{I_r}; B_{I_r} | \Pi^\rho) - 2\log n\\ &= H(A_{I_r} | \Pi^\rho) - 2\log n\\ &= H(B_{I_r} | \Pi^\rho) - 2\log n, \end{aligned}$$

since $H(A_{I_r} | B_{I_r}, \Pi^\rho) = H(A_{I_r} | B_{I_r}) = 0$ as $A_{I_r} = B_{I_r}$ for all inputs in the support of μ. It then follows that $H(B_{I_r} | \Pi^\rho, R_{\mathtt{Pub}}) \le \ell + C - L + 2\log n$, establishing (4.8).

Next, (4.7) and the data processing inequality give us that $\Delta((R_{\mathtt{Pub}}, B_{I_r}, \Pi^\rho), (R_{\mathtt{Pub}}, B_{I_r}, (\hat{\Pi}_{\mathtt{A}})^\rho)) \le 6\epsilon\rho$. [GS19, Corollary 6.5] and (4.8) then give that

$$\begin{aligned} &H(B_{I_r} | (\hat{\Pi}_{\mathtt{A}})^\rho, R_{\mathtt{Pub}})\\ &\le H(B_{I_r} | (\hat{\Pi}_{\mathtt{A}})^\rho)\\ &\le \ell + C - L + 2\log n + 36\epsilon\rho\ell + 1. \end{aligned}$$

Since $K'_{\mathtt{A}} = \hat{\Pi}^\rho$, we get that

$$I(B_{I_r}; K'_{\mathtt{A}}) \ge L - (C + 1 + 2\log n + 36\epsilon\rho\ell),$$

which establishes point (2).

Finally, to establish point (3), first notice that some inputs $(X,Y) \sim \mu_X \otimes \mu_Y$ may not be in the support of μ. We may extend the protocol Π' to be defined for all pairs of inputs $(X,Y) \in \mathcal{X} \times \mathcal{Y}$, by choosing an arbitrary behavior (e.g., terminating immediately) whenever there is a partial transcript $(\Pi')^{t-1}$ for which the distribution of the next message Π'_t has not been defined.

Recall that $(\Pi'_1, \ldots, \Pi'_\rho)$ denotes the transcript of communication of Π' and $R_{\mathtt{Pub}}$ is the public randomness of Π', so that when $(X,Y) \sim \mu_X \otimes \mu_Y$,

$$\begin{aligned} &I_{\mu_X \otimes \mu_Y}((\Pi')^\rho, X, R_{\mathtt{Pub}}; Y)\\ &= I_{\mu_X \otimes \mu_Y}((\Pi')^\rho; Y | X, R_{\mathtt{Pub}})\\ &\le H_{\mu_X \otimes \mu_Y}((\Pi')^\rho) \le \frac{C + 5\rho}{\epsilon} + O(\rho \log 1/\epsilon). \end{aligned}$$

Recalling that $K'_{\mathtt{A}} = \hat{\Pi}^\rho_{\mathtt{A}}$, by construction of Π' (and $\hat{\Pi}^\rho_{\mathtt{A}}$) from Theorem 4.3, it follows that

$$(K'_{\mathtt{A}}, R_{\mathtt{Pub}}, (\Pi')^\rho) - (X, (\Pi')^\rho, R_{\mathtt{Pub}}) - Y$$

is a Markov chain. It then follows from the data processing inequality that

$$\begin{aligned} &I_{\mu_X \otimes \mu_Y}(K'_{\mathtt{A}}, R_{\mathtt{Pub}}, (\Pi')^\rho; B_1, \ldots, B_n)\\ &\le I_{\mu_X \otimes \mu_Y}(\hat{K_{\mathtt{A}}}', R_{\mathtt{Pub}}, (\Pi')^\rho; Y)\\ &\le \frac{C + 5\rho}{\epsilon} + O(\rho \log 1/\epsilon), \end{aligned}$$

which gives (4.5); (4.6) follows in a similar manner. □

Roughly speaking, the next lemma, Lemma 4.2, shows how the protocol Π' constructed in Lemma 4.1 can use the properties (2) and (3) of Lemma 4.1 to distinguish between the distributions $\mu = \mu_{r,n,L}$ (which corresponds to ν_1 in the below statement) and $\mu_X \otimes \mu_Y$ (corresponding to ν_2 in the below statement). This, in combination with the result from Theorem 3.1 stating that μ and $\mu_X \otimes \mu_Y$ are indistinguishable to protocols with little communication, will ultimately complete the proof of Theorem 2.3.

LEMMA 4.2. *Suppose ν_1, ν_2 are distributions over tuples of random variables $(Z_1, \ldots, Z_n, I, K, \tilde{K})$, where $Z_1, \ldots, Z_n \in \{0,1\}^\ell$, $I \in [n]$, and $K \in \mathcal{K}$, where $\mathcal{K}$ is a finite set. Suppose that the marginal distribution*

of $Z_1, \dots, Z_n, I$ over each of ν_1, ν_2 is uniform over $\{0,1\}^{n\ell} \times [n]$. Finally suppose that $0 < \xi < 1$ and C satisfy $\log n \le C \le \frac{(1-\xi)^3 \ell}{1620}$ as well as:

1. $I_{\nu_1}(K; Z_1, \dots, Z_n) \le C$.
2. $I_{\nu_2}(K; Z_I) \ge \ell(1-\xi)$.
3. $\min\left\{\mathbb{P}_{\nu_2}[K = \tilde{K}], \mathbb{P}_{\nu_1}[K = \tilde{K}]\right\} \ge 1 - (1-\xi)^2/72$.

Then there is some function $f : \mathcal{K} \times \{0,1\}^{n\ell} \to \{0,1\}$ such that

$$\left|\mathbb{E}_{\nu_1}[f(\tilde{K}, Z_1, \dots, Z_n)] - \mathbb{E}_{\nu_2}[f(\tilde{K}, Z_1, \dots, Z_n)]\right| \ge p/2,$$

where $p = (1-\xi)^2/18$.

The idea of the proof of Lemma 4.2 is as follows. The second condition of the lemma can be shown to imply that $H_{\nu_2}(Z_I|K)$ is small, which means that for each value of $k \in \mathcal{K}$, there is a small subset $\mathcal{T}_k$ of $\{0,1\}^\ell$ to which Z_I belongs with high probability conditioned on the event $K = k$. Then the function f can be chosen to be 1 whenever *any* Z_i belongs to $\mathcal{T}_K$. That f is 1 with high probability under ν_2 follows from construction of f, and that f is 1 with not too high probability under ν_1 follows from the fact that all $\mathcal{T}_k$ are small.

We first establish some basic lemmas before proving Lemma 4.2 rigorously. Lemma 4.3, an immediate consequence of Markov's inequality, states that for a random variable with low entropy, it belongs to some small set with high probability.

LEMMA 4.3. *Suppose $W \in \{0,1\}^\ell$ is a random variable, and $H(W) = c$. For any $\delta \in (0,1]$ there is some set $\mathcal{S} \subset \{0,1\}^\ell$ such that $|\mathcal{S}| \le 2^{c/\delta}$ and $\mathbb{P}[W \notin \mathcal{S}] \le \delta$.*

Proof. Set

$$\mathcal{S} = \{w \in \{0,1\}^\ell : \mathbb{P}[W = w] \ge 2^{-c/\delta}\}.$$

We know that $c = H(W) = \mathbb{E}_{w \sim W}[\log(1/\mathbb{P}[W = w])]$, so the probability that $\mathbb{P}[W = w] < 2^{-c/\delta}$, i.e. that $\log(1/\mathbb{P}[W = w]) > c/\delta$, over $w \sim W$ is at most δ. Thus $\mathbb{P}[W \notin \mathcal{S}] \le \delta$. Clearly, by the definition of $\mathcal{S}$, we have that $|\mathcal{S}| \le 2^{c/\delta}$. ∎

Lemma 4.4, a sort of converse to Lemma 4.3, states that for a random variable with high entropy, it *does not* belong to any small set with high probability.

LEMMA 4.4. *Suppose that $W \in \{0,1\}^\ell$ is a random variable with $H(W) = h \le \ell$. Let $\mathcal{S} \subset \{0,1\}^\ell$ be a subset with size $|\mathcal{S}| \le 2^c$, for some $c < \ell$. Then $\mathbb{P}[W \in \mathcal{S}] \le \frac{\ell+1-h}{\ell-c}$.*

Proof. Write $p = \mathbb{P}[W \in \mathcal{S}]$. Let $J = \mathbf{1}[W \in \mathcal{S}]$. Then $pc + (1-p)\ell \ge pc + (1-p)\log(2^\ell - 2^c) \ge H(W|J) \ge H(W) - 1 = h - 1$. Hence $p(c-\ell) \ge h - 1 - \ell$, so $p \le \frac{\ell+1-h}{\ell-c}$. ∎

Lemma 4.5 is needed in order to reason about the random variable Z_I in Lemma 4.2.

LEMMA 4.5. *Suppose that random variables $I, Z_1, \dots, Z_n$ are distributed jointly so that the marginal of $Z_1, \dots, Z_n \in \{0,1\}^\ell$ is uniform on $\{0,1\}^{n\ell}$. Then $H(Z_I) \ge \ell - \log n$.*

Proof. Notice that

$$\begin{aligned} & H(Z_I, Z_{I+1}, \dots, Z_{I+n-1}) \\ \ge\ & H(Z_I, \dots, Z_{I+n-1}|I) \\ =\ & \mathbb{E}_{i \sim I}\left[H(Z_i, \dots, Z_{i+n-1}|I = i)\right] \\ =\ & \mathbb{E}_{i \sim I}\left[H(Z_1, \dots, Z_n|I = i)\right] \\ =\ & H(Z_1, \dots, Z_n|I) \\ \ge\ & \ell n - \log n, \end{aligned} \tag{4.10}$$

where addition of subscripts is taken modulo n. Since $(Z_{I+1}, \dots, Z_{I+n-1}) \in \{0,1\}^{\ell n - \ell}$, we get that

$$\begin{aligned} H(Z_I) &\ge H(Z_I|Z_{I+1}, \dots, Z_{I+n-1}) \\ &\ge H(Z_I, \dots, Z_{I+n-1}) - (\ell n - \ell) \\ &\ge \ell - \log n, \end{aligned}$$

as desired. ∎

Now we prove Lemma 4.2.

Proof. (of Lemma 4.2) We will first define f and determine a lower bound on $\mathbb{E}_{\nu_2}[f(K, Z_1, \dots, Z_n)]$. By assumption, $H_{\nu_2}(Z_I) = \ell$, so $H_{\nu_2}(Z_I|K) \le \xi\ell$. For each $k \in \mathcal{K}$, let $\gamma_k = H(Z_I|K = k)/\ell$, so that $\mathbb{E}_{k \sim K}[\gamma_k] \le \xi$. Pick some $\eta > 1, \zeta > 1$ to be specified later. By Lemma 4.3, for each $k \in \mathcal{K}$, there is a set $\mathcal{T}_k \subset \{0,1\}^\ell$ of size at most $2^{\eta\gamma_k\ell}$ such that $\mathbb{P}_{\nu_2}[Z_I \notin \mathcal{T}_k|K = k] \le 1/\eta$. Next, set $\mathcal{S} = \{k \in \mathcal{K} : \gamma_k \le \zeta\xi\}$. By Markov's inequality, $\mathbb{P}_{\nu_2}[K \in \mathcal{S}] \ge 1 - 1/\zeta$. Thus $\mathbb{P}_{\nu_2}[K \in \mathcal{S}] \cdot \mathbb{P}_{\nu_2}[Z_I \in T_K | K \in \mathcal{S}] \ge (1-1/\zeta)\cdot(1-1/\eta)$, and for all $k \in \mathcal{S}$, $|\mathcal{T}_k| \le 2^{\eta\zeta\xi\ell} < 2^{\eta\zeta\ell}$.

We now set

$$f(K, Z_1, \dots, Z_n) = \begin{cases} \bigvee_{i \in [n]} \mathbf{1}[Z_i \in \mathcal{T}_K] & : \quad K \in \mathcal{S} \\ 0 & : \quad \text{else.} \end{cases}$$

Since $\mathbb{P}[K \in \mathcal{S}] \cdot \mathbb{P}[Z_I \in \mathcal{T}_K | K \in \mathcal{S}] \leq \mathbb{E}\left[\vee_{i\in[n]} \mathbf{1}[Z_i \in \mathcal{T}_K]\right]$,

$$\mathbb{E}_{\nu_2}[f(K, Z_1, \ldots, Z_n)] \geq (1 - 1/\eta) \cdot (1 - 1/\zeta).$$

Next we determine an upper bound on $\mathbb{E}_{\nu_1}[f(K, Z_1, \ldots, Z_n)]$. Define a random variable $\hat{I} = \hat{I}(Z_1, \ldots, Z_n, K)$, by $\hat{I} = \min\{i : Z_i \in \mathcal{T}_K\}$, if the set $\{i : Z_i \in \mathcal{T}_K\}$ is nonempty, else $\hat{I} = 1$. Thus $H(\hat{I}) \leq \log n$. Consider the random variable $Z_{\hat{I}} \in \{0,1\}^\ell$. It follows that $f(K, Z_1, \ldots, Z_n) \leq \mathbf{1}[Z_{\hat{I}} \in \mathcal{T}_K]$. By [GS19, Lemma 6.2] and the data processing inequality, we have that

$$\begin{aligned} &I_{\nu_1}(K; Z_{\hat{I}}) - \log n \leq I_{\nu_1}(K; Z_{\hat{I}} | \hat{I}) \\ &\leq I_{\nu_1}(K; Z_1, \ldots, Z_n | \hat{I}) \\ &\leq I_{\nu_1}(K; Z_1, \ldots, Z_n) + \log n \\ &\leq C + \log n. \end{aligned}$$

Lemma 4.5 gives that $H_{\nu_1}(Z_{\hat{I}}) \geq \ell - \log n$, so $H_{\nu_1}(Z_{\hat{I}} | K) \geq \ell - C - 3\log n$. For each $k \in \mathcal{K}$, let $h_k = H_{\nu_1}(Z_{\hat{I}} | K = k)$, so that $\mathbb{E}_{\nu_1}[h_K] \geq \ell - C - 3\log n$. By Lemma 4.4, for each $k \in \mathcal{K}$ with $\eta\gamma_k < 1$, $\mathbb{P}[Z_{\hat{I}} \in \mathcal{T}_K | K = k] \leq \frac{\ell + 1 - h_k}{\ell(1 - \eta\gamma_k)}$, by our upper bound $|\mathcal{T}_k| \leq 2^{\eta\gamma_k \ell}$.

Recall that $\mathbb{E}_{\nu_2}[\gamma_k] \leq \xi$. For $i \in \{1,2\}$, let K_{ν_i} be the marginal distribution of K according to ν_i. We must have that $\Delta(K_{\nu_2}, K_{\nu_1}) < p$, else we could choose f to be a function of only K and would get that $|\mathbb{E}_{\nu_1}[f] - \mathbb{E}_{\nu_2}[f]| \geq p$. Thus $1 - 1/\zeta - p \leq \mathbb{P}_{\nu_1}[K \in \mathcal{S}] \leq 1$. Next notice that $\mathbb{E}_{\nu_1}[\ell - h_K] \leq C + 3\log n$, and that $\ell - h_K \geq 0$ with probability 1. Therefore, $\mathbb{E}_{\nu_1}[\ell - h_K | K \in \mathcal{S}] \leq \frac{C + 3\log n}{1 - 1/\zeta - p}$. Since $\gamma_k \leq \zeta\xi$ for all $k \in \mathcal{S}$, it follows that

$$\begin{aligned} &\mathbb{E}_{\nu_1}[f(K, Z_1, \ldots, Z_n)] \\ \leq\ &\mathbb{E}_{\nu_1}[f(K, Z_1, \ldots, Z_n) | K \in \mathcal{S}] \\ \leq\ &\mathbb{P}_{\nu_1}[Z_{\hat{I}} \in \mathcal{T}_K | K \in \mathcal{S}] \\ \leq\ &\frac{1 + \frac{C + 3\log n}{1 - 1/\zeta - p}}{\ell(1 - \eta\zeta\xi)}. \end{aligned}$$

Thus

$$\begin{aligned} &\mathbb{E}_{\nu_2}[f(K, Z_1, \ldots, Z_n)] - \mathbb{E}_{\nu_1}[f(K, Z_1, \ldots, Z_n)] \\ &\geq (1 - 1/\zeta) \cdot \left((1 - 1/\eta) - \frac{1}{1 - 1/\zeta} \cdot \frac{1 + \frac{C + 3\log n}{1 - 1/\zeta - p}}{\ell(1 - \eta\zeta\xi)} \right). \end{aligned}$$

Now, choose $\eta = \zeta = \xi^{-1/3}$, and let $\xi' = 1 - \xi$, so that $p \leq \xi'/6 \leq \frac{1 - (1 - \xi')^{1/3}}{2} = \frac{1 - 1/\zeta}{2}$. Using the inequality $ax \leq 1 - (1 - x)^a \leq x$ for $0 < a < 1$, $x \in [0,1]$ and $C \geq \log n$ gives

$$\begin{aligned} &\mathbb{E}_{\nu_2}[f(K, Z_1, \ldots, Z_n)] - \mathbb{E}_{\nu_1}[f(K, Z_1, \ldots, Z_n)] \\ \geq\ &\xi'/3 \cdot \left(\xi'/3 - \frac{1}{(1 - 1/\zeta)(1 - 1/\zeta - p)} \cdot \frac{15C}{\xi'\ell} \right) \\ \geq\ &\xi'/3 \cdot \left(\xi'/3 - \frac{270C}{(\xi')^2 \ell} \right) \\ \geq\ &(\xi')^2/18 = p, \end{aligned}$$

where the last inequality follows from $C \leq \frac{(\xi')^3 \ell}{1620}$.

Since $K, \tilde{K}$ are nonequal with probability at most $p/4$ over each of ν_1, ν_2, it follows that

$$\mathbb{E}_{\nu_2}[f(\tilde{K}, Z_1, \ldots, Z_n)] - \mathbb{E}_{\nu_1}[f(\tilde{K}, Z_1, \ldots, Z_n)] \geq p/2,$$

as desired. □

4.3 Proof of Theorem 2.3 Theorem 2.3 is proven as follows: Theorem 4.2 (see also Corollary 4.1) states that if the rate (C, L) is achievable under the source $\mu = \mu_{r,n,\ell}$, then there is a protocol Π with internal information cost bounded above by C and bounded below by L under μ. Lemma 4.1 implies that such a protocol Π can be compressed to a protocol Π' with communication not much more than C, with the same number of rounds as Π, and for which Alice and Bob can output keys at the end which "give some information" about the strings A_{I_r}, B_{I_r} under μ. Lemma 4.2 shows how this information can be used to distinguish μ and $\mu_X \otimes \mu_Y$. Theorems 3.1 and 4.4 state that unless C is sufficiently large, this is impossible, thus establishing the desired lower bound.

Proof. (of Theorem 2.3) The first part of Theorem 2.3 follows in the same way as the non-amortized case. To prove the second part, first suppose r is odd. We take $\mu = \mu_{r,n,\ell}$ and set $\epsilon = \min\{\gamma/(54(r+1)), \gamma^2/(2592 \cdot 6r)\}$.

We argue by contradiction. Suppose the theorem statement is false: namely, that for some $C \leq n/\log^{c_0} n$ and $L > \gamma\ell$, the tuple (C, L) is $\lfloor (r+1)/2 \rfloor$-achievable from μ. We can assume without loss of generality that $L < \ell$. By Theorem 4.2 (and in particular, Corollary 4.1), since $I_\mu(X;Y) = \ell > L$, there is a $\lfloor (r+1)/2 \rfloor$-round protocol Π such that $\mathrm{IC}^{\mathrm{int}}_\mu(\Pi) \leq C$ and $\mathrm{IC}^{\mathrm{ext}}_\mu(\Pi) \geq L$.

By Lemma 4.1, there is an $\lfloor (r+1)/2 \rfloor$-round public-coin protocol Π' with inputs $(X, Y) \sim \mu$ and communication at most $\frac{C + 3 + 5r/2}{\epsilon} + O(r \log 1/\epsilon)$ such that at the end of Π' with inputs $(X, Y) \sim \mu$, Alice

and Bob output keys K'_A, K'_B, respectively, which satisfy $\mathbb{P}_\mu[K'_A \neq K'_B] \leq 6\epsilon r$ and $I_\mu(K'_B; B_{I_r}) \geq L - (C + 1 + 2\log n + 18\epsilon(r+1)\ell)$. Moreover, when $(X,Y) \sim \mu_X \otimes \mu_Y$,

$$\begin{aligned}&\max\{I_{\mu_X\otimes\mu_Y}(K'_A; B_1, \ldots, B_n),\\ &I_{\mu_X\otimes\mu_Y}(K'_B; A_1, \ldots, A_n)\}\\ &\leq \frac{C+3+5r/2}{\epsilon} + O(r\log 1/\epsilon).\end{aligned}$$

Next, let Π'' be the protocol where the parties run Π', and the last party (suppose it is Alice, for concreteness) to speak in Π' sends over a random hash $h(K'_A)$ of length $O(\log 1/\gamma)$, so that for any $K'_A \neq K'_B$, $\mathbb{P}_h[h(K'_A) = h(K'_B)] \leq \gamma^2/1296$, and the other party, Bob, outputs a final bit equal to $\mathbf{1}[h(K'_A) = h(K'_B)]$. For sufficiently large n, we have that

$$\begin{aligned}&\mathrm{CC}(\Pi'')\\ &\leq \frac{C+3+5r/2}{\epsilon} + O(r\log 1/\epsilon) + O(\log 1/\gamma)\\ &\leq n/\log^{(c_0-1)} n. \qquad (4.11)\end{aligned}$$

CLAIM 4.1. *Π'' distinguishes μ and $\mu_X \otimes \mu_Y$ with advantage at least $\gamma^2/2592$.*

Proof. To prove Claim 4.1, we consider two cases.

The first case is that $\mathbb{P}_{\mu_X\otimes\mu_Y}[K'_A \neq K'_B] \geq \gamma^2/648$. In this case, the last bit output by Bob will be 0 with probability at least $\gamma^2/1296$ when $(X,Y) \sim \mu_X \otimes \mu_Y$. Since $K'_A = K'_B$ with probability at least $1 - 6\epsilon r \geq 1 - \gamma^2/2592$ when $(X,Y) \sim \mu$, it follows that Π'' distinguishes between the two distributions with advantage at least $\gamma^2/2592$ in this case.

The second case is that $\mathbb{P}_{\mu_X\otimes\mu_Y}[K'_A \neq K'_B] \leq \gamma^2/648$. Here we will use Lemma 4.2. Since $18\epsilon(r+1) \leq \gamma/3$, and since for sufficiently large n, $C + 1 + 2\log n \leq \gamma n/3 = \gamma\ell/3$, we see that $I_\mu(K'_B; B_{I_r}) \geq \gamma\ell - 2\gamma\ell/3 = \gamma\ell/3$.

We apply Lemma 4.2, with $(Z_1, \ldots, Z_n) = (B_1, \ldots, B_n), I = I_r, K = K'_A, \tilde{K} = K'_B, \nu_1 = \mu_X \otimes \mu_Y, \nu_2 = \mu$, $\xi = 1 - \gamma/3$ and $L = n/\log^{(c_0-1)} n$. Here we use that $n/\log^{(c_0-1)} n \leq \frac{(\gamma/3)^3 n}{1620}$ for sufficiently large n (depending on γ), as well as $\mathbb{P}_{\mu_X\otimes\mu_Y}[K'_A \neq K'_B] \leq \gamma^2/648 = (1-\xi)^2/72$. Then Lemma 4.2 gives that Bob can output a bit as a deterministic function of $K'_B, B_1, \ldots, B_n$ (all of which Bob holds at the conclusion of Π'), that distinguishes μ and $\mu_X \otimes \mu_Y$ with advantage at least $\gamma^2/324$. ∎

By Theorem 4.4 below (which is analogous to Theorem 3.1), with $\epsilon = \gamma^2/2592$, and as long as c_0 is large enough so that (4.11) holds for $n \geq c_0$, and such that $c_0 - 1 \geq \beta$ (where β is chosen from Theorem 4.4, given $\epsilon = \gamma^2/2592$), we arrive at a contradiction.

THEOREM 4.4. ([BGGS19], LEMMA 4.5) *For every $\epsilon > 0$ and odd r there exists $\beta > 0$ such that for every $n \geq \beta$ and ℓ, the distributions $\mu = \mu_{r,n,\ell}$ and $\mu_X \otimes \mu_Y$ are $(\epsilon, (r+3)/2, n/\log^\beta n)$-indistinguishable.*

The case of even r and the proof of part (3) of the theorem are handled similarly; details may be found in [GS19]. ∎

4.4 Separations in MIMK

In this section we use Theorem 2.3 to derive separations in the MIMK for the pointer chasing source $\mu_{r,n,\ell}$ (recall Definition 4.3). The below Theorem 4.5 generalizes a result of Tyagi [Tya13], which established a constant-factor separation in the MIMK for 2-round and 1-round protocols for a certain source.

THEOREM 4.5. *For each $r \in \mathbb{N}$, there is a c_0 such that for each $n \geq c_0$, the pointer chasing source $\mu_{r,n,n}$ satisfies:*

1. $\mathscr{I}_{r+2}(X;Y) \leq (r+2)\lceil \log n \rceil$.
2. $\mathscr{I}_{\lfloor (r+1)/2 \rfloor}(X;Y) > n/\log^{c_0} n$.
3. $\mathscr{I}_r(X;Y) > \sqrt{n}/\log^{c_0} n$.

Theorem 4.5 is a consequence of Theorem 2.3 and the proof is given in [GS19].

5 Discussion: CBIB, KBIB

Notice that the MIMK deals with very large rates of communication; in particular, communication at rates larger than the MIMK is no longer interesting, as, for instance, the entropy rate L for SKG is fixed at $I(X;Y)$. One can ask, on the other hand, whether Theorem 2.3 allows us to determine a separation in some measure that determines the efficiency of CRG and SKG at very *small* rates of communication. Formally, we consider the *common random bits per r-round interaction bit (r-round CBIB)* and the *secret key bits per r-round interaction bit (r-round KBIB)*:

DEFINITION 5.1. ([LCV17], COROLLARY 2[5]) *For a source $(X,Y) \sim \nu$ and $r \in \mathbb{N}$, define:*

$$\Gamma^{cr}_r(X,Y) = \sup\left\{\frac{L}{C} : (C,L) \in \mathcal{T}_r(X,Y), C > 0\right\}$$

and

$$\Gamma_r^{\mathtt{sk}}(X,Y) = \sup\left\{\frac{L}{C} : (C,L) \in \mathcal{S}_r(X,Y), C > 0\right\}.$$

Notice that $\Gamma_r^{\mathtt{cr}}(X,Y)$ and $\Gamma_r^{\mathtt{sk}}(X,Y)$ can be infinite, if, for instance, there are functions $f_{\mathtt{A}} : \mathcal{X} \to \{0,1\}$ and $f_{\mathtt{B}} : \mathcal{Y} \to \{0,1\}$ such that $\mathbb{P}_\nu[f_{\mathtt{A}}(X) = f_{\mathtt{B}}(Y)] = 1$ and $L := H(f_{\mathtt{A}}(X)) = H(f_{\mathtt{B}}(Y)) > 0$. In such a case, $(0,L) \in \mathcal{T}_r(X,Y)$. It is easy to see that whenever $\Gamma_r^{\mathtt{sk}}(X,Y)$ or $\Gamma_r^{\mathtt{cr}}(X,Y)$ is finite, we have $\Gamma_r^{\mathtt{cr}}(X,Y) = 1 + \Gamma_r^{\mathtt{sk}}(X,Y)$.

Intuitively, the r-round CBIB (KBIB, respectively) can be roughly interpreted as the maximum number of additional bits of common randomness (secret key, respectively) that Alice and Bob can obtain by communicating an additional bit, where the maximum is over "all protocols and any communication rate".

We also remark that it follows from Theorem 4.2 and [GS19, Lemma 5.4] that $\Gamma_r^{\mathtt{cr}}(X,Y)$ is the derivative of the function $\mathscr{C}_r^{\mathtt{am-cr}}(C)$ at $C = 0$.

Next we would like to derive similar separations for the r-round interactive CBIB and KBIB to that in Theorem 4.5 for the r-round MIMK. Notice that from the first item of Theorem 2.3 we have immediately that $\Gamma_{r+2}^{\mathtt{cr}}(X,Y) \geq \frac{n}{(r+2)\lceil \log n\rceil}$. We might hope to use the second and third items of Theorem 2.3 to derive upper bounds on $\Gamma_{\lfloor (r+1)/2\rfloor}^{\mathtt{cr}}(X,Y)$ and $\Gamma_r^{\mathtt{cr}}(X,Y)$ that grow as $\log^{c_0} n$ and $\sqrt{n}\log^{c_0} n$, respectively. However, such upper bounds do not immediately follow from Theorem 2.3 since Theorem 2.3 requires a lower bound on L in order to show that certain tuples (C,L) are not achievable. In particular, Theorem 2.3 leaves open the possibility that tuples such as $(\log n, \sqrt{n})$, or even $(2^{-n}, 1)$ are $\lfloor (r+1)/2\rfloor$-achievable for CRG from $\mu_{r,n,n}$. This limitation of Theorem 2.3 results from the fact that Lemmas 4.1 and 4.2 give vacuous bounds on the disintuishability of $\mu = \mu_{r,n,n}$ and $\mu_X \otimes \mu_Y$ when the tuple (C,L) is such that L is small compared to n. We leave the problem of remedying this issue for future work:

PROBLEM 5.1. *For each $r \in \mathbb{N}$, show (perhaps using Theorem 2.3) that there is a c_0, such that for each $n \geq c_0$, the pointer chasing source $(X,Y) \sim \mu_{r,n,n}$ satisfies:*

1. $\Gamma_{\lfloor (r+1)/2\rfloor}^{\mathtt{cr}}(X,Y) \leq \log^{c_0} n$.
2. $\Gamma_r^{\mathtt{cr}}(X,Y) \leq \sqrt{n}\log^{c_0} n$.

It seems that in fact the even stronger result $\Gamma_{r+1}^{\mathtt{cr}}(X,Y) \leq 1 + o_n(1)$ holds.

Problem 5.1 seems to be quite difficult; a result that $\Gamma_{r'}^{\mathtt{cr}}(X,Y) < f(n)$, for $(X,Y) \sim \mu_{r,n,n}$, some $r' \in \mathbb{N}$, and some function $f(n)$ would imply, since the function $C \mapsto \mathscr{C}_r^{\mathtt{am-cr}}(C)$ is concave ([GS19, Lemma 5.4]), that for any $C \geq 1$, the tuple $(C, f(n)\cdot C)$ is not r'-achievable for CRG from $\mu_{r,n,n}$. For $r' = \lfloor (r+1)/2\rfloor$ and $f(n) = \operatorname{poly}\log(n)$, this would imply part (2) of Theorem 2.3, and for $r' = r$ and $f(n) = \sqrt{n}\operatorname{poly}\log(n)$, this would imply part (3) of Theorem 2.3.

Acknowledgements

We are grateful to Badih Ghazi for useful discussions, to Salil Vadhan for helpful comments on an earlier version of this work, and to anonymous reviewers for helpful suggestions. N.G. would like to thank Venkat Anantharam for an insightful conversation.

References

[AC93] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. part i: secret sharing. *IEEE Transactions on Information Theory*, 39(4), 1993.

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

[AD89a] R. Ahlswede and G. Dueck. Identification in the presence of feedback-a discovery of new capacity formulas. *IEEE Transactions on Information Theory*, 35(1):30–36, January 1989.

[AD89b] R. Ahlswede and G. Dueck. Identification via channels. *IEEE Transactions on Information Theory*, 35(1):15–29, January 1989.

[BBB+92] Charles H Bennett, Franqois Bessette, Gilles Brassard, Louis Salvail, and John Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5:26, 1992.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[BBT60] David Blackwell, Leo Breiman, and A. J. Thomasian. The Capacities of Certain Channel Classes Under Random Coding. *Ann. Math. Statist.*, 31(3):558–567, September 1960.

[BGGS19] Mitali Bafna, Badih Ghazi, Noah Golowich, and Madhu Sudan. Communication-Rounds Tradeoffs for Common Randomness and Secret Key Generation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2019.

[BGI14] Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In *Automata, Languages, and Programming*, pages 150–162. Springer, 2014.

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From Information to Exact Communication. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 151–160, New York, NY, USA, 2013. ACM.

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

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 748–757. IEEE, 2011.

[Bra05] Mark Braverman. On the complexity of real functions. In *46th IEEE Symposium on Foundations of Computer Science*, 2005.

[Bra12] Mark Braverman. Interactive information complexity. In *In Proceedings of the 44th annual ACM Symposium on Theory of Computing, STOC '12*, pages 505–524, 2012.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. *Automata, Languages, and Programming*, 7965:232–243, 2013.

[CGMS17] Clément L Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *IEEE Transactions on Information Theory*, 63(10):6799–6818, 2017.

[CK81] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Academic Press, 1981.

[CMN14] Siu On Chan, Elchanan Mossel, and Joe Neeman. On extracting common random bits from correlated sources on large alphabets. *Information Theory, IEEE Transactions on*, 60(3):1630–1637, 2014.

[CN91] I. Csiszar and P. Narayan. Capacity of the Gaussian arbitrarily varying channel. *IEEE Transactions on Information Theory*, 37(1):18–26, January 1991.

[CN00] Imre Csiszár and Prakash Narayan. Common randomness and secret key generation with a helper. *Information Theory, IEEE Transactions on*, 46(2):344–366, 2000.

[CN04] Imre Csiszár and Prakash Narayan. Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory*, 50(12):3047–3061, 2004.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd IEEE Symposium on Foundations of Computer Science, 2001*, pages 270–278. IEEE, 2001.

[CT12] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[DGS84] Pavol Duris, Zvi Galil, and Georg Schnitger. Lower Bounds on Communication Complexity. In *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, STOC '84, pages 81–91, New York, NY, USA, 1984. ACM.

[DH76] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, November 1976.

[GA10a] A. A. Gohari and V. Anantharam. Information-Theoretic Key Agreement of Multiple Terminals—Part I. *IEEE Transactions on Information Theory*, 56(8):3973–3996, August 2010.

[GA10b] Amin Aminzadeh Gohari and Venkat Anantharam. Information-theoretic Key Agreement of Multiple Terminal: Part II: Channel Model. *IEEE Trans. Inf. Theor.*, 56(8):3997–4010, August 2010.

[GJ18] Badih Ghazi and TS Jayram. Resource-efficient common randomness and secret-key schemes. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1834–1853. Society for Industrial and Applied Mathematics, 2018.

[GK73] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

[GKS16] Badih Ghazi, Pritish Kamath, and Madhu Sudan. Communication complexity of permutation-invariant functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1902–1921, 2016.

[GO13] Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. In *2013 IEEE Conference on Computational Complexity*, pages 287–298. IEEE, 2013.

[GR16] Venkatesan Guruswami and Jaikumar Radhakrishnan. Tight bounds for communication-assisted agreement distillation. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 6:1–6:17, 2016.

[GS17] Badih Ghazi and Madhu Sudan. The Power of Shared Randomness in Uncertain Communication. *arXiv:1705.01082*, May 2017.

[GS19] Noah Golowich and Madhu Sudan. Round complexity of common randomness generation: The amortized setting. *arXiv:1909.00323*, 2019.

[HAD+95] Richard J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. Schauer. Quantum Cryptography. *Contemporary Physics*, 36(3), April 1995.

[Han03] Te Sun Han. *Information-Spectrum Methods in Information Theory*. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin Heidelberg, 2003.

[HJMR07] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The Communication Complexity of Correlation. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 10–23, San Diego, CA, June 2007. IEEE.

[JPY12] Rahul Jain, Attile Pereszlenyi, and Penghui Yao. A direct product theorem for bounded-round public-coin randomized communication complexity. In *2012 IEE Symposium on Foundations of Computer Science*. IEEE, 2012.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A Direct Sum Theorem in Communication Complexity via Message Compression. In Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger, editors, *Automata, Languages and Programming*, Lecture Notes in Computer Science, pages 300–315. Springer Berlin Heidelberg, 2003.

[LCV15] Jingbo Liu, Paul Cuff, and Sergio Verdú. Secret key generation with one communicator and a one-shot converse via hypercontractivity. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 710–714. IEEE, 2015.

[LCV17] Jingbo Liu, Paul W. Cuff, and Sergio Verdú. Secret key generation with limited interaction. *IEEE Transactions on Information Theory*, 63, 2017.

[Liu16] Jingbo Liu. Rate region for interactive key generation and common randomness generation. *Manuscript*, 2016.

[Mau91] Ueli M. Maurer. Perfect cryptographic security from partially independent channels. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing - STOC '91*, pages 561–571, New Orleans, Louisiana, United States, 1991. ACM Press.

[Mau92] UeliM. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *Journal of Cryptology*, 5(1), 1992.

[Mau93] Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993.

[MI08] Nan Ma and Prakash Ishwar. Distributed Source Coding for Interactive Function Computation. *arXiv:0801.0756*, January 2008.

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

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(1):211–219, 1993.

[Orl90] A. Orlitsky. Worst-case interactive communication. I. Two messages are almost optimal. *IEEE Transactions on Information Theory*, 36(5):1111–1126, September 1990.

[Orl91] A. Orlitsky. Worst-case interactive communication. II. Two messages are not optimal. *IEEE Transactions on Information Theory*, 37(4):995–1005, July 1991.

[PS82] Christos H. Papadimitriou and Michael Sipser. Communication Complexity. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, STOC '82, pages 196–200, New York, NY, USA, 1982. ACM. event-place: San Francisco, California, USA.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-key Cryptosystems. *Commun. ACM*, 21(2):120–126, February 1978.

[STW19] Madhu Sudan, Himanshu Tyagi, and Shun Watanabe. Communication for Generating Correlation. *CoRR*, abs/1904.09563, 2019.

[Tya13] Himanshu Tyagi. Common information and secret key capacity. *IEEE Transactions on Information Theory*, 59(9):5627–5640, 2013.

[Wit75] Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975.

[Yan07] Ke Yang. On the (im)possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157–166, August 2007.

[Yao79] Andrew Chi-Chih Yao. Some Complexity Questions Related to Distributive Computing(Preliminary Report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[Ye05] Chunxuan Ye. *Information Theoretic Generation of Multiple Secret Keys*. PhD thesis, University of Maryland, 2005.

[ZC11] Lei Zhao and Yeow-Kiang Chia. The efficiency of common randomness generation. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2011.