Some Results on Distributed Source Simulation with no Communication

Tomer Berg, Ofer Shayevitz, Young-Han Kim, and Lele Wang

***Abstract*—We consider the problem of distributed source simulation with no communication, in which Alice and Bob observe sequences U^n and V^n respectively, drawn from a joint distribution $p_{UV}^{\otimes n}$, and wish to locally generate sequences X^n and Y^n respectively with a joint distribution that is close (in KL divergence) to $p_{XY}^{\otimes n}$. We provide a single-letter condition under which such a simulation is asymptotically possible with a vanishing KL divergence. Our condition is nontrivial only in the case where the Gàcs-Körner (GK) common information between U and V is nonzero, and we conjecture that only scalar Markov chains $X-U-V-Y$ can be simulated otherwise. Motivated by this conjecture, we further examine the case where both p_{UV} and p_{XY} are doubly symmetric binary sources with parameters $p,q\leq 1/2$ respectively. While it is trivial that in this case $p\leq q$ is both necessary and sufficient, we show that when p is close to q then any successful simulation is close to being scalar in the total variation sense.**

I. Introduction and Main Results

Let us consider the following distributed simulation problem. Assume that (U^n,V^n) are drawn by nature according to some i.i.d distribution p_{UV}. Alice has access to U^n and she outputs some sequence X^n, while Bob has access to V^n and he outputs some sequence Y^n, such that (X^n,Y^n) are approximately distributed according to some i.i.d distribution p_{XY}. There is no communication between the parties nor do they share any common randomness. Our goal is to characterize the set of distributions p_{XY} that can be reliably simulated using this scheme. To make this more formal, Let p_{UV} be some joint discrete distribution, and let $(U^n,V^n)\sim p_{UV}^{\otimes n}$. We say that a joint distribution p_{XY} *is* (n,ϵ)*-simulable from* p_{UV}, if there exist conditional probability distributions $p_{X^n|U^n}$ and $p_{Y^n|V^n}$ such that the distribution

$$\begin{aligned}&p_{X^nY^n}(x^n,y^n)\\&=\sum_{u^n,v^n}p_{UV}^{\otimes n}(u^n,v^n)p_{X^n|U^n}(x^n|u^n)p_{Y^n|V^n}(y^n|v^n)\end{aligned}$$

is ϵ-close in relative entropy to $p_{XY}^{\otimes n}$, i.e.,

$$D\left(p_{X^nY^n}\|p_{XY}^{\otimes n}\right)\leq\epsilon.$$

We say that p_{XY} *is simulable from* p_{UV} if it is (n,ϵ)-simulable from p_{UV} for every $\epsilon>0$ and n sufficiently large.

For $U=V$, our question was already answered by Wyner:

Theorem 1 ([1]). *If $H(U)\geq C_{\mathsf{W}}(X;Y)$, where*

$$C_{\mathsf{W}}(X;Y)\overset{def}{=}\min_{W:X-W-Y}I(X,Y;W)$$

then p_{XY} is simulable from p_U.

$C_{\mathsf{W}}(X;Y)$ is the so-called *Wyner common information*, defined as the minimum common randomness that needed for Alice and Bob to be able to locally create X^n and Y^n respectively, where p_{X^n,Y^n} is arbitrarily close (in KL divergence) to $p_{XY}^{\otimes n}$, in the limit of large n. Note that this solution is "digital", in the sense that it uses codebooks. One naive approach that comes to mind is a reduction to Wyner's setup, by generating a "common part" $f(U)=g(V)$ from $U\neq V$. This corresponds to using the so-called *Gàcs-Körner (GK) common information* [2], which is defined as

$$C_{\mathsf{GK}}(U;V)\overset{\text{def}}{=}\max_{f,g:\,\Pr(f(U)=g(V))=1}H(f(U)).$$

$C_{\mathsf{GK}}(U;V)$ is the maximum amount of randomness that can be agreed upon by two separate agents, Alice and Bob, observing U or V respectively. The (unique) random variable $K=f(U)=g(V)$ that attains the maximum above is called the *GK common part* of (U,V). It is well known that if $(U^n,V^n)\sim p_{UV}^{\otimes n}$, then $C_{\mathsf{GK}}(U^n;V^n)=nC_{\mathsf{GK}}(U;V)$. In other words, the GK common part of (U^n,V^n) is simply the vector of scalar common parts pertaining to each (U_i,V_i). Moreover, this tensorization is asymptotically valid even if a vanishing error is allowed [2].

Combining the two results, Alice and Bob can both extract the GK common part K^n from U^n and V^n respectively, and use Wyner coding, which leads to the following simple solution:

Proposition 1 (digital solution)**.** *If $H(K)=C_{\mathsf{GK}}(U;V)\geq C_{\mathsf{W}}(X;Y)$, then p_{XY} is simulable from p_{UV}.*

This digital approach is viable only when $C_{\mathsf{GK}}(U;V)>0$. There is an even simpler analog approach that does not use common information – Alice and Bob pass their corresponding sequences through memoryless channels $p_{X^n|U^n}=p_{X|U}^{\otimes n}$ and $p_{Y^n|V^n}=p_{Y|V}^{\otimes n}$, respectively, symbol-by-symbol.

Proposition 2 (analog solution)**.** *If $X-U-V-Y$ form a Markov chain, then p_{XY} is simulable from p_{UV}.*

The first contribution of this work is the following characterization of a generally larger set of simulable distributions, which we prove in Section III.

Theorem 2. *Let K be the GK common part of (U,V). Suppose*

$$W-K-(U,V) \tag{1}$$
$$X-(U,W)-(V,W)-Y \tag{2}$$

are Markov chains, and

$$C_{\mathsf{GK}}(U;V)\geq I(X,Y;K,W).$$

Then p_{XY} is simulable from p_{UV}.

978-1-5386-6900-6/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 17,2022 at 17:41:41 UTC from IEEE Xplore. Restrictions apply.

We note that the theorem easily implies both analog and digital approaches. For the analog approach, choose W to be independent of (U,V,X,Y). The Markov chain (1) is satisfied and also $I(X,Y;K,W) = I(X,Y;K) \leq H(K) = C_{\mathsf{GK}}(U;V)$ holds. The only additional condition is the Markov chain (2), which in this case reduces to $X - U - V - Y$. A proper choice of W also implies the digital solution.

Let $\mathcal{S}_{\mathrm{dig}}(p_{UV})$, $\mathcal{S}_{\mathrm{ana}}(p_{UV})$, and $\mathcal{S}(p_{UV})$ denote the collections all p_{XY} simulable from p_{UV} via a digital scheme (Proposition 1), an analog scheme (Proposition 2) and a hybrid scheme (Theorem 2), respectively. The following proposition shows that the statement of Theorem 2 is not trivial.

Proposition 3. *$\mathcal{S}_{\mathrm{dig}}(p_{UV}) \cup \mathcal{S}_{\mathrm{ana}}(p_{UV}) \subseteq \mathcal{S}(p_{UV})$, and the inclusion is strict for some p_{UV}. Moreover, $\mathcal{S}(p_{UV})$ is strictly larger than $\mathcal{S}_{\mathrm{ana}}(p_{UV})$ if and only if $C_{\mathsf{GK}}(U;V) > 0$.*

In Theorem 2, our agents ability to cooperate stems from having some common information. No common part means no perfect cooperation, and this motivates us to conjecture that only analog simulation is possible in such a case.

Conjecture 1. *If $C_{\mathsf{GK}}(U;V) = 0$ then p_{XY} is simulable from p_{UV} if and only if there exists a joint distribution p_{XYUV} such that $X - U - V - Y$ form a Markov chain.*

We are currently unable to prove or refute this conjecture. Note, however, that in some simple restricted cases the conjecture holds due to other impossibility results. For example, if (U,V) is a $\mathsf{DSBS}(p)$ for some $p < 1/2$ (hence $C_{\mathsf{GK}}(U;V) = 0$) and we are only interested in simulating $\mathsf{DSBS}(q)$, then it is easy to see that $q \in [p, 1-p]$ is both necessary and sufficient, and can be attained by a scalar Markov chain. Our next result shows that this is true in a stronger way; namely, when q is close to p, then not only is the scalar Markov chain optimal, but it is essentially the only way to simulate a $\mathsf{DSBS}(q)$.

Let σ be a permutation on $[n]$. With some abuse of notation, we refer to σ as a *coordinate permutation* when applied to any n-vector, i.e., $\sigma(x^n) \stackrel{\text{def}}{=} (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$. We write $d_{\mathsf{TV}}(P,Q)$ to denote the total variation distance between the probability distributions P and Q.

Theorem 3. *Let (U,V) and (X,Y) be $\mathsf{DSBS}(p)$ and $\mathsf{DSBS}(p+\delta)$ respectively, where $0 \leq p \leq p+\delta \leq \frac{1}{2}$. Suppose that p_{XY} is (n,ϵ)-simulable from p_{UV} via $p_{X^n|U^n}$ and $p_{Y^n|V^n}$. Then there exists a coordinate permutation σ and scalar conditional distributions $q_{X_i|U_i}$ and $q_{Y_i|V_i}$ such that*

$$D_1 \triangleq d_{\mathsf{TV}}\left(p_{\sigma(X^n)|U^n}(\cdot \mid U^n), \prod_{i=1}^{n} q_{X_i|U_i}(\cdot|U_i)\right) \to 0,$$

$$D_2 \triangleq d_{\mathsf{TV}}\left(p_{\sigma(Y^n)|V^n}(\cdot \mid V^n), \prod_{i=1}^{n} q_{Y_i|V_i}(\cdot|V_i)\right) \to 0,$$

as $n \to \infty$ in probability, provided that $\epsilon, \delta = o(1/n)$. Conversely, if $\delta = \omega(1/\sqrt{n})$ or $\epsilon = \omega(1/\sqrt{n})$ then no such guarantee can be made, i.e., it is possible for D_1, D_2 to be bounded away from zero in probability as $n \to \infty$ for any scalar conditional distributions $q_{X_i|U_i}$ and $q_{Y_i|V_i}$.

Loosely speaking, the above result means that if δ and ϵ are $o(1/n)$, then the actual mechanism under the hood of any successful simulation scheme is truly scalar, in the sense that no statistical experiment can tell it apart from a scalar one. We note that Theorem 3 is well known in combinatorics for the case where $\epsilon = \delta = 0$ (see e.g. [3]), hence our result can be interpreted as a stable version of that combinatorial fact. Furthermore, our result is close to being tight; when $\epsilon = \omega(1/\sqrt{n})$ or $\delta = \omega(1/\sqrt{n})$, successful simulation is possible using vector operations, for example by using other coordinates as noise. In the following, note that all skipped/abbreviated proofs are available in an extended version of the paper [4].

II. Related work

In their classical paper on common randomness generation [5], Ahlswede and Körner considered a setup in which Alice and Bob observe correlated i.i.d. r.v. pairs, and a noiseless channel with capacity R from Alice to Bob is given. They defined the so-called *CR capacity* as the maximum entropy rate that Alice and Bob can agree upon with probability approaching one. The case of $R = 0$ is related to our problem, but their setup is in some sense weaker since they only care about generating randomness, and not about simulating specific distributions. Cuff *et al.* [6] studied the joint distributions that can be generated by nodes in a network under communication constraints in which some of the nodes actions are randomly selected by nature. The main concern in the paper is the notion of *empirical coordination*, which is the total variation between the joint type of the actions and some prescribed distribution. Cuff [7], [8] considered the problem of channel simulation, where an i.i.d sequence X^n is available to Alice, who can send a rate R quantized description of X^n to Bob, and rate R_0 common randomness is shared between the parties. He fully characterized the rate pairs (R, R_0) for which Bob can generate Y^n such that the channel from X^n to Y^n is arbitrarily close in total variation to a given memoryless channel. Haddadpour *et al.* [9] studied similar problem, but where the channel from Alice to Bob is a noisy memoryless one, instead of a bit pipe. Ghazi *et al.* [10] and De *et al.* [11] studied the computational-theoretic problem of deciding whether certain distributions can be simulated from a given sequence of i.i.d. pairs in a setup similar to ours (but when the target distributions are more general), and gave conditions for decidability.

III. Proofs of Theorem 2 and Proposition 3

The proof of Theorem 2 is based on the following "soft-covering" lemma.

Lemma 1. *Let p_{XUW} be given and $U^n \sim p_U^{\otimes n}$. If $H(U) > I(X;U,W)$, then there exists a sequence $a_n : \mathcal{U}^n \to \mathcal{W}^n$, such that if we draw $X^n \sim p_{X|UW}^{\otimes n}(\cdot \mid U^n, a_n(U^n))$ then*

$$\lim_{n\to\infty} D\left(p_{X^n} \parallel p_X^{\otimes n}\right) = 0.$$

This lemma was proved by Cuff [7] for a weaker convergence in total variation. However, as noted in [7], an inequality

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 17,2022 at 17:41:41 UTC from IEEE Xplore. Restrictions apply.

from [12] can be used to show that in this case convergence in total variation also implies convergence in KL divergence. We note that it is also possible to prove the latter directly.

We proceed to prove Theorem 2. Our construction is based on *hybrid coding* in the spirit of [13], [14]. We use the GK common part as the digital part, and U^n (resp. V^n) as the analog part. Alice and Bob both remotely compute the GK common part K^n of (U^n, V^n) from their respective components, and create $W^n = a(K^n)$ using some encoding $a : \mathcal{K}^n \to \mathcal{W}^n$. Alice then generates $X^n \sim p_{X|UW}^{\otimes n}(\cdot \mid U^n, a(K^n))$ and Bob generates $Y^n \sim p_{Y|VW}^{\otimes n}(\cdot \mid V^n, a(K^n))$ using local randomness. This setup is depicted in Figure 1, which also hints that a good protocol combines both the common information between U and V and the structure of the channel between them in the simulation process.

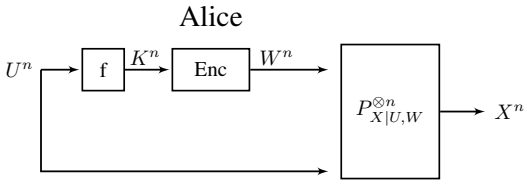

Figure 1: Theorem 2 - Hybrid coding

Define $\widetilde{U}^n = (U^n, V^n)$ and $\widetilde{X}^n = (X^n, Y^n)$. Note that we cannot use Lemma 1 directly on $\widetilde{X}^n, \widetilde{U}^n, W^n$, since W^n is generated from K^n and not from the entire $\widetilde{U}^n$. Instead, we show that $\widetilde{X}^n$ is generated from K^n and $W^n = a(K^n)$ in a memoryless fashion via $p_{X|KW}^{\otimes n}$.

$$\begin{aligned}
p_{\widetilde{X}^n}(\widetilde{x}^n) &= \sum_{\widetilde{u}^n, k^n} p_{\widetilde{U}K}^{\otimes n}(\widetilde{u}^n, k^n) p_{\widetilde{X}|\widetilde{U}W}^{\otimes n}(\widetilde{x}^n|\widetilde{u}^n, a(k^n)) \\
&= \sum_{k^n} p_K^{\otimes n}(k^n) \sum_{\widetilde{u}^n} p_{\widetilde{U}|K}^{\otimes n}(\widetilde{u}^n|k^n) p_{\widetilde{X}|\widetilde{U}W}^{\otimes n}(\widetilde{x}^n|\widetilde{u}^n, a(k^n)) \\
&= \sum_{k^n} p_K^{\otimes n}(k^n) \sum_{\widetilde{u}^n} p_{\widetilde{U}|KW}^{\otimes n}(\widetilde{u}^n|k^n, a(k^n)) \\
&\quad \times p_{\widetilde{X}|\widetilde{U}WK}^{\otimes n}(\widetilde{x}^n|\widetilde{u}^n, a(k^n), k^n) \\
&= \sum_{k^n} p_{K^n}(k^n) \sum_{\widetilde{u}^n} p_{\widetilde{X}\widetilde{U}|K,W}^{\otimes n}(\widetilde{x}^n, \widetilde{u}^n|k^n, a(k^n)) \\
&= \sum_{k^n} p_K^{\otimes n}(k^n) p_{\widetilde{X}|KW}^{\otimes n}(\widetilde{x}^n|k^n, a(k^n)),
\end{aligned}$$

where we have used the fact that $\widetilde{U} - K - W$ and $\widetilde{X} - (\widetilde{U}, W) - K$ are Marokv chains. Applying Lemma 1 with $(X, U, W) \leftarrow (\widetilde{X}, K, W)$, we find that if $H(K) > I(\widetilde{X}; K, W) = I(X, Y; K, W)$, then there exists encodings such that the statement of the Theorem holds.

We now proceed to prove Proposition 3. Due to lack of space, we only show the second claim of the proposition. Assume first that $C_{\mathsf{GK}}(U; V) = 0$. Then clearly W is independent of (U, V) and also independent of (X, Y). Now fix any $w_0 \in \mathcal{W}$, and write

$$\begin{aligned}
p_{XY}(x, y) &= p_{XY|W}(x, y|w_0) \\
&= \sum_{u,v} p_{UV}(u, v) p_{XY|UVW}(x, y|u, v, w_0) \\
&= \sum_{u,v} p_{UV}(u, v) p_{X|UW}(x|u, w_0) p_{Y|VW}(y|v, w_0)
\end{aligned}$$

Hence, considering the r.v.s $(\tilde{X}, \tilde{Y})$ generated via

$$\begin{aligned}
\tilde{p}_{\tilde{X}|U}(\tilde{x}|u) &\stackrel{\text{def}}{=} p_{X|U,W}(\tilde{x}|u, w_0) \\
\tilde{p}_{\tilde{Y}|V}(\tilde{y}|v) &\stackrel{\text{def}}{=} p_{Y|V,W}(\tilde{y}|v, w_0),
\end{aligned}$$

we have that $\tilde{X} - U - V - \tilde{Y}$ forms a Markov chain, and also $(\tilde{X}, \tilde{Y}) \sim p_{XY}$. Hence, the distributions guaranteed by the theorem are only scalar Markov in this case.

Now suppose that $C_{\mathsf{GK}}(U; V) = \epsilon > 0$, and let us show that the Theorem covers more than scalar Markov distributions. Consider the set of simulable distributions generated by some scalar Markov chain $X - U - V - Y$. Each of these distributions can be written in matrix form as $\mathbf{P}_{XY} = \mathbf{P}_{X|U} \mathbf{P}_{UV} \mathbf{P}_{Y|V}^T$, hence in particular, recalling that $\mathrm{rank}(AB) \le \min(\mathrm{rank}(A), \mathrm{rank}(B))$, it must hold that $\mathrm{rank}(\mathbf{P}_{XY}) \le \mathrm{rank}(\mathbf{P}_{UV})$. Now, appealing to the digital approach, it suffices to show that there exists a Markov chain $X - W - Y$ such that $\mathrm{rank}(\mathbf{P}_{XY}) \ge \mathrm{rank}(\mathbf{P}_{UV})$. To that end, choose W to have support over an alphabet of cardinality $M > \mathrm{rank}(\mathbf{P}_{UV})$, and let $|\mathcal{X}| = |\mathcal{Y}| = M$ as well. The Markov structure implies that $\mathbf{P}_{XY} = \mathbf{P}_{X|W} \mathbf{P}_W \mathbf{P}_{Y|W}^T$. Since $\mathrm{rank}(\mathbf{P}_W) = M$ by construction, it suffices to show one can choose $\mathbf{P}_{X|W}$ and $\mathbf{P}_{X|W}$ to have full rank, while keeping $I(X, Y; W) \le I(X; W) + I(Y; W) \le \epsilon$. This is a consequence of the fact that mutual information is continuous w.r.t. the L^∞ metric, whereas matrix rank is not (the details are omitted).

IV. Proof of Theorem 3

We prove the theorem in steps, starting with deterministic schemes and then moving to randomized schemes. Before we proceed, we provide some necessary background.

A. Boolean Functions and Fourier Analysis

A Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ on the Hamming cube can be uniquely expressed [15] as

$$f(u^n) = \sum_{S \subseteq [n]} \hat{f}_S u^S$$

where $u^S = \prod_{i \in S} u_i$. This is the *Fourier expansion* of f w.r.t. the orthonormal basis of parity functions $\left(U^S\right)_{S \subseteq [n]}$, and the real numbers $\hat{f}(S)$ are called the *Fourier coefficients* of f. Note that $\hat{f}_\emptyset = \mathbb{E}[f] = 2\Pr(f = 1) - 1$. The *Fourier weight of f at degree k* is defined as

$$W^k[f] = \sum_{|S| = k} \hat{f}_S^2,$$

hence $\sum_{k=0}^n W^k[f] = \sum_{S \subseteq [n]} \hat{f}_S^2 = 1$.

Lemma 2. *Let $(U^n, V^n) \sim p_{UV}^{\otimes n}$ where p_{UV} is $\mathsf{DSBS}(p)$. Then*

$$\mathbb{E}[f(U^n) g(V^n)] \le \frac{1}{2} \sum_{k=0}^n (1 - 2p)^k (W^k[f] + W^k[g]).$$

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 17,2022 at 17:41:41 UTC from IEEE Xplore. Restrictions apply.

B. Deterministic Schemes with $\delta = 0$

We begin by limiting our discussion to deterministic simulation schemes, i.e., where X^n (resp. Y^n) is a deterministic function of U^n (resp. V^n).

1) Exact Simulation ($\epsilon = 0$): Let $X^n = f(U^n)$ and $Y^n = g(V^n)$ be such that $p_{X^nY^n} = p_{XY}^{\otimes n}$, where p_{XY} is a $\mathsf{DSBS}(p)$. In particular, X^n (resp. Y^n) are uniformly distributed over the entire Hamming cube, hence it is clear that f and g must be permutations of the Hamming cube. Thus on the one hand, by assumption, we have that

$$p_{X^nY^n}(x^n, y^n) = 2^{-n} p^{d_H(x^n,y^n)} (1-p)^{n - d_H(x^n,y^n)}$$

and on the other hand

$$\begin{aligned} &p_{X^nY^n}(x^n, y^n) = \Pr\left(f(U^n) = x^n, g(V^n) = y^n\right) \\ &= 2^{-n} p^{d_H(f^{-1}(x^n), g^{-1}(y^n))} (1-p)^{n - d_H(f^{-1}(x^n), g^{-1}(y^n))} \end{aligned}$$

hence it must be that $d_H(x^n, y^n) = d_H(f(x^n), g(y^n))$ for any $x^n, y^n \in \{0,1\}^n$. Substituting $y^n = x^n$ in the above, we see that $d_H(f(x^n), g(x^n)) = 0$ for any x^n. We thus conclude that $f = g$ must hold. The problem is now reduced to establishing the following Lemma.

Lemma 3. *A bijection $f : \{0,1\}^n \to \{0,1\}^n$ preserves the Hamming distance if and only if f is a signed coordinate permutation.*

Proof. As mentioned, this is well known, but we nevertheless provide a short proof. A signed coordinate permutation is clearly a bijection that preserves the Hamming distance. To prove the other direction, assume first that $f(0^n) = 0^n$. Then it must be that f preserves the Hamming weight, and specifically, it permutes the vectors of weight one, hence it must be a coordinate permutation. The case where $f(0^n)$ is mapped to any other nonzero vector is similar, with the exception that f is now a signed coordinate permutation, flipping exactly those coordinates where $f(0^n)$ is one. □

2) Almost Exact Simulation ($\epsilon > 0$): Now we allow the KL divergence between the simulated distribution and a $\mathsf{DSBS}(p)$ to be at most ϵ, and show that both the functions $f(x^n)$ and $g(y^n)$ will be almost equal to the same signed permutation.

First, we expand the divergence into two nonnegative quantities: one that captures the deviation of the sequence from being i.i.d., and the other that captures the deviation of the marginals from p_{XY}. The (straightforward) proof is omitted.

Lemma 4. *It holds that*

$$\begin{aligned} D\left(p_{X^nY^n} \| p_{XY}^{\otimes n}\right) = &\sum_{i=1}^{n} I\left(X_i, Y_i; X^{i-1}, Y^{i-1}\right) \\ &+ \sum_{i=1}^{n} D\left(p_{X_iY_i} \| p_{XY}\right). \end{aligned}$$

Next, we provide a useful lower bound on the KL divergence between the simulation $P_{X^nY^n}$ and the desired i.i.d. distribution P_{XY}^n, in terms of the expected Hamming distance.

Lemma 5. *Let $p_{UV} = p_{XY}$ be $\mathsf{DSBS}(p)$. Let $X^n = f(U^n)$ and $Y^n = g(V^n)$. Then*

$$D\left(p_{X^nY^n} \| p_{XY}^{\otimes n}\right) \geq \log \frac{1-p}{p} \cdot \left(\mathbb{E}\, d_H(X^n, Y^n) - np\right).$$

with equality if and only if both f and g are bijections.

Proof. Write

$$\begin{aligned} &D\left(p_{X^nY^n} \| p_{XY}^{\otimes n}\right) = \sum_{x^n, y^n} p_{X^nY^n}(x^n, y^n) \log \frac{p_{X^nY^n}(x^n, y^n)}{p_{XY}^{\otimes n}(x^n, y^n)} \\ &= \sum_{x^n, y^n} \left(\sum_{u^n \in f^{-1}(x^n), v^n \in g^{-1}(y^n)} p_{UV}^{\otimes n}(u^n, v^n) \right) \\ &\times \log \frac{\sum_{u^n \in f^{-1}(x^n), v^n \in g^{-1}(y^n)} p_{UV}^{\otimes n}(u^n, v^n)}{p_{XY}^{\otimes n}(x^n, y^n)} \\ &\geq \sum_{u^n, v^n} p_{UV}^{\otimes n}(u^n, v^n) \log \frac{p_{UV}^{\otimes n}(u^n, v^n)}{p_{XY}^{\otimes n}(f(u^n), g(v^n))} \\ &= \sum_{u^n, v^n} p_{UV}^{\otimes n}(u^n, v^n) \log \frac{p^{d_H(u^n, v^n)} q^{n - d_H(u^n, v^n)}}{p^{d_H(f(u^n), g(v^n))} q^{n - d_H(f(u^n), g(v^n))}} \\ &= \log \frac{1-p}{p} \cdot \mathbb{E}\, d_H(f(U^n), g(V^n)), \end{aligned}$$

where $q = 1 - p$. It is easy to see that the inequality holds with equality if and only if both f and g are bijections. □

The next lemma follows easily from Lemma 4, the data-processing inequality, and Pinsker's inequality.

Lemma 6. *The following two claims hold:*

(i) $\frac{1}{2 \ln 2} \sum_{i=1}^{n} \left(\mathbb{E}(X_i Y_i) - (1 - 2p_{X \neq Y})\right)^2 \leq \epsilon$

(ii) $\frac{1}{2 \ln 2} \sum_{i=1}^{n} \mathbb{E}^2(X_i) \leq \epsilon$

Let us now write $X_i = f_i(U^n), Y_i = g_i(V^n)$, where f_i, g_i are the Boolean functions generating the ith coordinate in the respective sequences. First, from Lemma 6 and the fact that $W^0[f] = \hat{f}_\emptyset$, it is clear that these functions are close to being unbiased, i.e. $\sum_{i \in [n]} W^0[f_i] + W^0[g_i] \leq \epsilon \cdot 4 \ln 2$.

Next, we show that most of the energy of each of these functions is concentrated on the first level.

Lemma 7. *It holds that*

$$W^1[f_i] = 1 - \varepsilon_i^f, \quad W^1[g_i] = 1 - \varepsilon_i^g,$$

where $\varepsilon_i^f, \varepsilon_i^g \geq 0$ and

$$\sum_{i \in [n]} \left(\varepsilon_i^f + \varepsilon_i^g\right) \leq \frac{2\epsilon}{p(1-2p)}.$$

Proof. Write

$$\begin{aligned} \epsilon &\geq D(p_{X^nY^n} \| p_{XY}^{\otimes n}) \\ &\geq \mathbb{E}\left[d_H(f(U^n), g(V^n))\right] - np \\ &= \sum_{i=1}^{n} \left(\Pr\left(f_i(U^n) \neq g_i(V^n)\right) - p\right) \\ &= \frac{1}{2} \sum_{i=1}^{n} \left(1 - 2p - \mathbb{E}[f_i(U^n) g_i(V^n)]\right) \end{aligned} \tag{3}$$

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 17,2022 at 17:41:41 UTC from IEEE Xplore. Restrictions apply.

$$\geq \frac{1}{2}\sum_{i=1}^{n}\left(1-2p-\frac{1}{2}\sum_{k=0}^{n}(1-2p)^k\left(W^k[f_i]+W^k[g_i]\right)\right) \tag{4}$$

$$\geq -\epsilon\cdot\ln 2+\frac{1-2p}{2}\sum_{i=1}^{n}\left(1-\frac{1}{2}\left(W^1[f_i]+W^1[g_i]\right)\right.$$
$$\left.-(1-2p)\left(1-\frac{1}{2}\left(W^1[f_i]+W^1[g_i]\right)\right)\right) \tag{5}$$
$$=-\epsilon\cdot\ln 2+p(1-2p)\sum_{i=1}^{n}\left(1-\frac{1}{2}\left(W^1[f_i]+W^1[g_i]\right)\right)$$

where (3) follows from Lemma 5, (4) follows from Lemma 2 and (5) follows since the functions are close to being unbiased and $\sum_{k=2}^{n}(1-2p)^k W^k[f]\leq(1-2p)^2\sum_{k=2}^{n}W^k[f]\leq(1-2p)^2\left(1-W^1[f]\right)$. Hence, we have

$$\sum_{i=1}^{n}\left(W^1[f_i]+W^1[g_i]\right)\geq 2n-\frac{1+\ln 2}{p(1-2p)}\cdot\epsilon.$$

The claim follows by recalling that since f_i, g_i are Boolean functions, then $W^1[f_i], W^1[g_i]\leq 1$ (and using $\ln 2<1$) $\square$

Now, appealing to the Friedgut-Kalai-Naor (FKN) Theorem [16], we conclude that f_i, g_i are close to some dictator function (a function dominated by one coordinate), i.e., for any $i\in[n]$ there exist $k_i,\ell_i\in[n]$ and $a_i,b_i\in\{-1,1\}$ such that

$$\Pr(X_i\neq a_iU_{k_i})\leq K\cdot\varepsilon_i^f,\quad \Pr(Y_i\neq b_iV_{\ell_i})\leq K\cdot\varepsilon_i^g$$

for some universal constant K. Our penultimate step is to show that the functions f and g are close to the sample signed permutation of the Hamming cube. In order to show that the mappings $k_i\leftarrow i$ and $\ell_i\leftarrow i$ are bijections from $[n]$ to $[n]$, assume without loss of generality that $k_1=k_2=1$, and that $a_1=a_2=1$. FKN Theorem then states that both X_1 and Y_1 are equal to U_1 with high probability, which implies that they are equal with high probability, in contradiction to Lemma 4, which implies $I(X_1;X_2)\leq\epsilon$. Next, in order to show that $k_i=\ell_i$ and $a_i=b_i$, without loss of generality we need to consider two cases: First, assume that $k_1=1,\ell_1=2$ and $a_1=b_1=1$. FKN Theorem then states that $X_1=U_1$, $Y_1=V_2$ with high probability, and since $\Pr(U_1\neq V_2)=\frac{1}{2}$, it follows that $\Pr(X_1\neq Y_1)$ is about $\frac{1}{2}$. The second case is where $k_1=\ell_1=1$ but $a_1=1,b_1=-1$. Now, $X_1=U_1$ and $Y_1=-V_1$ with high probability, implying that $\Pr(X_1\neq Y_1)$ is about $1-p$. However, both of these cases contradict Lemma 6, which states implicitly that $\left(\Pr(X_1\neq Y_1)-p\right)^2\leq\epsilon$.

The only thing left to show is the relation to the total variation distance. Assume without loss of generality that the coordinate permutation induced by $k_i\leftarrow i$ is the identity one, i.e., that $k_i=i$, and that $a_i=1$ for all i. Set the scalar noiseless channels $q_{X_i|U_i}(x_i|v_i)=\mathbb{1}(x_i=u_i)$ and $q_{Y_i|V_i}(y_i|v_i)=\mathbb{1}(y_i=v_i)$. The TV distance is then

$$\mathbb{E}\,d_{\mathsf{TV}}\left(p_{X^n|U^n}(\cdot\mid U^n),\prod_{i=1}^{n}q_{X_i|U_i}(\cdot|U_i)\right)$$
$$=\Pr(X^n\neq U^n)\leq\sum_{i=1}^{n}\Pr\left(X_i\neq U_i\right)\leq\frac{2K\epsilon}{p(1-2p)}.$$

By appealing to Markov's inequality, we conclude the proof.

C. Randomized Schemes- sketch of proof

In the following, we give a brief overview of the randomized scheme, in which Alice and Bob are allowed to use local randomness (the complete proof is available in the extended paper). The functional representation lemma [17], states that $p_{X^n|U^n}$ and $p_{Y^n|V^n}$ can be replaced by random functions, i.e., we can write $X_i=f_i(U^n,A),Y_i=g_i(V^n,B)$ where A,B and (U^n,V^n) are mutually independent, and where $f_i(\cdot,a),g_i(\cdot,b)$ are Boolean functions. From the DPI and tensorization properties of the *maximal correlation* between X and Y, $\rho_m(X;Y)$, combined with the fact that the maximal correlation of a $\mathsf{DSBS}(p)$ is $1-2p$, we have that $\rho_m(X_i;Y_i|A,B)\leq 1-2p$. This allows us to show that, with high probability over the local randomness, $f_i(\cdot,a),g_i(\cdot,b)$ are close to some dictator functions, which implies Theorem 3 for $\delta=0$, given that $\epsilon=o(1/n)$. Then, by allowing δ to be of the same order, we can incorporate it into the simulation distortion ϵ and conclude the proof.

References

[1] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[2] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[3] R. Frucht, "On the groups of repeated graphs," *Bulletin of the American Mathematical Society*, vol. 55, no. 4, pp. 418–420, 1949.

[4] T. Berg, O. Shayevitz, Y.-H. Kim, and L. Wang, "Distributed source simulation with no communication," *arXiv preprint arXiv:1906.06970*, 2019. Available online: https://arxiv.org/abs/1906.06970.

[5] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[6] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[7] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[8] P. Cuff, "Communication requirements for generating correlated random variables," in *2008 IEEE International Symposium on Information Theory*, pp. 1393–1397, July 2008.

[9] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref, "Simulation of a channel with another channel.," *IEEE Trans. Information Theory*, vol. 63, no. 5, pp. 2659–2677, 2017.

[10] B. Ghazi, P. Kamath, and M. Sudan, "Decidability of non-interactive simulation of joint distributions," *CoRR*, vol. abs/1607.04322, 2016.

[11] A. De, E. Mossel, and J. Neeman, "Non interactive simulation of correlated distributions is decidable," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2728–2746, SIAM, 2018.

[12] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[13] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding.," *IEEE Trans. Information Theory*, vol. 61, no. 4, pp. 1509–1523, 2015.

[14] R. Soundararajan and S. Vishwanath, "Hybrid coding for gaussian broadcast channels with gaussian sources," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 2790–2794, IEEE, 2009.

[15] R. O'Donnell, *Analysis of boolean functions*. Cambridge University Press, 2014.

[16] E. Friedgut, G. Kalai, and A. Naor, "Boolean functions whose fourier transform is concentrated on the first two levels," *Advances in Applied Mathematics*, vol. 29, no. 3, pp. 427–437, 2002.

[17] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 17,2022 at 17:41:41 UTC from IEEE Xplore. Restrictions apply.