Resource-Efficient Common Randomness and Secret-Key Schemes

Badih Ghazi* T.S. Jayram†

Abstract

We study *common randomness* where two parties have access to i.i.d. samples from a known random source, and wish to generate a shared random key using limited (or no) communication with the largest possible probability of agreement. This problem is at the core of secret key generation in cryptography, with connections to communication under uncertainty and locality sensitive hashing. We take the approach of treating correlated sources as a critical resource, and ask whether common randomness can be generated *resource-efficiently*.

We consider two notable sources in this setup arising from correlated bits and correlated Gaussians. We design the first *explicit* schemes that use only a *polynomial* number of samples (in the key length) so that the players can generate shared keys that agree with constant probability using optimal communication. The best previously known schemes were both non-constructive and used an exponential number of samples. In the *amortized* setting, we characterize the largest achievable ratio of key length to communication in terms of the *external* and *internal* information costs, two well-studied quantities in theoretical computer science. In the relaxed setting where the two parties merely wish to improve the *correlation* between the generated keys of length k, we show that there are no interactive protocols using $o(k)$ bits of communication having agreement probability even as small as $2^{-o(k)}$. For the related communication problem where the players wish to compute a joint function f of their inputs using i.i.d samples from a known source, we give a *simultaneous message passing* protocol using $2^{O(c)}$ bits where c is the *interactive* randomized public-coin communication complexity of f. This matches the lower bound shown previously while the best previously known upper bound was doubly exponential in c.

Our schemes reveal a new connection between common randomness and *unbiased error-correcting codes*, e.g., dual-BCH codes and their analogues in Euclidean space.

*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA 02139. badih@mit.edu. Part of the work done while at IBM Research - Almaden.

†IBM Research - Almaden. jayram@us.ibm.com.

1 Introduction

Common randomness plays a fundamental role in various problems of cryptography and information theory. We study this problem in the basic two-party communication setting in which Alice and Bob wish to agree on a (random) key by drawing i.i.d. samples from a known source such as correlated bits or correlated Gaussians. If we further require that an eavesdropper, upon seeing the communication only, gains no information about the shared key, then this defines a *secret key* scheme. This information-theoretic approach to security was introduced in the seminal works of Mauer [33] and Ahlswede and Csiszár [1]. Both common randomness and secret-key generation have been extensively studied in information theory [2, 15, 16, 18, 30, 31, 42, 45, 47]. Common randomness has applications to identification capacity [3] and hardware-based procedures for extracting a unique random ID from process variations [28, 38, 46] that can be used in authentication [28, 40].

Randomness is a powerful tool as well in the algorithm designer's arsenal. Shared keys (aka public randomness) are used crucially in the design of efficient communication protocols with immediate applications to diverse problems in streaming, sketching, data structures and property testing. Common randomness is thus a natural model for studying how shared keys can be generated in settings where it is not available directly [10, 12, 14, 22, 34, 35]. In this paper, we take the approach of treating correlated sources as a critical algorithmic resource, and ask whether common randomness can be generated *efficiently*.[1]

For $-1 \leq \rho \leq 1$, we say that $(X, Y) \sim \mathsf{DSBS}(\rho)$ (*doubly symmetric binary source*) if X, Y are both uniform over $\{\pm 1\}$ and their correlation (and covariance) $\mathbb{E}[XY] = \rho$ (i.e., a *binary symmetric channel* with uniform input). We say that $(X, Y) \sim \mathsf{BGS}(\rho)$ (*bivariate Gaussian source*) if $X, Y \sim \mathcal{N}(0, 1)$, the standard normal distribution, and their correlation is again ρ.[2]

[1] Notably, the schemes that we design can also be easily transformed into secret key schemes, as shown later.

[2] Note that $\mathsf{BGS}(\rho)$ is uniquely defined. To see this, note that any linear combination of X and Y is also Gaussian; thus, (X, Y) has a multivariate Gaussian distribution and is hence characterized by its first and second moments.

Bogdanov and Mossel [10] gave a common randomness scheme for $\mathsf{DSBS}(\rho)$ with zero-communication to generate k-bit keys that agree with probability $2^{-\frac{1-\rho}{1+\rho}\cdot k}$, up to lower order inverse $\mathsf{poly}(k, 1-\rho)$ factors (which we suppress henceforth). Using the hypercontractive properties of the noise operator [9, 11], they also proved the "converse" that the bound for agreement (probability) is essentially the best possible. In followup work, Guruswami and Radhakrishnan [22] recently gave a one-way scheme that achieves an *optimal* tradeoff between communication and agreement.[3] Note that a simple scheme in which Alice just sends her input requires $k - O_\rho(1)$ bits of communication for constant agreement. In contrast, their scheme can guarantee the same agreement using only $(1-\rho^2)\cdot k$ bits of communication. This is a nontrivial *amortized* bound since for $\rho > 0$, the ratio of entropy to communication $(=1/(1-\rho^2))$ is strictly bounded away from 1 as $k \to \infty$. On the other hand, the above schemes are *non-explicit* (i.e., proved using the probabilistic method) and use an *exponential* number of samples in k. Bogdanov and Mossel [10] asked whether an *explicit* and *efficient* scheme can be designed, motivating the definition below.

We say that a common randomness scheme to generate k-bit keys (with k as input) is *resource-efficient*, if it (i) is explicitly[4] defined, (ii) uses $\mathsf{poly}(k)$ samples, (iii) has constant agreement probability, and (iv) achieves an amortized ratio of entropy to communication bounded away from 1. We give the *first* efficient scheme for correlated bits and Gaussians, answering the question of [10].

THEOREM 1.1. *There exist resource-efficient one-way common randomness schemes for* $\mathsf{DSBS}(\rho)$ *and* $\mathsf{BGS}(\rho)$ *using* $(1-\rho^2)\cdot k$ *bits of communication. For zero-communication, there exist explicit schemes for* $\mathsf{DSBS}(\rho)$ *and* $\mathsf{BGS}(\rho)$ *using* $\mathsf{poly}(k)$ *samples with agreement probability* $2^{-\frac{1-\rho}{1+\rho}\cdot k}$*, up to polynomial factors.*

More generally, we obtain one-way schemes with *optimal* tradeoff between communication and agreement, matching [22], while using only $\mathsf{poly}(k)$ samples. Below is the formal statement.

THEOREM 1.2. *Let* $0 < \rho < 1$ *and* $0 \le \delta \le \sqrt{\frac{1-\rho}{1+\rho}}$ *be arbitrary. Set* $\varphi = \rho + \delta\sqrt{1-\rho^2}$*. Then there exist explicit one-way common randomness schemes for* $\mathsf{DSBS}(\rho)$ *and* $\mathsf{BGS}(\rho)$ *using* $\mathsf{poly}(k)$ *samples such that:*

1. *the entropy of the key is at least* $k - o(k)$*;*[5]
2. *the agreement probability is at least* $2^{-\delta^2 k}$*, up to polynomial factors; and*
3. *the communication is* $O((1-\varphi^2)\cdot k)$ *bits.*

We point out that our schemes are resource efficient but *computationally inefficient*. One representative challenge that arises here is in decoding dual-BCH codes, which are an explicit algebraic family of error-correcting codes, from a *very large* number of errors.

The above schemes follow a template that generalizes the approach taken by [10, 22]. It relies on a carefully constructed codebook $\mathcal{C} \subseteq \mathbb{R}^n$ of size 2^k, where n is the number of samples. Alice outputs the codeword in $\mathcal{C}$ with the largest projection while Bob does the same on a subcode of $\mathcal{C}$ based on Alice's message. The analysis of the template reduces it to the problem of obtaining good tail bounds on the joint distribution induced by these projections. For $\mathsf{BGS}(\rho)$, we use a codebook consisting of an explicitly defined large family of nearly-orthogonal vectors in $\mathbb{R}^n$ due to Tao [41], who showed their near-orthogonality property using the Weil bound for curves. The novel part of the analysis involves getting precise conditional probability tail bounds on trivariate Gaussians induced by the projections, whose covariance matrix has a special structure. Standard methods only give asymptotic bounds on such tails which is inadequate in the low-communication regime. Here, the best possible agreement is exponentially small in k. Our analysis determines the exact constant in the exponent by carefully evaluating the underlying triple integrals.

The resource-efficient scheme for $\mathsf{DSBS}(\rho)$ is based on Dual-BCH codes that can be seen as an $\mathbb{F}_2$-analogue of Tao's construction. The Weil bound for curves implies that dual-BCH codes are "unbiased", in the sense that any two distinct codewords are at distance $\approx n/2$ (with n being the block length)[6]. Analogous to the Gaussian case, the analysis involves getting precise bounds on the (conditional) tail probabilities of various correlated binomial sums. Since $n = \mathsf{poly}(k)$, we cannot handle these binomial sums using the (two-dimensional) Berry-Esseen theorem, since the incurred additive error of $1/\sqrt{n}$ would overwhelm target agreement. Moreover, crude concentration and anti-concentration bounds cannot be used since they do not determine the exact constant in the exponent. We directly handle these correlated binomial sums, which turns out to involve some

[3] They also use hypercontractivity to prove the converse, which extends to other sources including $\mathsf{BGS}(\rho)$.

[4] By an *explicit* scheme, we mean that its existence is not proved using the probabilistic method.

[5] We follow [22] who actually consider the *min-entropy* of Alice's output, which is justifiable on technical grounds. For more details on this choice, we refer the reader to the full version of the paper [20].

[6] For more on unbiased codes, we refer the reader to the work of Kopparty and Saraf [27].

tedious calculations related to the binary entropy function.

Interactive Common Randomness and Information Complexity. Ahlswede and Csiszár [1, 2] studied common randomness in their seminal work using an *amortized* communication model. They defined it as the maximum achievable ratio a/c, such that for every large enough number of samples n, Alice and Bob can agree on a key of $a \cdot n$ bits using $c \cdot n$ bits of communication, where the agreement probability tends to 1 (as n tends to infinity). This more stringent linear relationship between the quantities is not obeyed by our explicit schemes. For one-way communication, they characterized this ratio in terms of the *Strong Data Processing Constant* of the source, which is intimately related to its hypercontractive properties [4,5]. More recently, Liu, Cuff and Verdu [29–31] extended this beyond one-way communication. In particular, [29] derives the "rate region" for r-round amortized common randomness.

In this work, we show that r-round amortized common randomness can be alternatively characterized in terms of two well-studied notions in theoretical computer science: the *internal* and *external* information costs of communication protocols. Recall that the internal information cost [6, 7] of a two-party randomized protocol is the total amount of information that each of the two players learns about the other player's input, whereas its external information cost [13] is the amount of information that an external observer learns about the inputs (see Section 5 for formal definitions). These measures have been extensively studied within the context of communication complexity. While being interesting measures in their own rights, they have also been the central tool in tackling direct-sum problems, with numerous applications, e.g., in data streams and distributed computation.

THEOREM 1.3. (INFORMAL STATEMENT) *Given an arbitrary distribution μ, let Γ_r denote the supremum over all r-round randomized protocols P of the ratio of the external information cost to the internal information cost of P with respect to μ. Then, for r-round amortized common randomness, Γ_r equals the largest achievable ratio H/R such that using μ as the source, for every large enough n, Alice and Bob can agree on a key of $H \cdot n - O(1)$ bits with probability $1 - o_n(1)$ using r rounds and $R \cdot n + O(1)$ bits of communication.*

For the proof, we use a direct-sum approach, a classical staple of information complexity arguments. Our setup is slightly different from the known direct-sum results because we need to lower bound the internal information cost of the n-input protocol as well as upper bound its external information cost (which is non-standard) *simultaneously*. The essential ingredients are the same: embed the input on a judiciously chosen coordinate but the argument works an round-by-round basis so as to keep the mutual information expressions intact. To prove the other direction, we use the rate region of [29, 31] to get a lower bound on Γ_r.

Finally, we outline various settings where common randomness plays an important role.

Secret Key Generation: While secret key generation requires common randomness, in the amortized setting they are known to imply each other [29, 31]: the rate pair (H, R), using the notation of Theorem 1.3, is achievable for common randomness if and only if $(H - R, R)$ is achievable for secret key generation. In particular, using the Strong Data Processing Constant for $\mathsf{DSBS}(\rho)$, the rate ratio $H/R = 1/(1-\rho^2)$ is achievable for common randomness and the rate ratio $\rho^2/(1-\rho^2)$ for secret key generation, but using non-explicit schemes. Our resource-efficient but non-amortized schemes given in Theorem 1.1 can be easily transformed into secret key schemes. See the remark at the end of Section 3.

General Sources: Theorem 1.2 also implies an explicit scheme for an *arbitrary* source μ in terms of its *maximal correlation* $\rho(\mu)$ [19, 23, 36]. For $(X, Y) \sim \mu$, recall that $\rho(\mu) \coloneqq \sup \mathbb{E}\, F(X)G(Y)$ over all real-valued functions F and G with $\mathbb{E}\, F(X) = \mathbb{E}\, G(Y) = 0$ and $\operatorname{Var} F(X) = \operatorname{Var} G(Y) = 1$. This uses the idea (implicit in [44]) that given i.i.d. samples from any source of maximal correlation ρ, there is a explicit strategy via CLT that allows Alice and Bob to use these samples in order to generate standard ρ-correlated Gaussians. The resulting scheme however is *not* resource-efficient.

Correlated Randomness Generation: In this relaxation proposed by [12], Alice and Bob are given access to $\mathsf{DSBS}(\rho)$ and wish to generate k bits that are jointly distributed i.i.d. according to $\mathsf{DSBS}(\rho')$ where $\rho < \rho'$? Note that the $\rho' = 1$ corresponds to the the common randomness setup studied above. We partially answer a question of [12] that even a modest improvement in the correlation requires substantial communication. Let $\varepsilon' \log(1/\epsilon') \ll \varepsilon < \frac{1}{2}$ be fixed. We show that for Alice and Bob to produce k samples according to $\mathsf{DSBS}(1-2\varepsilon')$ using $\mathsf{DSBS}(1-2\varepsilon)$ as the source requires $\Omega(\varepsilon \cdot k)$ bits of communication (even for interactive protocols and even when the agreement probability is as small as $2^{-o(k)}$). See Section 9 for a detailed description.

Communication with Imperfect Shared Randomness: In this framework [8, 12] (see also [21]), Alice and Bob wish to compute a joint function of their inputs and

have access to i.i.d samples from a known source. For example, with $\mathsf{DSBS}(\rho)$ this setup interpolates between the well-studied public randomness ($\rho = 1$) and private randomness ($\rho = 0$) models. Communication complexity lower bounds for imperfect shared randomness give one approach to rule out low-communication common randomness schemes. In particular, [8] exhibit a (partial) function whose zero-communication complexity using $\mathsf{DSBS}(\rho)$ for all $\rho < 1$ is exponentially larger than the one using public randomness. We prove that this separation is tight. We show a stronger result that every function having *interactive* communication c bits using public randomness has a zero-communication protocol with $2^{O(c)}$ bits using $\mathsf{DSBS}(\rho)$ for every $\rho < 1$. This answers a question of Sudan [39]. See Section 10 for a detailed description.

Locality Sensitive Hashing (LSH): A surprising "universality" feature of our schemes (as well as previous ones) for $\mathsf{DSBS}(\rho)$ and $\mathsf{BGS}(\rho)$ using zero-communication is that their definition is oblivious to ρ; only the analysis for every fixed ρ shows that they have near-optimal agreement. This has a close resemblance to schemes used in LSH. Indeed, we show that our common randomness scheme leads to an improvement in the "$\bar{\rho}$-parameter" [24] that governs one aspect of the performance of an LSH scheme. While this is mathematically interesting, we caution the reader that this does not lead to better nearest-neighbor data structures since the improvement is only qualitatively better and our scheme is computationally inefficient. See Section 11.

Organization. Section 2 describes the template used for the one-way schemes and sets up the structure of the analysis. Section 3 and Section 4 describe the schemes for $\mathsf{BGS}(\rho)$ and $\mathsf{DSBS}(\rho)$ and their analysis. In Section 5, we show the connection between amortized common randomness and information complexity. In Section 6, we conclude with some very intriguing open questions.

1.1 Preliminaries

Notation. For a tuple $U = (U_1, U_2, \ldots, U_n)$, let $U_i^j := (U_i, U_{i+1}, \ldots, U_j)$, when $1 \le i \le j \le n$, and empty otherwise; we may drop the subscript when $i = 1$. For a distribution μ, let $\mu^{\otimes n}$ be obtained by taking i.i.d. samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ from μ. Abusing notation, we say that $(X^n, Y^n) \sim \mu^{\otimes n}$. Let $\bullet$ denote the standard inner product and let $\|\cdot\|$ denote the Euclidean norm over $\mathbb{R}$. For any positive integer n, let $[n] := \{1, \ldots, n\}$. Let $a \lesssim b$ denote $a \le Cb$ for some positive global constant C.

Bivariate Gaussians. Let $(X, Y) \sim \mathsf{BGS}(\rho)$. Let $Q(t) := \Pr[X > t]$ denote the Gaussian tail probability and $L(t, \varphi; \rho) := \Pr[X > t, Y > \varphi t]$ denote the (asymmetric) orthant probability. In Section 7, we prove the following, which also uses some seemingly new properties of $Q(t)$.

PROPOSITION 1.1. *Let $t, \delta \ge 0$. Set $\varphi := \rho + \delta\sqrt{1-\rho^2}$ and $\lambda_0 := \sqrt{\frac{2}{\pi}}$. Then:*

$$(a)\ \frac{e^{-t^2/2}}{t+\lambda_0} \lesssim Q(t) \lesssim \frac{e^{-t^2/2}}{t+1/\lambda_0} \le e^{-t^2/2};$$

$$(b)\ \frac{Q(t)^{\delta^2}}{\delta t+\lambda_0} \lesssim Q(\delta t) \lesssim Q(t)^{\delta^2}(t+\lambda_0)^{c^2};$$

$$(c)\ L(t, \varphi; \rho) \ge Q(t)Q(\delta t); \quad \textit{and}$$

$$(d)\ Q(t) \le Q(\delta t) \le Q(t)^{\delta^2}, \quad \textit{if} \quad \delta \le 1$$

PROPOSITION 1.2. (ELLIPTICAL SYMMETRY) *Let $(X, Y) \sim \mathsf{BGS}(\rho)^{\otimes n}$ and $v, w \in \mathbb{R}^n$ have unit norm. Then, $(v \bullet X, w \bullet Y) \sim \mathsf{BGS}\big(\rho(v \bullet w)\big)$.*

2 Template One-Way Scheme and its Analysis

The one-way schemes (including zero-communication as a special case) have the following template. Let μ denote the source on $\mathbb{R} \times \mathbb{R}$. Alice and Bob will generate n iid samples from μ and use them to output k bit keys. This is achieved by the players using a special codebook $\mathcal{C}$ of 2^k points in $\mathbb{R}^n$ where each codeword $u \in \mathcal{C}$ corresponds to a k-bit message $D(v)$, where $D : \mathcal{C} \to \{0,1\}^k$. For $c \ge 1$, the players also agree on a coloring χ of $\mathcal{C}$ using 2^c colors such that each color class has size at most $|\mathcal{C}| \cdot 2^{-c} + 1$. In addition, let $\diamond$ denote an auxiliary color. Thus, each color can be specified using $c + 1$ bits. For the special case of zero-communication, we assume wlog that all codewords are colored $\diamond$ and we set $c = 0$.

Let t and s be parameters that govern the achievable min-entropy and agreement probability. Let κ_A and κ_B be any explicit mappings such that $\kappa_A(X)$ and $\kappa_B(Y)$ are each uniformly distributed over $\{0,1\}^k$.

The pseudocode is given in Algorithm 1. For the analysis, define the following quantities:

1. **Univariate tail:** $\mathcal{U} := \max_{v \in \mathcal{C}} \Pr[v \bullet X > t]$; 2. **Bivariate tail:** $\mathcal{B} := \min_{v \in \mathcal{C}} \Pr[v \bullet X > t, v \bullet Y > s]$

3. **Conditional trivariate tails:**
(a) $\mathcal{T}_A := \max_{v \ne w \in \mathcal{C}} \Pr[w \bullet X > t \mid v \bullet X > t, v \bullet Y > s]$ and
(b) $\mathcal{T}_B := \max_{v \ne w \in \mathcal{C}} \Pr[w \bullet Y > s \mid v \bullet X > t, v \bullet Y > s]$

THEOREM 2.1. *The min-entropy of the basic scheme is at least $-\log(\mathcal{U} + 2^{-k})$. Assume that $|\mathcal{C}| \cdot \mathcal{T}_A \le \frac{1}{4}$ and $|\mathcal{C}| \cdot \mathcal{T}_B \le \frac{1}{4} \cdot 2^c$. Then the probability of agreement is at least $\frac{1}{2}|\mathcal{C}| \cdot \mathcal{B}$.*

Algorithm 1 One-way scheme for source μ

procedure CR(k;μ) ▷ Generate k-bit common random key using source μ.
  Let $(X, Y) \sim \mu^{\otimes n}$. ▷ Alice gets X and Bob gets Y.
  if $\exists$ unique $v \in \mathcal{C}$ such that $v \bullet X > t$ **then** Alice outputs $D(v)$ and sends $\chi(v)$.
  else Alice outputs $\kappa_A(X)$ and sends $\diamond$.
  Bob receives the color τ.
  if $\exists$ unique $w \in \mathcal{C}$ such that $\chi(w) = \tau$ and $w \bullet Y > s$ **then** Bob outputs $D(w)$.
  else Bob outputs $\kappa_B(Y)$.

Proof. If Alice outputs $a \in \{0,1\}^k$ then either there exists a unique $v \in \mathcal{C}$ such that $D(v) = a$ and $v \bullet X > t$, which happens with probability at most $\Pr[v \bullet X > t] \le \mathcal{U}$, or $\kappa_A(X) = a$, which happens with probability 2^{-k}. The min-entropy guarantee follows.

For the agreement, fix $v \in \mathcal{C}$. Define event $E_v := A_v \wedge \overline{B_v} \wedge \overline{C_v}$ where $A_v := \{v \bullet X > t \wedge v \bullet Y > s\}$, $B_v := \{\exists w \neq v : w \bullet X > t\}$, and $C_v := \{\exists w \neq v : \chi(v) = \chi(w) \wedge w \bullet Y > s\}$.

Note that the event E_v ensures that both players output $D(v)$. By the union bound:

$$\begin{aligned}\Pr[E_v] &\ge \Pr[A_v] \cdot \big(1 - \Pr[B_v \vee C_v \mid A_v]\big) \\ &\ge \Pr[A_v] \cdot \Big(1 - \sum_{w \neq v} \Pr[w \bullet X > t \mid A_v] \\ &\qquad - \sum_{w \neq v} 1\{\chi(w) = \chi(v)\} \cdot \Pr[w \bullet Y > s \mid A_v]\Big) \\ &\ge \mathcal{B}\big(1 - |\mathcal{C}| \cdot \mathcal{T}_A - |\mathcal{C}| \cdot 2^{-c} \cdot \mathcal{T}_B\big) \;\ge\; \tfrac{1}{2}\mathcal{B},\end{aligned}$$

where the last two inequalities follow from the definition of $\mathcal{B}$ and $\mathcal{T}$ and then invoking the premise of the lemma. Thus the agreement probability is at least $\sum_v \Pr[E_v] \ge \frac{1}{2}|\mathcal{C}| \cdot \mathcal{B}$.

As an illustration, we present an explicit one-way scheme for the $\mathsf{BGS}(\rho)$ using an exponential number of samples. Let k be a large enough constant and let $n = 2^k$. Let $\mathcal{C}$ consist of the n standard basis vectors $\{e_i : i \in [n]\}$ in $\mathbb{R}^n$. Choose $t > 0$ so that the Gaussian tail probability $Q(t) = \frac{1}{4} \cdot 2^{-k}$. Let $\rho \le \varphi \le 1$ be arbitrary and set $s = \varphi t$. (Choose $\varphi = 1$ for zero-communication.)

For the analysis, note that for each i, we have $e_i \bullet X = X_i$ and $e_i \bullet Y = Y_i$. Therefore $\Pr[X_i > t] = Q(t)$ and so by Theorem 2.1, the min-entropy of Alice's output is at least $-\log(Q(t) + 2^{-k}) \ge k - 1$.

We now analyze the agreement probability. To bound the bivariate tail, first by Proposition 1.1 (a), we have $t = \Theta(\sqrt{k})$. Let δ satisfy $\varphi = \rho + \delta\sqrt{1-\rho^2}$. Observe that $0 \le \delta \le 1$. Applying Proposition 1.1 (b,c), we obtain:

$$\begin{aligned}\mathcal{B} &= \min_{i \in [n]} \Pr[X_i > t, Y_i > \varphi t] = L(t, \varphi; \rho) \\ &\gtrsim \frac{Q(t)^{1+\delta^2}}{\delta t + \Theta(1)} \gtrsim \frac{Q(t)^{1+\delta^2}}{\delta\sqrt{k} + \Theta(1)} \qquad (2.1)\end{aligned}$$

For $i \neq j$, the trivariate tail probability $\Pr[X_j > t \mid X_i > t, Y_i > \varphi t] = \Pr[X_j > t] = Q(t)$, by independence of components of (X, Y). Similarly, $\Pr[Y_j > \varphi t \mid X_i > t, Y_i > \varphi t] = Q(\varphi t)$. Therefore:

$$\mathcal{T}_A \le Q(t) \qquad \text{and} \qquad \mathcal{T}_B \le Q(\varphi t) \qquad (2.2)$$

Now $Q(t) = \frac{1}{4} \cdot 2^{-k}$, so $|\mathcal{C}| \cdot \mathcal{T}_A \le \frac{1}{4}$. Next, $Q(\varphi t) \le Q(t)^{\varphi^2}$, using Proposition 1.1 (d). Therefore, $\mathcal{T}_B \le Q(t)^{\varphi^2}$. If we choose $c \ge (1 - \varphi^2)(k + 2)$, then it can be verified that $|\mathcal{C}| \cdot \mathcal{T}_B \le \frac{1}{4} \cdot 2^c$. This ensures that the conditions of Theorem 2.1 for agreement are satisfied.

By Theorem 2.1, the agreement probability is $\frac{1}{2}|\mathcal{C}| \cdot \mathcal{B} \gtrsim 2^{-\delta^2 k}/(\delta\sqrt{k} + \Theta(1))$ and the scheme uses $O((1 - \varphi^2)k)$ bits of communication. In particular, set $\varphi = \rho$ and $\delta = 0$; we obtain an explicit one-way scheme with constant probability and $O((1 - \rho^2)k)$ bits of communication.

3 Efficient Scheme for $\mathsf{BGS}(\rho)$

In this section, we give a resource-efficient one-way scheme for $\mathsf{BGS}(\rho)$ with optimal communication $(1 - \rho^2)k$ bits. More generally, the tradeoff between the communication and agreement probability is similar to the one obtained with the scheme presented in Section 2.

The analysis of the template given previously suggests the following scheme to reduce the sample complexity to $k = \mathsf{poly}(n)$: use a codebook such that the projections are only 3-wise independent. Unfortunately, this does not work since a multivariate Gaussian distribution is completely characterized by its first and second moments, so even pairwise independence would imply full independence! Instead, we use a codebook consisting of an explicitly defined large family of nearly-orthogonal vectors in $\mathbb{R}^n$ due to Tao [41], who showed their near-orthogonality property using the Weil bound for curves.

Let p be a prime number and $n = 2 \cdot p$. We identify $\mathbb{R}^n$ with the complex vector space $\mathcal{V}$ of functions from $\mathbb{F}_p$ to $\mathbb{C}$, where $\mathbb{C}$ denotes the complex plane. Thus $v \in \mathcal{V}$ will also denote an element of $\mathbb{R}^n$. With this identification, we have $v \bullet w = \mathfrak{Re}\big(\sum_{x \in \mathbb{F}_p} v(x)\overline{w(x)}\big)$ for $v, w \in \mathcal{V}$.

Let d be a positive integer. Let $\omega := e^{2\pi i/p}$ denote the p-th root of unity. For every $a \in \mathbb{F}_p^d$, let $v_a \in \mathcal{V}$

be defined as $v_a(x) = \frac{1}{\sqrt{p}} \cdot \omega^{a_d x^d + \cdots + a_1 x}$. We set $\mathcal{C} \coloneqq \{v_a : a \in \mathbb{F}_p^d\}$. Note that the all elements of $\mathcal{C}$ have unit norm. The Weil bound for curves then implies that for every $a \neq b \in \mathbb{F}_p^d$, we have that $|v_a \bullet v_b| \leq (d-1)/\sqrt{p}$ [43] (for a recent exposition see [26]).

Choose $d = o(n^{1/4}/\sqrt{\log n})$ and $k = d \cdot \log(n/2)$ in Tao's construction. We use the same parameters t, s, φ and δ for Algorithm 1 as in the previous scheme described in Section 2.

By elliptical symmetry (Proposition 1.2), $(v \bullet X, v \bullet Y) \sim \mathsf{BGS}(\rho)$, for every $v \in \mathcal{C}$. Therefore the bounds in Section 2 for the univariate and bivariate tails (see eq. (2.1)) also hold here. The key difference is in the analysis of the trivariate probabilities because we no longer have independence amongst the various pairs $(v \bullet X, v \bullet Y)$. This requires a new analysis of the conditional tails involving trivariate Gaussians whose covariances have a special structure. Below, we show that a slightly weaker bound than eq. (2.2): $\mathcal{T}_A \leq Q(t) \cdot (1 + o_n(1))$ and $\mathcal{T}_B \leq Q(\varphi t) \cdot (1 + o_n(1))$. Nevertheless, we can apply the same argument following eq. (2.2) in Section 2 which implies again that: (a) the min-entropy at least $k - 1$; (b) the agreement probability is $\gtrsim 2^{-\delta^2 k}/(\delta\sqrt{k} + \Theta(1))$; and (c) the communication is $O((1-\varphi^2)k)$ bits. In particular, with $\varphi = \rho$ and $\delta = 0$; we obtain the main result of this section, namely a resource-efficient one-way scheme using $O((1-\rho^2) \cdot k)$ bits of communication.

It remains to prove that $\mathcal{T}_A \leq Q(t) \cdot (1 + o_n(1))$ and $\mathcal{T}_B \leq Q(\varphi t) \cdot (1 + o_n(1))$. Fix $v \neq w \in \mathcal{C}$. The construction ensures that $|v \bullet w| \leq \theta$ with $\theta = (d-1)/\sqrt{p} = O(k/(\sqrt{n} \cdot \log n))$. For $k = o(n^{1/4} \cdot \sqrt{\log n})$, we have $\theta = o_n(1)$.

Now observe that $(w \bullet X, v \bullet X, v \bullet Y)$ can be written as a linear transform on (X, Y), so jointly they have the trivariate Gaussian distribution, Their joint distribution is fully given by the first two moments. By stability, the marginals are standard normal and by elliptical symmetry, the covariances can be calculated as (i) $\mathbb{E}[(w \bullet X)(v \bullet X)] = w \bullet v \leq \theta$, (ii) $\mathbb{E}[(v \bullet X)(v \bullet Y)] = \rho$, and (iii) $\mathbb{E}[(w \bullet X)(v \bullet Y)] = \rho(w \bullet v) \leq \rho\theta$. Observe that $(w \bullet Y, v \bullet Y, v \bullet X)$ is also trivariate with an identical mean and covariance matrix.

LEMMA 3.1. *Let (U, V, W) be a trivariate Gaussian with standard normal marginals and covariances $\mathbb{E}[UV] = \sigma$, $\mathbb{E}[VW] = \rho$, and $\mathbb{E}[UW] = \sigma\rho$. Let $r, r' \geq 0$. Then for all $b \geq 1$:*

$$\Pr[U > r \mid V > r, W > r'] \leq Q\left(\frac{1 - b\sigma}{\sqrt{1-\sigma^2}} r\right) + \frac{Q(br)}{\Pr[V > r, W > r']}$$

Proof. We have:

$$\Pr[U > r \mid V > r, W > r'] = \frac{\Pr[U > r, V > r, W > r']}{\Pr[V > r, W > r']} \tag{3.3}$$

For the numerator, we split the range of V into two intervals:

$$\begin{aligned}&\Pr[U > r, V > r, W > r'] \\ &= \Pr[U > r, r < V \leq br, W > r'] + \Pr[U > r, V > br, W > r']\end{aligned}$$

The second term is at most $\Pr[V > br] = Q(br)$. For the first term, note that the covariance structure implies that U and W are independent conditioned on V, so we can write $U = \sigma V + \sqrt{1-\sigma^2} Z$. where $Z \sim \mathcal{N}(0, 1)$ is independent of (V, W). The event $\{U > r\}$ can be rewritten as $\left\{Z > \frac{r - \sigma V}{\sqrt{1-\sigma^2}}\right\}$ which under the assumption $\{V \leq br\}$ implies that $\{Z > ar\}$ where $a \coloneqq \frac{1 - b\sigma}{\sqrt{1-\sigma^2}}$. By independence:

$$\begin{aligned}&\Pr[U > r, r < V \leq br, W > r'] \\ &\leq \Pr[Z > ar] \Pr[r < V \leq br, W > r'] \\ &\leq Q(ar) \Pr[V > r, W > r']\end{aligned}$$

Substituting these bounds in eq. (3.3) finishes the proof.

Apply Lemma 3.1 to the triples $(w \bullet X, v \bullet X, v \bullet Y)$ with $r \coloneqq t$, $r' \coloneqq \varphi t$ and $(w \bullet Y, v \bullet Y, v \bullet X)$ with $r \coloneqq \varphi t$, $r' \coloneqq t$. In both cases, $\sigma \coloneqq v \bullet w \leq \theta$. Since $Q(\cdot)$ is decreasing:

$$\begin{aligned}&\Pr[w \bullet X > t \mid v \bullet X > t, v \bullet Y > \varphi t] \\ &\leq Q\left(\frac{1 - b\theta}{\sqrt{1-\theta^2}} t\right) + \frac{Q(bt)}{L(t, \varphi; \rho)}, \quad \forall b \geq 1\end{aligned} \tag{3.4}$$

$$\begin{aligned}&\Pr[w \bullet Y > \varphi t \mid v \bullet Y > \varphi t, v \bullet X > t] \\ &\leq Q\left(\frac{1 - b\theta}{\sqrt{1-\theta^2}} \varphi t\right) + \frac{Q(b\varphi t)}{L(t, \varphi; \rho)}, \quad \forall b \geq 1\end{aligned} \tag{3.5}$$

Set $b \coloneqq 2/\phi$. By Proposition 1.1 (a), $Q(b\varphi t) \lesssim e^{-b^2\varphi^2 t^2/2} = e^{-2t^2}$ and $Q(t) \gtrsim e^{-t^2/2}/(t + \lambda_0)$. Because $t = \Theta(\sqrt{k})$, for large enough k, we have $Q(b\varphi t) \lesssim Q(t)^3 e^{-t^2/2}(t + \lambda_0)^3 = Q(t)^3 o_n(1)$.

Using this bound and Proposition 1.1 (c,d), we obtain:

$$\begin{aligned}Q(bt) \leq Q(b\varphi t) &\leq Q(t)^3 \cdot o_n(1) \leq L(t, \varphi; \rho) Q(t) \cdot o_n(1) \\ &\leq L(t, \varphi; \rho) Q(\varphi t) \cdot o_n(1)\end{aligned}$$

Thus the second term in eq. (3.4) (resp. eq. (3.5)) is at most $Q(t) \cdot o_n(1)$ (resp. $Q(\varphi t) \cdot o_n(1)$).

For the first terms in the right side of eq. (3.4) and eq. (3.5), let $a \coloneqq \frac{1 - b\theta}{\sqrt{1-\theta^2}}$. Note that $a \leq 1$. Now

$Q(at) \le Q(t)^{a^2}$ by Proposition 1.1 (d). We calculate $1-a^2 = \left(\frac{2b-(1+b^2)\theta}{1-\theta^2}\right)\theta \le 4b\theta$, since $\theta \ll 2b/(1+b^2)$. For the choice of d we have $kb\theta = o_n(1)$. Thus $Q(at)/Q(t) \le Q(t)^{a^2-1} \lesssim 2^{k(1-a^2)} \le 2^{4kb\theta} = 2^{o_n(1)} = 1+o_n(1)$.

By Lemma 7.2, $Q(at)/Q(t)$ is increasing in t, so $Q(a\varphi t)/Q(\varphi t) \le Q(at)/Q(t) \le Q(at)/Q(t) = 1+o_n(1)$. Thus the first term in eq. (3.4) (resp. eq. (3.5)) is at most $Q(t)\cdot(1+o_n(1))$ (resp. $Q(\varphi t)\cdot(1+o_n(1))$). Combine the above bounds for the two terms in eqs. (3.4) and (3.5) to complete the analysis. This completes the proofs of Theorems 1.1 and 1.2 for the BGS source.

Remark We modify the above resource-efficient scheme that uses $c = (1-\rho^2)\cdot k$ bits of communication to generate secret keys. Assume wlog that codewords within the same color class are encoded with the same prefix of c bits. Now Alice just outputs the $k-c = \rho^2\cdot k$-bit suffix of her output. We briefly sketch the analysis as follows. Using the min-entropy property as well as a similar lower bound on the probability that Alice outputs a particular key (which essentially follows from the same bounds on bivariate tails used above) it can be shown that the communicated bits are nearly uniform as well and that the suffix of the output is nearly uncorrelated with the prefix. This ensures the secrecy of the key from the eavesdropper.

4 Efficient Scheme for DSBS(ρ)

We give a resource-efficient one-way scheme for DSBS(ρ) with optimal communication $(1-\rho^2)\cdot k$ using the template of Algorithm 1. It is based on dual-BCH codes which can be seen as finite field analogues of the nearly-orthogonal vectors used in Section 3. It is more natural here but still equivalent to work with $\{0,1\}^n$ instead of $\{\pm1\}^n$ and the Hamming distance Δ instead of inner-product.

Let $\mathcal{C}_{dBCH} = \mathcal{C}_{dBCH(d,m)}$ be the dual-BCH code with parameters $m = \log(n+1)$ and d being any polynomial in n that satisfies $d = o(n^{1/4}/\sqrt{\log n})$. Then, $|\mathcal{C}_{dBCH}| = 2^k$ where $k = d\cdot\log(n+1)$ is a polynomial in n. Let $\mathcal{C}$ be an arbitrary subset of $\mathcal{C}_{dBCH}$ of size $2^{k'} = 2^k/(\gamma\cdot n)$ where $\gamma > 0$ is a sufficiently large absolute constant to be chosen later on. We denote $\mathcal{C} = \{v_a : a \in \{0,1\}^{k'}\}$. We set $r \triangleq n/2 - t\sqrt{n}/2$ where $t > 0$ satisfies $Q(t) = (1/4)\cdot 2^{-k}$. Similar to before, let $\rho \le \varphi \le 1$ so that the communication is $O((1-\varphi^2)k)$ bits. Recall that δ satisfies $\varphi = \rho + \delta\sqrt{1-\rho^2}$.

In the following, we prove the appropriate uni-, bi- and trivariate tail bounds for the scheme. These are stated in Proposition 4.1 and Lemmas 4.1 and 4.2. The proof follows the same structure that was used for BGS(ρ). It requires some bounds on binomial sums proved in Section 8. By incorporating them into Theorem 2.1, we obtain the desired performance of the scheme. Let $(X,Y) \sim \mathsf{DSBS}(\rho)^{\otimes n}$.

PROPOSITION 4.1. *For any $u \in \mathbb{R}$ (possibly depending on n), $\Pr[|\mathsf{wt}(X) - n/2| \ge u\sqrt{n}/2] \le \mathsf{poly}(n)\cdot Q(u)$.*

LEMMA 4.1. *For every $a \in \{0,1\}^{k'}$: $\Pr[\Delta(v_a, X) \le r, \Delta(v_a, Y) \le r'] \ge \frac{1}{\Theta(n^2)}\cdot 2^{-k}\cdot 2^{-k\cdot\delta^2}$.*

Proof. Follows from Proposition 8.1 and Proposition 8.2.

LEMMA 4.2. *Let $v, v' \in \{0,1\}^n$ satisfy $|\Delta(v,v') - n/2| \le \theta\cdot n/2$, where $\theta = O(k/(\sqrt{n}\cdot\log n))$. Then:*

$$\Pr[\Delta(v',X) \le r \mid \Delta(v,X) \le r, \Delta(v,Y) \le r'] \le O(n)\cdot Q(t)$$
$$\Pr[\Delta(v',Y) \le r' \mid \Delta(v,X) \le r, \Delta(v,Y) \le r'] \le O(n)\cdot Q(\bar{\phi}t)$$

where $\bar{\phi} = 1-\phi$

Proof. Let $\ell \triangleq n/2 - \theta\cdot n/2$. Without loss of generality, we assume that $v = 0^n$ is the all-zeros vector and that $v' = 1^\ell 0^{n-\ell}$. Then,

$$\begin{aligned}
&\Pr[\Delta(v',X) \le r \mid \Delta(v,X) \le r, \Delta(v,Y) \le r'] \\
&= \Pr[\Delta(v',X) \le r \mid \mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r'] \\
&= \frac{\Pr[\Delta(v',X) \le r, \mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \\
&= \frac{1}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \\
&\quad\cdot \sum_{r_1=0}^{r}\sum_{r_2=0}^{r}\sum_{r_3=0}^{r'} \Pr[\Delta(v',X) = r_1, \mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \\
&= \frac{1}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']}\cdot \sum_{r_2=0}^{r}\sum_{r_3=0}^{r'} p(r_2,r_3) \\
&\qquad\cdot \sum_{r_1=0}^{r} \Pr[\Delta(v',X) = r_1 \mid \mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \\
&= \frac{1}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \\
&\quad\cdot \sum_{r_2=0}^{r}\sum_{r_3=0}^{r'} p(r_2,r_3)\cdot \sum_{r_1=0}^{r} \Pr[\Delta(v',X) = r_1 \mid \mathsf{wt}(X) = r_2] \qquad (4.6) \\
&= \frac{1}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \\
&\quad\cdot \sum_{r_2=0}^{r}\sum_{r_3=0}^{r'} p(r_2,r_3)\cdot \Pr[\Delta(v',X) \le r \mid \mathsf{wt}(X) = r_2],
\end{aligned}$$

where $p(r_2,r_3) \triangleq \Pr[\mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3]$, and where the penultimate equality follows from the fact that $\Delta(v',X) - \mathsf{wt}(X) - \mathsf{wt}(Y)$ is a Markov chain.

For every non-negative integer t_2 satisfying $t_2 = o(n^{1/4})$ and $\theta \cdot t \cdot t_2 = o_n(1)$, we have that

$$\begin{aligned}
&\Pr[\Delta(v', X) \le r \mid \mathsf{wt}(X) = n/2 - t_2\sqrt{n}/2] \\
&\quad = \sum_{a=0}^{a_{max}} \psi(a) \\
&\quad \overset{(A)}{\le} (a_{max} + 1) \cdot \psi(a_{max}) \\
&\quad \overset{(B)}{\le} O(n) \cdot \Theta\left(\frac{1}{\sqrt{n}}\right) \cdot e^{-\frac{t^2}{2}} \\
&\quad \overset{(C)}{\le} O(n) \cdot Q(t), \qquad (4.7)
\end{aligned}$$

where (A) follows from Proposition 8.3, (B) from Proposition 8.4 and the fact that $\theta = o_n(1)$, and (C) from Proposition 1.1 (a) and the facts that $t = \Theta(\sqrt{k})$ and $k \le n$. Note that by assumption $\theta = O(k/(\sqrt{n} \cdot \log n))$. Thus, for any $k = o(n^{1/4} \cdot \sqrt{\log n})$, there exists a function $\nu(t, \theta) = \omega_n(1)$ satisfying $\nu(t, \theta) = o_n(\min(n^{1/4}, 1/(t \cdot \theta)))$ and

$$\Theta(n^2) \cdot 2^{k + k \cdot \delta^2} \cdot \exp(-\nu(t, \theta)^2) \le Q(t). \qquad (4.8)$$

We fix such a function $\nu(t, \theta)$ and set $\tau(t, \theta) \triangleq n/2 - \nu(t, \theta)\sqrt{n}/2$. Equation (4.6) now becomes:

$$\begin{aligned}
&\Pr[\Delta(v', X) \le r \mid \Delta(v, X) \le r, \Delta(v, Y) \le r'] \qquad (4.9) \\
&\quad = \frac{1}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \cdot (\alpha + \beta),
\end{aligned}$$

where

$$\begin{aligned}
\alpha \triangleq &\sum_{r_2 = \tau(t,\theta)}^{r} \sum_{r_3 = 0}^{r'} \Pr[\mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \\
&\cdot \Pr[\Delta(v', X) \le r \mid \mathsf{wt}(X) = r_2],
\end{aligned}$$

and

$$\begin{aligned}
\beta \triangleq &\sum_{r_2 = 0}^{\tau(t,\theta)} \sum_{r_3 = 0}^{r'} \Pr[\mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \\
&\cdot \Pr[\Delta(v', X) \le r \mid \mathsf{wt}(X) = r_2].
\end{aligned}$$

Using eq. (4.7) and the fact that

$$\nu(t, \theta) = o(\min(n^{1/4}, 1/(t \cdot \theta))),$$

we get that α is at most

$$\begin{aligned}
&\sum_{r_2 = \tau(t,\theta)}^{r} \sum_{r_3 = 0}^{r'} \Pr[\mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \cdot O(n) \cdot Q(t) \qquad (4.10) \\
&\quad \le O(n) \cdot Q(t) \cdot \Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r'].
\end{aligned}$$

We also have that

$$\begin{aligned}
\beta &\le \sum_{r_2 = 0}^{\tau(t,\theta)} \sum_{r_3 = 0}^{r'} \Pr[\mathsf{wt}(X) = r_2, \mathsf{wt}(Y) = r_3] \qquad (4.11) \\
&\le \Pr[\mathsf{wt}(X) \le \tau(t, \theta)] \le \exp(-\nu(t, \theta)^2),
\end{aligned}$$

where the last inequality uses the fact that $\nu(t, \theta) = \omega_n(1)$ and follows from Proposition 4.1 and Proposition 1.1 (a). Combining eq. (4.9), eq. (4.10) and eq. (4.11), we get that

$$\begin{aligned}
&\Pr[\Delta(v', X) \le r \mid \Delta(v, X) \le r, \Delta(v, Y) \le r'] \\
&\le O(n) \cdot Q(t) + \frac{\exp(-\nu(t, \theta)^2)}{\Pr[\mathsf{wt}(X) \le r, \mathsf{wt}(Y) \le r']} \\
&\le O(n) \cdot Q(t) + \Theta(n^2) \cdot 2^k \cdot 2^{-k \cdot \delta^2} \cdot \exp(-\nu(t, \theta)^2) \\
&\le O(n) \cdot Q(t),
\end{aligned}$$

where the second inequality follows from Proposition 8.1 and Proposition 8.2, and the third inequality follows from the fact that $\nu(t, \theta)$ satisfies eq. (4.8). This completes the proof of the first part of the lemma. The proof of the second part follows along the same lines.

We note that the above bounds imply the desired result for agreement probabilities up to $1/\mathsf{poly}(k)$. The result also holds for *constant* agreement probability. The main idea is to combine the constant agreement scheme for the Gaussian source along with a multi-dimensional Berry-Esseen Theorem (e.g., Theorem 67 of [32]).[7] The details are deferred to a future version.

5 Information Complexity and Common Randomness

In this section, we show an intimate relationship between the achievable regions for amortized common randomness generation and the *internal* and *external* information costs of communication protocols, two well-studied notions in theoretical computer science. For any random variable X, let $H(X)$ denote its Shannon entropy. We say that a triple (H, R_1, R_2) of non-negative real numbers is *r-achievable* for a distribution μ if for every $\varepsilon > 0$ there exists an r-round common randomness scheme Π with $(X^n, Y^n) \sim \mu^{\otimes n}$ as inputs, for some $n = n(\varepsilon)$, where $n \to \infty$ as $\varepsilon \to 0$, such that the following holds: let M_t denote the message sent in round t in Π, and let K_A (resp. K_B) denote the output of Alice (resp. Bob). Then (1) $\sum_{t \text{ odd}} H(M_t) \le (R_1 + \varepsilon)n$, (2) $\sum_{t \text{ even}} H(M_t) \le (R_2 + \varepsilon)n$, (3) $H(K_A), H(K_B) \ge (H - \varepsilon)n$ and (4) K_A and K_B both belong to a domain

[7] Since we are dealing with constant error probabilities, the additive error from the Berry-Esseen theorem is negligible.

of size cn for some absolute constant c independent of ε and n. (The min-entropy guarantee in our basic definition is stronger than the combination of parts (3) and (4).)

DEFINITION 5.1. *Let P be a two-player randomized communication protocol with both public and private coins and let R_{pub} denote the public randomness. With a slight abuse in notation, given $(X,Y) \sim \mu$, let P also denote the transcript of the protocol on input (X,Y). Define the following measures for the protocol with respect to μ: (i) the* external information cost $\mathrm{IC}^{\mathrm{ext}}(P)$ *equals $I(X,Y;P \mid R_{pub})$; (ii) the* marginal internal information cost $\mathrm{IC}_A^{\mathrm{int}}(P)$ *for Alice equals $I(X;P \mid YR_{pub})$ and analogously $\mathrm{IC}_B^{\mathrm{int}}(P) = I(Y;P \mid XR_{pub})$ for Bob. The (total)* internal information cost *equals the sum of the two marginal costs.*

We now characterize the achievable region for a fixed source distribution μ in terms of internal and external information costs of protocols with respect to μ.

Converse. We extend the ideas present in several works, e.g. [2, 25, 31]. We need the following direct-sum property (Lemma 5.1 below) for information costs of randomized protocols that we crucially use in our analysis. This property differs from the known direct-sum results in that it simultaneously bounds the internal and external information costs of the single-coordinate protocol. Its proof uses the following tool.

PROPOSITION 5.1. ([2, LEMMA 4.1]) *Let S,T,X^n,Y^n be arbitrary random variables. Then:*

$$I(X^n;S \mid T) - I(Y^n;S \mid T) = \sum_{j=1}^{n} I(X_j;S \mid X^{j-1}Y_{j+1}^n T) - I(Y_j;S \mid X^{j-1}Y_{j+1}^n T).$$

Proof. We have by telescoping:

$$I(X^n;S \mid T) - I(Y^n;S \mid T) = \sum_{j=1}^{n} I(X^jY_{j+1}^n;S \mid T) - I(X^{j-1}Y_j^n;S \mid T). \tag{5.12}$$

By the chain rule for mutual information, for each $j \in [n]$, we have that

$$I(X^jY_{j+1}^n;S \mid T) = I(X^{j-1}Y_{j+1}^n;S \mid T) + I(X_j;S \mid X^{j-1}Y_{j+1}^n T)$$

and

$$I(X^{j-1}Y_j^n;S \mid T) = I(X^{j-1}Y_{j+1}^n;S \mid T) + I(Y_j;S \mid X^{j-1}Y_{j+1}^n T)$$

The proposition now follows by substituting the last two equations in Equation (5.12).

LEMMA 5.1. (DIRECT SUM) *Fix a distribution μ and an r-round randomized protocol Π with inputs $(X^n,Y^n) \sim \mu^{\otimes n}$. Then there exists an r-round randomized protocol P with inputs $(X,Y) \sim \mu$ such that (a) $\mathrm{IC}_A^{\mathrm{int}}(\Pi) = n \cdot \mathrm{IC}_A^{\mathrm{int}}(P)$, (b) $\mathrm{IC}_B^{\mathrm{int}}(\Pi) = n \cdot \mathrm{IC}_B^{\mathrm{int}}(P)$, and (c) $\mathrm{IC}^{\mathrm{ext}}(\Pi) \le n \cdot \mathrm{IC}^{\mathrm{ext}}(P)$.*

Proof. For ease of presentation we suppress the public randomness of Π in the expressions appearing in the proof below. Let M_t be the message sent in Π during round $t \in [r]$; set $M_{r+1} \coloneqq \emptyset$. We will be using the following properties of Π:

I. For every odd $t \le r$, $I(Y^n;M_t \mid X^nM^{t-1}) = I(X^n;M_{t+1} \mid Y^nM^t) = 0$.

II. For all $j \in [n]$ and odd $t \le r$, $I(Y_j;M_t \mid X^jY_{j+1}^nM^{t-1}) = I(X_j;M_{t+1} \mid X^{j-1}Y_j^nM^t) = 0$. This can also be shown, see, e.g., [25, Eqns. 3.10–3.13].

We present the argument for the marginal internal information cost for Alice; a similar argument can be carried out for Bob's case as well. Observe that:

$$\mathrm{IC}_A^{\mathrm{int}}(\Pi) = I(X^n;M^r \mid Y^n) = \sum_{t \le r} I(X^n;M_t \mid Y^nM^{t-1}) = \sum_{t \text{ odd}} I(X^n;M_t \mid Y^nM^{t-1}), \tag{5.13}$$

by part (I) above. Fix an odd t in the above sum. Again by part (I) above:

$$I(X^n;M_t \mid M^{t-1}) = I(X^nY^n;M_t \mid M^{t-1}) = I(Y^n;M_t \mid M^{t-1}) + I(X^n;M_t \mid Y^nM^{t-1}), \tag{5.14}$$

and therefore,

$$\begin{aligned} &I(X^n;M_t \mid Y^nM^{t-1}) \\ &= I(X^n;M_t \mid M^{t-1}) - I(Y^n;M_t \mid M^{t-1}) \\ &\overset{(a)}{=} \sum_{j=1}^{n} I(X_j;M_t \mid X^{j-1}Y_{j+1}^nM^{t-1}) \\ &\qquad - I(Y_j;M_t \mid X^{j-1}Y_{j+1}^nM^{t-1}) \\ &\overset{(b)}{=} \sum_{j=1}^{n} I(X_j;M_t \mid Y_jX^{j-1}Y_{j+1}^nM^{t-1}) \\ &\overset{(c)}{=} \sum_{j=1}^{n} I(X_j;M_tM_{t+1} \mid Y_jX^{j-1}Y_{j+1}^nM^{t-1}) \end{aligned}$$

where (a) follows from Proposition 5.1, and each of (b) and (c) follows from the chain rule and by invoking

part (II). We now substitute the last expression above in eq. (5.13), and sum over all odd t.

(5.15)
$$\begin{aligned}\mathrm{IC}_A^{\mathrm{int}}(\Pi) &= I(X^n; M^r \mid Y^n)\\ &= \sum_{t \text{ odd}} \sum_{j=1}^{n} I(X_j; M_t M_{t+1} \mid Y_j X^{j-1} Y_{j+1}^n M^{t-1})\\ &= \sum_{j=1}^{n} I(X_j; M^r \mid Y_j X^{j-1} Y_{j+1}^n)\\ &= n \cdot I(X_J; M^r \mid Y_J X^{J-1} Y_{J+1}^n J),\end{aligned}$$

using the chain rule and then defining J to be uniform over $[n]$ and independent of all the other random variables. Similarly for Bob:

$$\mathrm{IC}_B^{\mathrm{int}}(\Pi) = n \cdot I(Y_J; M^r \mid X_J X^{J-1} Y_{J+1}^n J). \tag{5.16}$$

We claim that the right side of eqs. (5.15) and (5.16) are respectively the marginal internal information costs for Alice and Bob in some protocol P with inputs $(X, Y) \sim \mu$. Specifically, on input pair (X, Y), the protocol P simulates the protocol Π by settting $X_J := X$ and $Y_J := Y$, and associating the public randomness with J, X^{J-1}, and Y_{J+1}^n. Part (II) above ensures that the messages in protocol P can be generated by the players using private randomness.

It remains to bound the external information cost of P. Observe that $\mathrm{IC}^{\mathrm{ext}}(P)$ equals

$$\begin{aligned}&I(X_J, Y_J; M^r \mid X^{J-1} Y_{J+1}^n J)\\ &= I(Y_J; M^r \mid X^{J-1} Y_{J+1}^n J)\\ &\qquad + I(X_J; M^r \mid Y_J X^{J-1} Y_{J+1}^n J).\end{aligned}$$

The term in the last line equals $\frac{1}{n} \cdot I(X^n; M^r \mid Y^n)$ via eq. (5.15). For the first term, using the independence of coordinates,

$$\begin{aligned}I(Y_J; M^r \mid X^{J-1} Y_{J+1}^n J) &= I(Y_J; M^r X^{J-1} \mid Y_{J+1}^n J)\\ &\geq I(Y_J; M^r \mid Y_{J+1}^n J)\\ &= \tfrac{1}{n} \cdot I(Y^n; M^r),\end{aligned}$$

where we expand over J and use the chain rule. Combining the bounds for the two terms, we conclude:

$$\begin{aligned}n \cdot \mathrm{IC}^{\mathrm{ext}}(P) &\geq I(Y^n; M^r) + I(X^n; M^r \mid Y^n)\\ &= I(X^n Y^n; M^r)\\ &= \mathrm{IC}^{\mathrm{ext}}(\Pi).\end{aligned}$$

THEOREM 5.1. *If a tuple (H, R_1, R_2) is r-achievable then for every $\varepsilon > 0$ there exists a randomized r-round protocol whose marginal internal information cost for Alice (resp. Bob) with respect to the distribution μ is at most $R_1 + O(\varepsilon) + 1/n$ (resp. $R_2 + O(\varepsilon) + 1/n$) and whose external information cost is at least $H - \varepsilon$.*

Proof. Fix $\varepsilon > 0$. Let n be such that there is an r-round protocol for common randomness generation Π on inputs $(X^n, Y^n) \sim \mu^{\otimes n}$. Let M_t denote the message sent in round t in Π. Let K_A (resp. K_B) denote the output of Alice (resp. Bob). We have (1) $\sum_{t \text{ odd}} H(M_t) \leq (R_1 + \varepsilon)n$, (2) $\sum_{t \text{ even}} H(M_t) \leq (R_2 + \varepsilon)n$, (3) $H(K_A), H(K_B) \leq (H - \varepsilon)n$ and (4) K_A and K_B both belong to a domain of size cn for some absolute constant c (independent of ϵ and n).

Consider the case where r is odd (the other case can be handled similarly) and define a new protocol Π' where Alice also sends K_A to Bob along with the last message. The number of rounds is still r. Applying Lemma 5.1, there exists an r-round randomized protocol P with inputs $(X, Y) \sim \mu$ such that $\mathrm{IC}_A^{\mathrm{int}}(\Pi') = n \cdot \mathrm{IC}_A^{\mathrm{int}}(P)$ and $\mathrm{IC}_B^{\mathrm{int}}(\Pi') = n \cdot \mathrm{IC}_B^{\mathrm{int}}(P)$. Now since Π' depends only on X^n and Y^n, we have that $\mathrm{IC}_A^{\mathrm{int}}(\Pi') = I(X^n; M^r K_A \mid Y^n) = I(X^n; M^r \mid Y^n) + I(X^n; K_A \mid Y^n M^r)$. Because $X^n \perp M_t \mid Y^n M^{t-1}$ for each even round t, by the chain rule, the first term equals

$$\begin{aligned}\sum_t I(X^n; M_t \mid Y^n M^{t-1}) &= \sum_{t \text{ odd}} I(X^n; M_t \mid Y^n M^{t-1})\\ &\leq \sum_{t \text{ odd}} H(M_t) \leq (R_1 + \varepsilon)n.\end{aligned}$$

The second term is at most $H(K_A \mid Y^n M^r)$. Now K_B is determined by Y^n and M^r and $\Pr[K_A \neq K_B] \leq \varepsilon$, so by Fano's inequality, $H(K_A \mid Y^n M^r) \leq \varepsilon cn + 1$. Therefore, $\mathrm{IC}_A^{\mathrm{int}}(P) \leq R_1 + \varepsilon(1 + c) + 1/n$. For Bob, the analysis is similar and even simpler because his messages are unchanged (from Π to Π') so $\mathrm{IC}_B^{\mathrm{int}}(P) \leq R_2 + \varepsilon$. (The bound stated in the lemma is weaker because Fano's inequality is used when r is even.) Finally, apply Lemma 5.1 to bound the external information cost of P as

$$\begin{aligned}n \cdot \mathrm{IC}^{\mathrm{ext}}(P) &\geq \mathrm{IC}^{\mathrm{ext}}(\Pi') = I(X^n Y^n; M^r K_A)\\ &= H(M^r K_A) - H(M^r K_A \mid X^n Y^n)\\ &= H(M^r K_A).\end{aligned}$$

But $H(M^r K_A) \geq H(K_A) \geq (H - \varepsilon)n$, so the desired bound follows.

Achievability. In [31], a sufficient condition using Markov chains on auxiliary random variables is given the existence of an interactive common randomness scheme. To fulfill this condition, their construction uses a random encoding argument. We connect these conditions to the existence of an r-round communication protocol with the appropriate information costs.

PROPOSITION 5.2. ([31]) *Let $(X, Y) \sim \mu$. Suppose there exist auxiliary random variables $U_1, U_2, \ldots, U_r$ for*

some r in some joint probability space with X and Y where the marginal distribution of (X,Y) is μ satisfying the following:

1. *For every odd t, $Y \perp U_t \mid XU^{t-1}$ and for every even t, $X \perp U_{t+1} \mid YU^t$.*
2. *$\sum_{t \text{ odd}} I(X;U_t \mid U^{t-1}) + \sum_{t \text{ even}} I(Y;U_t \mid U^{t-1}) \geq H$.*
3. *$\sum_{t \text{ odd}} I(X;U_t \mid U^{t-1}) - \sum_{t \text{ odd}} I(Y;U_t \mid U^{t-1}) \leq R_1$.*
4. *$\sum_{t \text{ even}} I(Y;U_t \mid U^{t-1}) - \sum_{t \text{ even}} I(X;U_t \mid U^{t-1}) \leq R_2$.*

Then, there exists an r-round interactive common randomness generation scheme $\Pi(X^n,Y^n)$ using n i.i.d. samples as input where Alice sends at most $R_1 n$ bits, Bob sends at most $R_2 n$ bits and the entropy of their output is at least Hn bits where the agreement probability tends to 1 as $n \to \infty$.

THEOREM 5.2. *If there exists a r-round randomized protocol with inputs $(X,Y) \sim \mu$ whose marginal internal information cost for Alice (resp. Bob) is at most R_1 (resp. R_2) and whose external information cost is at least H, then (H,R_1,R_2) is r-achievable.*

Proof. Let P be a randomized protocol with inputs $(X,Y) \sim \mu$ whose marginal internal information cost for Alice (resp. Bob) is at most R_1 (resp. R_2) and whose external information cost is at least H. Without loss of generality, we assume that P uses no public randomness. For every $t \in [r]$, we let U_t denote the message sent in P during round r. We claim that the U_t's satisfy the conditions in Proposition 5.2. First, note that the conditional independencies given in part 1 of Proposition 5.2 are equivalent to the message structure of an r-round randomized protocol, and are thus satisfied by the U_t's.

For every odd t, by part 1, $I(Y;U_t \mid XU^{t-1}) = 0$, so

$$\begin{aligned} I(X;U_t \mid U^{t-1}) &= I(XY;U_t \mid U^{t-1}) \\ &= I(Y;U_t \mid U^{t-1}) + I(X;U_t \mid YU^{t-1}). \end{aligned}$$

Therefore, $I(X;U_t \mid YU^{t-1}) = I(X;U_t \mid U^{t-1}) - I(Y;U_t \mid U^{t-1})$. By the chain rule,

$$\begin{aligned} \mathrm{IC}_A^{\mathrm{int}}(P) = I(X;U^r \mid Y) &= \sum_t I(X;U_t \mid YU^{t-1}) \\ &= \sum_{t \text{ odd}} I(X;U_t \mid U^{t-1}) - I(Y;U_t \mid U^{t-1}) \leq R_1, \end{aligned}$$

via part 1 where we used $I(X;U_t \mid YU^{t-1}) = 0$ for every even t. Using the given assumption that $\mathrm{IC}_A^{\mathrm{int}}(P) \leq R_1$, we deduce that the U_t's satisfy part 3 of Proposition 5.2. A similar argument using the given assumption that $\mathrm{IC}_B^{\mathrm{int}}(P) \leq R_2$ implies that the U_t's satisfy part 4 of Proposition 5.2.

Applying a similar reasoning, we also obtain that:

$$\begin{aligned} &\sum_{t \text{ odd}} I(X;U_t \mid U^{t-1}) + \sum_{t \text{ even}} I(Y;U_t \mid U^{t-1}) \\ &= \sum_{t \text{ odd}} I(XY;U_t \mid U^{t-1}) + \sum_{t \text{ even}} I(XY;U_t \mid U^{t-1}) \\ &= I(XY;U^r) \\ &= \mathrm{IC}^{\mathrm{ext}}(P). \end{aligned}$$

The given assumption that $\mathrm{IC}^{\mathrm{ext}}(P) \geq H$ now implies that the U_t's satisfy part 2 of Proposition 5.2. Therefore, we conclude that (H,R_1,R_2) is r-achievable.

Combining Theorem 5.1 and Theorem 5.2, we obtain the the formal version of Theorem 1.3.

THEOREM 5.3. *Let Γ_r denote the supremum over all r-round randomized protocols Π of the ratio of the external information cost to the internal information cost of Π with respect to μ. Then, Γ_r equals the supremum of $H/(R_1+R_2)$ such that (H,R_1,R_2) is r-achievable for μ.*

6 Conclusion and Open Questions

The most important open question raised in this work is to obtain *computationally* efficient schemes for common randomness. In particular, is there a resource-efficient scheme that also has time complexity $\mathsf{poly}(k)$? For our schemes, it not at all clear how to implement the decoding phase time-efficiently (either over $\mathbb{F}_2$ or in Euclidean space). In fact, even the slightly subexponential time algorithm of [27] for decoding dual-BCH codes falls short of working for the error radii that are needed to achieve near-optimal agreement probability!

The sample complexity $n = o(k^4)$ of our explicit schemes is polynomial but still far from the linear non-explicit sample schemes arising from amortized common randomness. The Kabatjanskii-Levenstein bound (cf. [41]) implies that no nearly-orthogonal families of vectors (including the one we used) will achieve a linear sample complexity in our setup. Can we rule out linear sample schemes altogether? One challenge is that such a proof cannot solely rely on hypercontractivity because they "tensorize" and are thus oblivious to the number n of used samples.

Our one-way scheme for general sources with *maximal correlation* ρ is explicit but not sample-efficient because it uses the CLT to reduce the problem to

$\mathsf{BGS}(\rho)$. Moreover, the tradeoff between communication and agreement is stated in terms of ρ, whereas the best known negative results are in terms of *hypercontractivity*. [5] give an example of a source separating its maximal correlation from its *Strong Data Processing Constant*, which is intimately related to its hypercontractive properties. Can such a source be used to prove that the tradeoff stated in Theorem 1.2 is not tight for general sources?

A characterization of amortized *correlated* randomness would be interesting even for one-way as it would generalize the notion of the Strong Data Processing Constant.

Finally, our paper shows that tools used in common randomness could also be useful for Locality Sensitive Hashing. Can one establish a formal connection between these two areas?

Acknowledgements

The authors would like to thank Madhu Sudan for several helpful discussions, in particular, related to the non-asymptotic bounds in Section 8. They would also like to thank Venkat Guruswami, Clément Canonne, Jingbo Liu and Ilya Razenshteyn for very helpful discussions and pointers.

References

[1] Ahlswede, R., and Csiszár, I. Common randomness in information theory and cryptography. part I: Secret sharing. *IEEE Transactions on Information Theory 39*, 4 (1993).

[2] Ahlswede, R., and Csiszár, I. Common randomness in information theory and cryptography. II. CR capacity. *IEEE Transactions on Information Theory 44*, 1 (1998), 225–240.

[3] Ahlswede, R., and Dueck, G. Identification via channels. *IEEE Trans. Information Theory 35*, 1 (1989), 15–29.

[4] Ahlswede, R., and Gács, P. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability* (1976), 925–939.

[5] Anantharam, V., Gohari, A., Kamath, S., and Nair, C. On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover. *arXiv preprint arXiv:1304.6133* (2013).

[6] Bar-Yossef, Z., Jayram, T. S., Kumar, R., and Sivakumar, D. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences 68*, 4 (2004), 702–732.

[7] Barak, B., Braverman, M., Chen, X., and Rao, A. How to compress interactive communication. *SIAM Journal on Computing 42*, 3 (2013), 1327–1363.

[8] Bavarian, M., Gavinsky, D., and Ito, T. On the role of shared randomness in simultaneous communication. In *Automata, Languages, and Programming*. Springer, 2014, pp. 150–162.

[9] Beckner, W. Inequalities in Fourier analysis. *Annals of Mathematics* (1975), 159–182.

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

[11] Bonami, A. Étude des coefficients de Fourier des fonctions de $L^p(g)$. In *Annales de l'institut Fourier* (1970), vol. 20, pp. 335–402.

[12] Canonne, C., Guruswami, V., Meka, R., and Sudan, M. Communication with imperfectly shared randomness. *ITCS* (2014).

[13] Chakrabarti, A., Shi, Y., Wirth, A., and Yao, A. Informational complexity and the direct sum problem for simultaneous message complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on* (2001), IEEE, pp. 270–278.

[14] Chan, S. O., Mossel, E., and Neeman, J. On extracting common random bits from correlated sources on large alphabets. *Information Theory, IEEE Transactions on 60*, 3 (2014), 1630–1637.

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

[16] Csiszár, I., and Narayan, P. Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory 50*, 12 (2004), 3047–3061.

[17] Duembgen, L. Bounding standard gaussian tail probabilities. *arXiv preprint arXiv:1012.2063* (2010).

[18] Gács, P., and Körner, J. Common information is far less than mutual information. *Problems of Control and Information Theory 2*, 2 (1973), 149–162.

[19] Gebelein, H. Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik 21*, 6 (1941), 364–379.

[20] Ghazi, B., and Jayram, T. S. Resource-efficient common randomness and secret-key schemes. *arXiv preprint arXiv:1707.08086* (2017).

[21] Ghazi, B., Kamath, P., and Sudan, M. Communication complexity of permutation-invariant functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (2016), SIAM, pp. 1902–1921.

[22] Guruswami, V., and Radhakrishnan, J. Tight bounds for communication-assisted agreement distillation. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan* (2016), pp. 6:1–6:17.

[23] Hirschfeld, H. O. A connection between correlation and contingency. In *Mathematical Proceedings of the Cambridge Philosophical Society* (1935), vol. 31, Cam-

bridge Univ Press, pp. 520–524.

[24] Indyk, P., and Motwani, R. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998* (1998), pp. 604–613.

[25] Kaspi, A. H. Two-way source coding with a fidelity criterion. *IEEE Trans. Information Theory 31*, 6 (1985), 735–740.

[26] Kaufman, T., and Lovett, S. New extension of the weil bound for character sums with applications to coding. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on* (2011), IEEE, pp. 788–796.

[27] Kopparty, S., and Saraf, S. Local list-decoding and testing of random linear codes from high error. *SIAM Journal on Computing 42*, 3 (2013), 1302–1326.

[28] Lim, D., Lee, J. W., Gassend, B., Suh, G. E., Van Dijk, M., and Devadas, S. Extracting secret keys from integrated circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems 13*, 10 (2005), 1200–1205.

[29] Liu, J. Rate region for interactive key generation and common randomness generation. *Manuscript available at http:// www.princeton.edu/~jingbo/ preprints/ RateRegionInteractiveKeyGen120415.pdf (visited on 02/13/2017)* (2016).

[30] Liu, J., Cuff, P., and Verdú, S. Secret key generation with one communicator and a one-shot converse via hypercontractivity. In *2015 IEEE International Symposium on Information Theory (ISIT)* (2015), IEEE, pp. 710–714.

[31] Liu, J., Cuff, P., and Verdú, S. Common randomness and key generation with limited interaction. *arXiv preprint arXiv:1601.00899* (2016).

[32] Matulef, K., O'Donnell, R., Rubinfeld, R., and Servedio, R. A. Testing halfspaces. *SIAM Journal on Computing 39*, 5 (2010), 2004–2047.

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

[34] Mossel, E., and O'Donnell, R. Coin flipping from a cosmic source: On error correction of truly random bits. *arXiv preprint math/0406504* (2004).

[35] Mossel, E., O'Donnell, R., Regev, O., Steif, J. E., and Sudakov, B. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse Bonami-Beckner inequality. *Israel Journal of Mathematics 154*, 1 (2006), 299–336.

[36] Rényi, A. On measures of dependence. *Acta mathematica hungarica 10*, 3-4 (1959), 441–451.

[37] Sheppard, W. On the application of the theory of error to cases of normal distribution and normal correlation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* (1899), 101–531.

[38] Su, Y., Holleman, J., and Otis, B. P. A digital 1.6 pj/bit chip identification circuit using process variations. *IEEE Journal of Solid-State Circuits 43*, 1 (2008), 69–77.

[39] Sudan, M. Personal communication, 2014.

[40] Suh, G. E., and Devadas, S. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of the 44th annual Design Automation Conference* (2007), ACM, pp. 9–14.

[41] Tao, T. A cheap version of the Kabatjanskii-Levenstein bound for almost orthogonal vectors. https://terrytao.wordpress.com/2013/07/18/a-cheap-version-of-the-kabatjanskii-levenstein-bound-for-almost-orthogonal-vectors/, 2013.

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

[43] Weil, A. *Sur les courbes algébriques et les variétés qui s' en déduisent*. No. 1041. Hermann, 1948.

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

[45] Wyner, A. D. The common information of two dependent random variables. *IEEE Transactions on Information Theory 21*, 2 (1975), 163–179.

[46] Yu, H., Leong, P. H. W., Hinkelmann, H., Moller, L., Glesner, M., and Zipf, P. Towards a unique FPGA-based identification circuit using process variations. In *2009 International Conference on Field Programmable Logic and Applications* (2009), IEEE, pp. 397–402.

[47] Zhao, L., and Chia, Y.-K. The efficiency of common randomness generation. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (2011).

7 Properties of Bivariate Gaussian Distribution

Proposition 7.1. *(Elliptical symmetry, Proposition 1.2 restated) Let* $(X, Y) \sim \mathsf{BGS}(\rho)^{\otimes n}$ *and* $v, w \in \mathbb{R}^n$ *have unit norm. Then,* $(v \bullet X, w \bullet Y) \sim \mathsf{BGS}\big(\rho(v \bullet w)\big)$.

Proof. Since $(v \bullet X, w \bullet Y)$ is a linear transform of (X, Y) it has the bivariate Gaussian distribution. Thus, we only need to determine the first and second moments. Since v and w have unit-norm, by stability, the marginals are standard normal. Finally, we verify that their covariance is $\rho(v \bullet w)$.

$$\begin{aligned}\mathbb{E}[v \bullet X w \bullet Y] &= \sum_{i,j=1}^{n} v(i)w(j) \cdot \mathbb{E}[X_i \cdot Y_j] \\ &= \rho \sum_{i=1}^{n} v(i)w(i) = \rho(v \bullet w).\end{aligned}$$

7.1 Tail bounds for Gaussians The following bounds are well-known; using Duembgen's approach [17], we prove them below to make it self-

contained. Let $\lambda(t) \coloneqq \frac{\phi(t)}{Q(t)}$ denote the *inverse Mills ratio*, i.e the ratio of the density function to the tail probability of a standard normal random variable. Let $\lambda_0 \coloneqq \lambda(0) = \sqrt{\frac{2}{\pi}}$.

LEMMA 7.1. *For all $t \geq 0$, $\max\{t, \lambda_0^2 \cdot t + \lambda_0\} \leq \lambda(t) \leq t + \min\{1/t, \lambda_0\}$. Equality holds only at $t = 0$.*

Proof. For all $t \geq 0$ and any function $\alpha\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{t\to\infty} \alpha(t) = \infty$, let

$$f_\alpha(t) \coloneqq \frac{\phi(t)}{\alpha(t)} - Q(t),$$

so that $\lim_{t\to\infty} f_\alpha(t) = 0$. Observing that $Q'(t) = -\phi(t)$ and $\phi'(t) = -t\phi(t)$, we have:

$$\frac{\partial f_\alpha}{\partial t} = \frac{\phi(t)}{\alpha(t)^2}\big(\alpha(t)^2 - t \cdot \alpha(t) - \alpha'(t)\big)$$

Thus, the sign of the partial derivative is determined by $g_\alpha(t) \coloneqq \alpha(t)^2 - t \cdot \alpha(t) - \alpha'(t)$.

1. $\alpha(t) = t + 1/t$: In this case, $g_\alpha(t) = 2/t^2 > 0$. Therefore, $f_\alpha(t)$ is strictly increasing in t; together with $f_\alpha(0) = -\frac{1}{2}$ and $\lim_{t\to\infty} f_\alpha(t) = 0$, it follows that that $f_\alpha(t) < 0$ for all $t \geq 0$.

2. $\alpha(t) = t + \lambda_0$: In this case, $g_\alpha(t) = \lambda_0 t + \lambda_0^2 - 1$ is linear in t. Set $d \coloneqq (1 - \lambda_0^2)/\lambda_0 > 0$, and it follows that $g_\alpha(t) < 0$ for $0 \leq t < d$ and $g_\alpha(t) > 0$ for $t > d$. Therefore, $f_\alpha(t)$ is decreasing in t over $[0, d]$ and increasing in t over $[d, \infty)$; the endpoint conditions imply that $f_\alpha(t) \leq 0$ for all $t \geq 0$ with equality only at $t = 0$.

3. $\alpha(t) = t$: In this case, $g_\alpha(t) = -1$ so $f_\alpha(t)$ is strictly decreasing in t. Now $\lim_{t\to 0} f_\alpha(t) = \infty$ therefore $f_\alpha(t) > 0$ for all $t \geq 0$.

4. $\alpha(t) = \lambda_0^2 \cdot t + \lambda_0$: In this case, $g_\alpha(t)$ is quadratic in t with a zero constant term. Set $d \coloneqq \frac{2\lambda_0^2 - 1}{\lambda_0(1-\lambda_0)} > 0$ and an easy calculation shows that $g_\alpha(t) > 0$ for $0 \leq t < d$ and $g_\alpha(t) \leq 0$ for $t > d$. An analogous argument implies that $f_\alpha(t) \geq 0$ for all $t \geq 0$ with equality only at $t = 0$.

We now show two interesting properties of the tail probability function. These seem to be new as far as we know.

LEMMA 7.2. *The function $Q(t)^{1/t^2}$ is increasing in t for $t \geq 0$. For every fixed $0 \leq a \leq 1$, the function $Q(at)/Q(t)$ is increasing in t for $t \geq 0$.*

Proof. We use the basic identities $(\ln Q(t))' = -\lambda(t)$ and $\lambda'(t) = \lambda(t)^2 - t\lambda(t)$.

For the first property, it suffices to show that the function $f(t) \coloneqq \frac{1}{t^2} \cdot \ln Q(t)$ is increasing in t for $t \geq 0$. We have:

$$\frac{df}{dt} = -\frac{t\lambda(t) + 2\ln(Q(t))}{t^3}.$$

Let $u(t) \coloneqq t\lambda(t) + 2\ln(Q(t))$ and observe that $u'(t) = \lambda(t)(t \cdot \lambda(t) - t^2 - 1) < 0$ by Lemma 7.1. Thus $f'(t) > 0$ and $f(t)$ is increasing in t.

For the second property, it suffices to show that the function $g(t, a) \coloneqq \ln Q(at) - \ln Q(t)$ for each fixed $0 \leq a \leq 1$ is increasing in t for $t \geq 0$. We have:

$$\frac{\partial g}{\partial t} = \lambda(t) - a \cdot \lambda(at)$$

At $t = 0$ the right side equals 0 and for $t > 0$ we will show that $\lambda(t) > a \cdot \lambda(bt)$. This would imply the desired property that $g(t, a)$ is increasing in t. Multiplying both sides by t, we need to show that $t \cdot \lambda(t) > at \cdot \lambda(at)$, that is, the function $h(x) \coloneqq x \cdot \lambda(x)$ is an increasing function of x for $x \geq 0$. This holds because $h'(x) = \lambda(x)(1 - x^2 + x\lambda(x)) > 0$ by Lemma 7.1.

We are ready to prove Proposition 1.1.

PROPOSITION 7.2. (PROPOSITION 1.1 RESTATED) *Let $t, \delta \geq 0$. Set $\eta \coloneqq \rho + \delta\sqrt{1-\rho^2}$ and $\lambda_0 \coloneqq \sqrt{\frac{2}{\pi}}$. Then:*

$$(a)\ \ \frac{e^{-t^2/2}}{t + \lambda_0} \lesssim Q(t) \lesssim \frac{e^{-t^2/2}}{t + 1/\lambda_0} \leq e^{-t^2/2};$$

$$(b)\ \ \frac{Q(t)^{\delta^2}}{\delta t + \lambda_0} \lesssim Q(\delta t) \lesssim Q(t)^{\delta^2}(t + \lambda_0)^{c^2};$$

$$(c)\ \ L(t, \eta; \rho) \geq Q(t)Q(\delta t); \quad \textit{and}$$

$$(d)\ \ Q(t) \leq Q(\delta t) \leq Q(t)^{\delta^2}, \quad \textit{if}\ \ \delta \leq 1$$

Proof. Substituting the definition of $\lambda(t)$ in Lemma 7.1 and simplifying the expression, we obtain (a). Applying these bounds appropriately on both sides of (b) proves that inequality as well.

Next, let $(X, Y) \sim \mathsf{BGS}(\rho)$ so that $L(t, \eta; \rho) = \Pr[X > t, Y > \eta t]$. When $\rho = 1$, we have $X = Y$ with probability 1 so $L(t, \eta; \rho) = Q(t)$, implying (c) trivially. Therefore, let $\rho < 1$.

Now $Y = \rho X + \sqrt{1-\rho^2}Z$ where $Z \sim \mathcal{N}(0, 1)$ is independent of (X, Y). Observe:

$$\begin{aligned}
\Pr[X > t, Y > \eta t] &= \Pr[X > t,\ \rho X + \sqrt{1-\rho^2}Z > \eta t] \\
&\geq \Pr[X > t,\ \rho t + \sqrt{1-\rho^2}Z > \eta t] \\
&= \Pr[X > t, Z > \delta t] \ \ (\text{because } \rho < 1) \\
&= Q(t) \cdot Q(\delta t),
\end{aligned}$$

proving (c). For the last inequality, because $\delta \leq 1$, we have $Q(t) \leq Q(\delta t)$, and the latter can be bounded from above using the first property in Lemma 7.2, which implies (d).

8 Non-Asymptotic Bounds on Correlated Binomials

We let $h(\cdot)$ denote the binary entropy function.

FACT 8.1. *Stirling's approximation of the factorial implies that for every integers $0 < \ell < m$, we have that*

$$\binom{m}{\ell} = \Theta\left(\sqrt{\frac{m}{\ell \cdot (m-\ell)}}\right) \cdot 2^{-m \cdot h(\frac{\ell}{m})}.$$

FACT 8.2. *(Taylor approximation of binary entropy function) For every $x \in [0,1]$, we have that*

$$h(1/2 - x/2) = 1 - \frac{1}{2\ln 2}x^2 - O(x^4).$$

We now prove Proposition 4.1.

Proof. [Proof of Proposition 4.1] We have that

$$\begin{aligned}
&\Pr_{X \in_R \{0,1\}^n}[\mathsf{wt}(X) \leq n/2 - u\sqrt{n}/2] \\
&= \sum_{i=0}^{n/2-u\sqrt{n}/2} \binom{n}{i} \cdot 2^{-n} \\
&\overset{(A)}{\leq} n \cdot 2^{-n} \cdot \binom{n}{n/2 - u\sqrt{n}/2} \\
&= n \cdot 2^{-n} \cdot \Theta\left(\sqrt{\frac{n}{(n/2 - u\sqrt{n}/2) \cdot (n/2 + u\sqrt{n}/2)}}\right) \\
&\cdot 2^{n \cdot h\left(\frac{n/2 - u\sqrt{n}/2}{n}\right)} \\
&\overset{(B)}{\leq} O(n) \cdot 2^{-n} \cdot 2^{n \cdot (1 - \frac{u^2}{2 \cdot \ln 2 \cdot n})} \\
&= O(n) \cdot e^{-\frac{u^2}{2}} \\
&\overset{(C)}{\leq} O(n^2) \cdot Q(u),
\end{aligned}$$

where (A) follows from Fact 8.1, (B) from Fact 8.2, and (C) from Proposition 1.1 (a). Since the distribution of $\mathsf{wt}(X)$ is symmetric around $n/2$, the other case follows as well.

We point out that in the statement of Proposition 4.1 we made no effort to optimize the multiplicative function of n since that would not be consequential for our purposes. Recall that $r := n/2 - t\sqrt{n}/2$.

PROPOSITION 8.1. *For any $k = o(\sqrt{n})$, we have that*

$$\Pr_{X \in_R \{0,1\}^n}[\mathsf{wt}(X) \leq r] \geq \frac{1}{\Theta(\sqrt{n})} \cdot Q(t).$$

The proof of Proposition 8.1 appears in the full version of the paper [20].

PROPOSITION 8.2. *Fix $\epsilon \in (0, 0.5]$. For every $n = \omega(k^3)$, we have that*

$$\begin{aligned}
&\Pr_{(X,Y) \sim \mathsf{DSBS}(1-2\epsilon)^{\otimes n}}[Y \in \mathsf{Ball}(0,r') | X \in \mathsf{Ball}(0,r)] \\
&\geq \Theta\left(\frac{1}{n^{1.5}}\right) \cdot 2^{-\delta^2 k},
\end{aligned}$$

where $\mathsf{Ball}(0,r)$ denotes the Hamming ball of radius r centered around the all-zeros vector.

The proof of Proposition 8.2 appears in the full version [20].

In order to prove Lemma 4.2, we will need the following propositions.

PROPOSITION 8.3. *Let $t_2 \geq 0$ and $a_{max} \triangleq n \cdot (1+\theta)/4 - (t+t_2) \cdot \sqrt{n}/4$. For every $a \in \{0, 1, \ldots, a_{max}\}$, let*

$$\psi(a) \triangleq \frac{\binom{n \cdot (1+\theta)/2}{a} \cdot \binom{n \cdot (1-\theta)/2}{n/2 - t_2\sqrt{n}/2 - a}}{\binom{n}{n/2 - t_2\sqrt{n}/2}}.$$

Then, $\psi(a)$ is monotonically increasing in a.

The proof of Proposition 8.3 appears in the full version [20].

PROPOSITION 8.4. *Assume that $t = o(n^{1/4})$, $t_2 = o(n^{1/4})$ and $\theta \cdot t \cdot t_2 = o_n(1)$. Then,*

$$\psi(a_{max}) \leq \Theta\left(\frac{1}{\sqrt{n}}\right) \cdot e^{-\frac{t^2}{2}}.$$

The proof of Proposition 8.4 appears in the full version [20].

9 Correlated Randomness Generation

We first recall that Canonne et al. [12] – using the converse bound of [10] – showed that for any $\epsilon > 0$, if Alice and Bob are given access to i.i.d. samples from $\mathsf{DSBS}(1-2\epsilon)$, then, *perfectly agreeing* on k random bits requires $\Omega_\epsilon(k)$ bits of communication even in the two-way model. They also raised the following intriguing question: "What if their goal is only to generate more correlated bits than they start with? What is possible here and what are the limits?"

We partially answer this question and show that for any $\epsilon > 0$ and $\epsilon' \gg \epsilon \cdot \log(1/\epsilon)$, if Alice and Bob are given access to i.i.d. samples from $\mathsf{DSBS}(1-2\epsilon')$, then, generating k random samples from $\mathsf{DSBS}(1-2\epsilon)$ requires $\Omega_{\epsilon,\epsilon'}(k)$ bits of communication.

DEFINITION 9.1. *(Correlated Randomness Generation) In the CorrelatedRandomness$_{\gamma,\epsilon',\alpha,k}$ problem, Alice and Bob are given access to i.i.d. samples from a known source/distribution μ. Their goal is to for Alice to output $w_A \in \{0,1\}^k$ and for Bob to output $w_B \in \{0,1\}^k$, that satisfy the following properties: (i) $\Pr[\Delta(w_A, w_B) \leq \epsilon' k] \geq \gamma$; (ii) $H_\infty(w_A) \geq \alpha \cdot k$; and (iii) $H_\infty(w_B) \geq \alpha \cdot k$.*

We point out that one can alternatively define Correlated Randomness Generation in terms of coming close, say in total variation distance, to $\mathsf{DSBS}(1-2\varepsilon')^{\otimes k}$. The results in this section apply to this variant as well. This is because of the next lemma which can be proved by a simple Chernoff bound and which says that if Alice and Bob are given access to i.i.d. samples from $\mathsf{DSBS}(1-2\epsilon')$, then they can generate two length-k binary strings that lie in a Hamming ball of radius $\approx \epsilon' \cdot k$ with high probability.

LEMMA 9.1. *Fix $0 < \delta < \epsilon'$ and let $\mathsf{DSBS}(1-2(\epsilon'-\delta))$ be the source. Then, there is a non-interactive protocol solving CorrelatedRandomness$_{\gamma,\epsilon',\alpha,k}$ with $\gamma = 1-\exp(-(\epsilon'-\delta)^2 \cdot k)$ and $\alpha = 1$.*

We are now ready to state the main result.

THEOREM 9.1. *(Interactive Correlated Randomness Generation) Any interactive protocol solving CorrelatedRandomness$_{\gamma,\epsilon',\alpha,k}$ for the source $\mathsf{DSBS}(1-2(\epsilon'-\delta))$ with $h(\epsilon') \leq 4 \cdot \epsilon \cdot (1-\epsilon) \cdot \alpha/(1+\Omega(1))$ should communicate at least $\Omega(\epsilon \cdot \alpha \cdot k) - O(\log(1/\gamma))$ bits.*

Theorem 9.2 says that *non-interactively* generating two strings with min-entropy k and that lie in a Hamming ball of radius $\approx \epsilon' \cdot k$ cannot be done with success probability $2^{-o_\epsilon(k)}$ when Alice and Bob are given access to i.i.d. samples from $\mathsf{DSBS}(1-2\epsilon)$ with $\epsilon = \omega(\epsilon' \cdot \log(1/\epsilon'))$.

THEOREM 9.2. *(Non-Interactive Correlated Randomness Generation) There is no non-interactive protocol solving CorrelatedRandomness$_{\gamma,\epsilon',\alpha,k}$ for the source $\mathsf{DSBS}(1-2\epsilon)$ with $h(\epsilon') \leq 4 \cdot \epsilon \cdot (1-\epsilon) \cdot \alpha$ and $\gamma > 2^{-\nu k}$ where*

$$\nu = \alpha \cdot \frac{[\sqrt{1-h(\epsilon')/\alpha} - (1-2\epsilon)]^2}{4 \cdot \epsilon \cdot (1-\epsilon)}.$$

Consequently, whenever $h(\epsilon') \leq 4 \cdot \epsilon \cdot (1-\epsilon) \cdot \alpha/(1+\Omega(1))$, there is no non-interactive protocol solving CorrelatedRandomness$_{\gamma,\epsilon',\alpha,k}$ given i.i.d. access to $\mathsf{DSBS}(1-2\epsilon)$ with $\gamma > 2^{-\Omega(\epsilon \cdot \alpha \cdot k)}$.

We point out that getting the tight bounds in Theorems 9.1 and 9.2 remains a very interesting open question. In order to prove Theorems 9.1 and 9.2, we next introduce a "list" version of Common Randomness which is implicit in several of the known converse results for Common Randomness Generation.

DEFINITION 9.2. *(List Common Randomness Generation) In the ListCommonRandomness$^k_{\gamma,b}$ problem, Alice and Bob are given access to i.i.d. samples from a known distribution μ over pairs of random variables. Their goal is for Alice to output an element w_A and for Bob to output a list L_B (over the same universe), such that (i) $\Pr[w_A \in L_B] \geq \gamma$; (ii) $H_{\min}(w_A) \geq k$; and (iii) $|L_B| \leq b$.*

We prove the following converse results for List Common Randomness Generation both in the non-interactive and two-way communication models:

THEOREM 9.3. *(Non-Interactive List Common Randomness Generation) There is no non-interactive protocol solving ListCommonRandomness$^k_{\gamma,b}$ for the source $\mathsf{DSBS}(1-2\epsilon)$ with $(\log b)/k \leq 4 \cdot \epsilon \cdot (1-\epsilon)$ and with $\gamma > 2^{-\nu k}$ where*

$$\nu = \frac{[\sqrt{1-(\log b)/k} - (1-2\epsilon)]^2}{4 \cdot \epsilon \cdot (1-\epsilon)}.$$

Consequently, whenever $(\log b)/k \leq 4 \cdot \epsilon \cdot (1-\epsilon)/(1+\Omega(1))$, there is no non-interactive protocol solving ListCommonRandomness$^k_{\gamma,b}$ with $\gamma > 2^{-\Omega(\epsilon \cdot k)}$.

Proof. The proof is very similar to that of the converse result of [22]. Let Π be a protocol solving ListCommonRandomness$^k_{\gamma,b}$. Let X be Alice's input and $w_A \triangleq f(X)$ be her output, and let Y be Bob's input and $L_B \triangleq (g_1(Y), g_2(Y), \ldots, g_b(Y))$ be his output. Here, $(X,Y) \sim \mathsf{DSBS}(1-2\epsilon)^{\otimes n}$, and f, g_1, g_2, $\ldots$, g_b are functions mapping $\{0,1\}^n$ to $\{0,1\}^k$. For every $y \in \{0,1\}^n$ and $z \in \{0,1\}^k$, denote $\beta(z|y) \triangleq \Pr[f(X) = z|Y=y]$. The success probability of the protocol Π is given by

$$\begin{aligned}
\Pr[w_A \in L_B] &= \Pr[f(X) \in \{g_1(Y), g_2(Y), \ldots, g_b(Y)\}] \\
&= \mathbb{E}_y[\Pr[f(X) \in L_B(y) \mid Y=y] \\
&= \mathbb{E}_y[\sum_{z \in L_B(y)} \beta(z|y)] \\
&\leq \mathbb{E}_y[(\sum_{z \in L_B(y)} \beta(z|y)^q)^{1/q}] \cdot b^{1-1/q} \\
&\leq \mathbb{E}_y[(\sum_z \beta(z|y)^q)^{1/q}] \cdot b^{1-1/q} \\
&\leq (\mathbb{E}_y[\sum_z \beta(z|y)^q])^{1/q} \cdot b^{1-1/q} \\
&= (\sum_z \mathbb{E}_y[\beta(z|y)^q])^{1/q} \cdot b^{1-1/q},
\end{aligned}$$

where the first inequality follows from Holder's inequality and the last inequality follows from the fact that the function $x \mapsto x^{1/q}$ for non-negative x is concave for every $q \geq 1$. Consider the function $h_z \colon \{0,1\}^n \to \{0,1\}$ given by $h_z(X) = \mathbb{1}[f(X) = z]$ for all $X \in \{0,1\}^n$. Hypercontractivity then implies that

$$\begin{aligned}\mathbb{E}_y[\beta(z|y)^q])^{1/q} &= \mathbb{E}_y[\mathbb{E}[h_z(X) \mid Y = y]^q] \\ &= \|\mathbb{E}[h_z(X) \mid Y]\|_q^q \\ &\leq \|h_z\|_p^q \\ &= (\mathbb{E}_x h_z(x))^{q/p} \\ &= \Pr[f(X) = z]^{q/p}.\end{aligned}$$

Thus, the success probability of Π satisfies

$$\begin{aligned}&\Pr[w_A \in L_B] \\ &\leq (\sum_z \Pr[f(X) = z]^{q/p})^{1/q} \cdot b^{1-1/q} \\ &= (\sum_z \Pr[f(X) = z]^{q/p-1} \cdot \Pr[f(X) = z])^{1/q} \cdot b^{1-1/q} \\ &\leq (2^{-k\cdot(\frac{q}{p}-1)} \cdot \sum_z \Pr[f(X) = z])^{1/q} \cdot b^{1-1/q} \\ &= 2^{-k\cdot(q/p-1)\cdot\frac{1}{q}} \cdot b^{1-1/q},\end{aligned}$$

where the inequality above follows from the fact that w_A has min-entropy at least k bits. Setting $p = 1 + (1-2\cdot\epsilon)^2 \cdot \delta$ and $q = 1 + \delta$ and optimizing for δ, we get that

$$\gamma \leq 2^{-k\cdot\frac{[-\sqrt{s}+\sqrt{1-(\log b)/k}]^2}{1-s}},$$

where $s = (1-2\epsilon)^2$ is the Strong Data Processing Constant of the $\mathsf{DSBS}(1-2\epsilon)$ source, and where the above bound holds assuming that $(\log b)/k \leq 1 - s$. The theorem statement now follows.

We point out that Theorem 9.3 implies a lower bound on the 1-way communication complexity of List Common Randomness Generation (by essentially increasing the list size by a factor of 2^c where c is the communication from Alice to Bob). It turns out that, by adapting a reduction of [12], one can also use Theorem 9.3 to get a lower bound on the *interactive* communication complexity of List Common Randomness Generation, which we state next.

THEOREM 9.4. *(Interactive List Common Randomness Generation) Let* $\mathsf{DSBS}(1-2\epsilon)$ *be the source. Then, any interactive protocol solving* $\mathit{ListCommonRandomness}^k_{\gamma,b}$ *with* $(\log b)/k \leq 4\cdot\epsilon\cdot(1-\epsilon)$ *should communicate at least*

$$k\cdot\frac{[\sqrt{1-(\log b)/k}-(1-2\epsilon)]^2}{8\cdot\epsilon\cdot(1-\epsilon)} - \frac{3}{2}\log(1/\gamma) - O(1) \quad \textit{bits.}$$

Consequently, whenever $(\log b)/k \leq 4\cdot\epsilon\cdot(1-\epsilon)/(1+\Omega(1))$*, any interactive protocol solving* $\mathit{ListCommonRandomness}^k_{\gamma,b}$ *should communicate at least* $\Omega(\epsilon\cdot k) - O(\log 1/\gamma)$ *bits.*

Proof. The proof will combine Theorem 9.3 with the approach of [12] for getting lower bounds on *interactive* Common Randomness Generation using lower bounds on *non-interactive* Common Randomness Generation.

Let Π be an interactive protocol solving $\mathrm{ListCommonRandomness}^k_{\gamma,b}$ with $(\log b)/k \leq (1-s)/(1+\Omega(1))$. Let X denote Alice's input and Y denote Bob's input. Consider now the non-interactive protocol Π where on input pair (X,Y):

- Alice samples Y' from the conditional distribution of μ given X, and she outputs the element that she would have output in the execution of Π on (X,Y').
- Bob samples X' from the conditional distribution of μ given Y, and he outputs the list that he would have output in the execution of Π on (X',Y).

Note that the non-inteactive protocol Π' satisfies the property that the min-entropy of Alice's output is at least k (since it is exactly equal to the min-entropy of Alice's output under Π). We next show that the success probability of the protocol Π' is at least $\Omega(\gamma^3\cdot 2^{-2\cdot c})$ where c is the two-way communication complexity of Π. Using Theorem 9.3, this would imply that

$$c \geq k\cdot\frac{[\sqrt{1-(\log b)/k}-(1-2\epsilon)]^2}{8\cdot\epsilon\cdot(1-\epsilon)} - \frac{3}{2}\log(1/\gamma) - O(1),$$

which implies the desired statement. We now lower-bound the success probability of Π'. Let $P_X(t)$ denote the probability over Y' conditioned on X that $\Pi(X,Y')$ is equal to the transcript t. Similarly, let $Q_Y(t)$ denote the probability over X' conditioned on Y that $\Pi(X',Y)$ is equal to the transcript t. Let G be the set of all input pairs (X,Y) such that, in the execution of $\Pi(X,Y)$, Alice's output element belongs to Bob's output list. Then, the success probability of Π is equal to

$$\gamma = \sum_{(X,Y)\in G} \mu(X,Y).$$

We say that a transcript t is unlikely for X if $P_X(t) < (\gamma/4)\cdot 2^{-c}$. Similarly, we say that a transcript t is unlikely for Y if $Q_Y(t) < (\gamma/4)\cdot 2^{-c}$. Let B be the set of all input-pairs (X,Y) such that the transcript $\Pi(X,Y)$

is either unlikely for X or unlikely for Y. Note that

$$\begin{aligned}
(9.17)\quad & \sum_{(X,Y):\ \Pi(X,Y) \text{ unlikely for } X} \mu(X,Y) \\
&= \sum_{X} \sum_{t \text{ unlikely for } X} \sum_{Y:\ \Pi(X,Y)=t} \mu(X,Y) \\
&= \sum_{X} \mu(X) \cdot \sum_{t \text{ unlikely for } X} P_X(t) \\
&< \sum_{X} \mu(X) \cdot \sum_{t \text{ unlikely for } X} \frac{\gamma}{4} \cdot 2^{-c} \\
(9.18)\quad &< \frac{\gamma}{4}.
\end{aligned}$$

An identical argument shows that

$$(9.19)\quad \sum_{(X,Y):\ \Pi(X,Y) \text{ unlikely for } Y} \mu(X,Y) < \frac{\gamma}{4}.$$

Combining Equation (9.17) and Equation (9.19), we get that

$$\sum_{(X,Y)\in B} \mu(X,Y) < \frac{\gamma}{2}.$$

The success probability of Π' can now be lower-bounded by

$$\begin{aligned}
&\sum_{(X,Y)\in G} \mu(X,Y) \cdot P_X(\Pi(X,Y)) \cdot Q_Y(\Pi(X,Y)) \\
&\geq \sum_{(X,Y)\in G\setminus B} \mu(X,Y) \cdot P_X(\Pi(X,Y)) \cdot Q_Y(\Pi(X,Y)) \\
&\geq \sum_{(X,Y)\in G\setminus B} \mu(X,Y) \cdot \frac{\gamma^2}{16} \cdot 2^{-2\cdot c} \\
&= \frac{\gamma^2}{16} \cdot 2^{-2\cdot c} \cdot \left(\sum_{(X,Y)\in G} \mu(X,Y) - \sum_{(X,Y)\in B} \mu(X,Y) \right) \\
&\geq \frac{\gamma^3}{32} \cdot 2^{-2\cdot c},
\end{aligned}$$

as desired.

We note that Theorems 9.3 and 9.4 also hold with the same bounds when the source is $\mathsf{BGS}(1-2\varepsilon)$ instead of $\mathsf{DSBS}(1-2\varepsilon)$. We now show how Theorem 9.3 implies Theorem 9.2, and how Theorem 9.4 implies Theorem 9.1.

Proof. [Proof of Theorem 9.2] Given a protocol Π for $\text{CorrelatedRandomness}_{\gamma,\epsilon',\alpha,k}$, we give a protocol Π' for $\text{ListCommonRandomness}^{\alpha\cdot k}_{\gamma,b}$ with $b \leq 2^{h(\epsilon')\cdot k}$ as follows:

- If w_A is the output of Alice under the protocol Π, then she also outputs w_A under the protocol Π'.
- If w_B is the output of Bob under the protocol Π, then he outputs the list $L_B \triangleq \mathsf{Ball}(w_B, \epsilon'\cdot k)$ under the protocol Π'.

Theorem 9.2 now follows from Theorem 9.3 and the fact that $|\mathsf{Ball}(w_a, \epsilon'\cdot k)| \leq 2^{h(\epsilon')\cdot k}$.

Proof. [Proof of Theorem 9.1] The proof is identical to that of Theorem 9.2 except that we use Theorem 9.4 instead of Theorem 9.3.

10 Communication with Imperfect Shared Randomness

We start by stating the most general result for this problem that applies to many sources of randomness including $\mathsf{DSBS}(\rho)$.

THEOREM 10.1. *Let $\rho \in (0,1]$ and μ be any source of randomness with maximal correlation ρ. Every (possibly partial) function f with $(1/3)$-error two-way communication c bits with perfect randomness has δ-error simultaneous message passing communication with μ-randomness at most $2^{O(c)} \cdot \log(1/\delta)/\rho^2$ bits for every $\delta > 0$.*

We point out that the above theorem yields $\mathsf{DSBS}(\rho)$ as a special case because of the fact (due to [44]) that the maximal correlation of $\mathsf{DSBS}(\rho)$ is equal to ρ.

In order to prove Theorem 10.1, we will give a simultaneous message passing protocol with μ-randomness (where μ is any source of randomness with maximal correlation ρ) solving the following problem which is equivalent to "sketching ℓ_2-norms on the unit sphere." This problem was studied by [12] to prove a 1-way (instead of simultaneous message passing) analogue of Theorem 10.1.

DEFINITION 10.1. ($\text{GAPINNERPRODUCT}^n_{r,s}$) *Let $-1 \leq s < r \leq 1$ be known to Alice and Bob. Alice is also given a unit vector $u \in \mathbb{R}^n$ and Bob is given a unit vector v in $\mathbb{R}^n$. The goal is for Alice and Bob to distinguish the case where $u \bullet v \geq r$ from the case where $u \bullet v \leq s$.*

The next lemma shows that GapInnerProduct is complete for functions with low interactive communication complexity.

LEMMA 10.1. ([12]) *Let f be a (possibly partial) two-party function $f\colon \{0,1\}^{2\cdot n} \to \{0,1\}$, such that f has $(1/3)$-error two-way communication complexity c bits with perfect randomness. Then, there exists a function $\ell(n) \in \mathbb{N}$ along with mappings $g_A\colon \{0,1\}^n \to \{\pm\frac{1}{\sqrt{\ell(n)}}\}^{\ell(n)}$ and $g_B\colon \{0,1\}^n \to \{\pm\frac{1}{\sqrt{\ell(n)}}\}^{\ell(n)}$ such that*

- *If $f(x,y)=0$, then $(g_A(x),g_B(y))$ is a NO instance of GapInnerProduct$^{\ell(n)}_{\frac{2}{3}\cdot 2^{-k}-1,\frac{1}{3}\cdot 2^{-k}-1}$. Namely, $g_A(x)\bullet g_B(y)\leq \frac{1}{3}\cdot 2^{-k}-1$.*
- *If $f(x,y)=1$, then $(g_A(x),g_B(y))$ is a YES instance of GapInnerProduct$^{\ell(n)}_{\frac{2}{3}\cdot 2^{-k}-1,\frac{1}{3}\cdot 2^{-k}-1}$. Namely, $g_A(x)\bullet g_B(y)\geq \frac{2}{3}\cdot 2^{-k}-1$.*

The following theorem gives a simultaneous message passing protocol with μ-randomness for GapInnerProduct (where μ is any source with maximal correlation ρ). It matches the performance of the one-way protocol of [12].

THEOREM 10.2. *(simultaneous message passing protocol for GapInnerProduct$^n_{r,s}$) Let $\rho\in(0,1]$ and $-1\leq s<r\leq 1$ be given, and let μ be any source of randomness with maximal correlation ρ. There is a simultaneous message passing protocol using μ-randomness that solves GapInnerProduct$^n_{r,s}$ using $O(\frac{1}{\rho^2(r-s)^2})$ bits of communication.*

We point out that Theorem 10.2 gives a protocol for *sketching* ℓ_2-norms using imperfectly shared randomness, which might be of independent interest. Theorem 10.1 now follows by combining Lemma 10.1 and Theorem 10.2. In the rest of this section, we prove Theorem 10.2. First, we recall the following observation of [44] which can be used to convert any source μ of randomness with maximal correlation ρ to $\mathsf{BGS}(\rho)$.

PROPOSITION 10.1. ([44]) *Let μ be a source of randomness with maximal correlation ρ. Given access to i.i.d. samples from μ, Alice and Bob can (without interaction) generate i.i.d. samples from $\mathsf{BGS}(\rho)$.*

Proposition 10.1 follows from the definition of maximal correlation and from the two-dimensional Central Limit Theorem. We also recall the following well-known fact.

FACT 10.1. (SHEPPARD'S FORMULA [37]) *If $(X,Y)\sim\mathsf{BGS}(\rho)$ then $\Pr[\mathsf{Sign}(X)\neq\mathsf{Sign}(Y)]=\frac{\arccos(\rho)}{\pi}$.*

The following lemma is based on the well-known hyperplane rounding technique.

LEMMA 10.2. *Let $\delta>0$ and $\gamma<0$ be given, and let $t=O(\log(1/\delta)/\gamma^2)$ be large enough. Let Alice be given $(X_1,X_2,\ldots,X_t)\in\mathbb{R}^t$ and Bob be given $(Y_1,Y_2,\ldots,Y_t)\in\mathbb{R}^t$ where $(X_i,Y_i)\sim\mathsf{BGS}(\eta)$ independently over $i\in[t]$. Then, there is a deterministic simultaneous message passing protocol that distinguishes the case where $\eta\geq 0$ from the case where $\eta\leq\gamma$ using $O(1/\gamma^2)$ bits of communication, and with probability at least $1-\delta$ (where the probability is over $(X_1,X_2,\ldots,X_t)$ and $(Y_1,Y_2,\ldots,Y_t)$).*

Proof. For every $i\in[t]$, Alice computes $\tilde{X}_i=\mathsf{Sign}(X_i)$ and sends the t bits $\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_t$ to the referee. Similarly, Bob computes $\tilde{Y}_i=\mathsf{Sign}(Y_i)$ for each $i\in[t]$, and sends the t bits $\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_t$ to the referee. Let $\tau=(\arccos(\gamma)/\pi-1/2)/2$. The referee computes the Hamming distance $\Delta(\tilde{X},\tilde{Y})$ and declares that $\eta\geq 0$ if $\Delta(\tilde{X},\tilde{Y})\leq\tau$, and declares that $\eta\leq\gamma$ otherwise. Note that if $\eta\geq 0$, then for each $i\in[t]$,

(10.20)
$$\Pr[\mathsf{Sign}(X_i)\neq\mathsf{Sign}(Y_i)]=\frac{\arccos(\eta)}{\pi}\leq\frac{\arccos(0)}{\pi}=\frac{1}{2}.$$

On the other hand, if $\eta\leq\gamma$, then for each $i\in[t]$,

$$\begin{aligned}\Pr[\mathsf{Sign}(X_i)\neq\mathsf{Sign}(Y_i)]&=\frac{\arccos(\eta)}{\pi}\geq\frac{\arccos(\gamma)}{\pi}\\ &=\frac{1}{2}-\Theta(\gamma)-O(\gamma^3),\end{aligned}\tag{10.21}$$

where the last equality follows from the Taylor series approximation of $\arccos(x)$ around $x=0$. The proof now follows by combining Equations (10.20) and (10.21) and an application of the Chernoff bound.

We are now ready to prove Theorem 10.2.

Proof. [Proof of Theorem 10.2]

Alice is given $u\in\mathbb{R}^n$ and Bob is given $v\in\mathbb{R}^n$ such that $\|u\|_2=\|v\|_2=1$. They are also given access to i.i.d. samples from a source μ of randomness with maximal correlation ρ. Using Proposition 10.1, Alice and Bob can (without interaction) generate arbitrarily many i.i.d. samples from $\mathsf{BGS}(\rho)$. We first assume that $r=0$. We will handle the more general case at the end of the proof. Set $\gamma=\rho\cdot s$ and let $t=O(\log(1/\delta)/\gamma^2)$ be as in the statement of Lemma 10.2. Draw t i.i.d vectors $(X^{(1)},Y^{(1)}),(X^{(2)},Y^{(2)}),\ldots,(X^{(t)},Y^{(t)})$ each from $\mathsf{BGS}(\rho)^{\otimes n}$. Then, by elliptical symmetry, we get that independently over $i\in[t]$, $(u\bullet X^{(i)},v\bullet Y^{(i)})\sim\mathsf{BGS}(\rho(u\bullet v))$. Lemma 10.2 now implies a simultaneous message passing protocol that distinguish the case where $u\bullet v\rangle\geq 0$ from the case where $u\bullet v\leq s$, using $O(\frac{1}{\rho^2(r-s)^2})$ bits of communication.

We now handle the case where r is not necessarily equal to 0. First, note that without loss of generality, we can assume that $r\geq 0$. This is because if $r<0$, then Alice can negate each coordinate in her input vector which would preserve its ℓ_2 norm and replace r by $-s\geq 0$ and s by $-r\geq 0$. Let $N\triangleq n\cdot(1+r)$. Bob will construct a vector $u'\in\mathbb{R}^N$, and Alice will construct a vector $v'\in\mathbb{R}^N$, such that $\|u'\|_2=\|v'\|_2=1$, and:

- If $u\bullet v\geq r$, then $u'\bullet v'\geq 0$.
- If $u\bullet v\leq s$, then $u'\bullet v'\leq\frac{s-r}{1+r}=-\Theta(r-s)$.

To do so, Alice sets $u'_i = u_i \cdot \sqrt{n/N}$ for every $i \in [n]$ and $u'_i = +1/\sqrt{N}$ for all $i \in \{n+1, \ldots, N\}$. On the other side, Bob sets $v'_i = v_i \cdot \sqrt{n/N}$ for all $i \in [n]$ and $v'_i = -1/\sqrt{N}$ for all $i \in \{n+1, \ldots, N\}$.

11 LSH and Common Randomness

An important parameter that governs the performance of an LSH hash family $\mathcal{H}$ is given by its $\bar{\rho}(\mathcal{H})$ parameter [24]. Let $0 \leq \alpha \leq 1$ and $c \geq 1$. Loosely speaking, if the hash family ensures that points at relative distance at most α collide with probability at least p_1 while points at relative distance at least $c\alpha$ collide with probability at most p_2, then $\bar{\rho}(\mathcal{H}, \alpha, c) \leq \log(1/p_1)/\log(1/p_2)$. Smaller values of $\bar{\rho}(\mathcal{H}, \alpha, c)$ can potentially lead to improvements in the data structure performance. For the Hamming cube $\{0,1\}^d$, there is a trivial scheme $\mathcal{H}_0$ such that $\bar{\rho}(\mathcal{H}_0) \leq \log(1/(1-\alpha))/\log(1/(1-c\alpha)) \to 1/c$ as $\alpha \to 0$.

We show that the zero-communication common randomness schemes considered here and in previous works [10,22] imply an LSH scheme with a strictly better $\bar{\rho}$ parameter. This is perhaps not surprising since the best strategy for a universal scheme is to map close-by points to the same output in order to achieve high-agreement probability, but to ensure high entropy it must map far-away points to different outputs.

Recall that in the trivial scheme $\mathcal{H}_0$ the hash function just outputs the bit at a random coordinate in $[d]$. When the relative distance between the two points is ε, this is tantamount to producing a single sample from $\mathsf{DSBS}(1-2\varepsilon)$. Thus the trivial LSH scheme is also a trivial common randomness scheme using one sample from $\mathsf{DSBS}(1-2\varepsilon)$. If we use k samples, i.e. take k independent hash functions, and use the trivial scheme we obtain an agreement $p_\rho \coloneqq \left(\frac{1+\rho}{2}\right)^k$. Let $f_0(\rho) = \log(1/p_\rho)/k = \log(2/(1+\rho)$. In contrast, if we use the mapping given by the common randomness scheme then for this hash family (call it $\mathcal{H}_1$), the analogous expression equals $f_{cr}(\rho) \coloneqq (1-\rho)/(1+\rho) + O(\log(k)/k$. For large k we can ignore the lower order term. So let $f_{cr}(\rho) = (1-\rho)/(1+\rho)$. To show that $\bar{\rho}(\mathcal{H}_2)$ is better we need to show for $\rho > \rho'$ that $f_{cr}(\rho)/f_{cr}(\rho') \leq f(\rho)/f(\rho')$. That is, $f(\rho)/f_{cr}(\rho)$ is an increasing function in $[0,1]$. This can be verified analytically. In fact it is always strictly increasing so the bound for the CR scheme is strictly better than the trivial one.