2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Decidability of Non-Interactive Simulation of Joint Distributions

Badih Ghazi
CSAIL, MIT
Cambridge, MA
badih@mit.edu

Pritish Kamath
CSAIL, MIT
Cambridge, MA
pritish@mit.edu

Madhu Sudan
SEAS, Harvard University
Cambridge, MA
madhu@cs.harvard.edu

***Abstract*—We present decidability results for a sub-class of "non-interactive" simulation problems, a well-studied class of problems in information theory. A *non-interactive simulation* problem is specified by two distributions $P(x, y)$ and $Q(u, v)$: The goal is to determine if two players, Alice and Bob, that observe sequences X^n and Y^n respectively where $\{(X_i, Y_i)\}_{i=1}^n$ are drawn i.i.d. from $P(x, y)$ can generate pairs U and V respectively (without communicating with each other) with a joint distribution that is arbitrarily close in total variation to $Q(u, v)$. Even when P and Q are extremely simple: e.g., P is uniform on the triples $\{(0, 0), (0, 1), (1, 0)\}$ and Q is a "doubly symmetric binary source", i.e., U and V are uniform ± 1 variables with correlation say 0.49, it is open if P can simulate Q.**

In this work, we show that whenever P is a distribution on a finite domain and Q is a 2×2 distribution, then the non-interactive simulation problem is *decidable*: specifically, given $\delta > 0$ the algorithm runs in time bounded by some function of P and δ and either gives a non-interactive simulation protocol that is δ-close to Q or asserts that no protocol gets $O(\delta)$-close to Q. The main challenge to such a result is determining explicit (computable) convergence bounds on the number n of samples that need to be drawn from $P(x, y)$ to get δ-close to Q. We invoke contemporary results from the analysis of Boolean functions such as the invariance principle and a regularity lemma to obtain such explicit bounds.

I. Introduction

Given a sequence of independent samples $(x_1, y_1), (x_2, y_2), \ldots$ from a joint distribution P on $\mathcal{A} \times \mathcal{B}$ where Alice observes $x_1, x_2, \ldots$ and Bob observes $y_1, y_2, \ldots$, what is the largest correlation that they can extract if Alice applies some function to her observations and Bob applies some function to his? The continuous version of this question – where the extracted correlation is required to be in *Gaussian* form – was solved by Witsenhausen in 1975 who gave (roughly) a $\mathrm{poly}(|\mathcal{A}|, |\mathcal{B}|, \log(1/\delta))$-time algorithm that estimates the best such correlation up to an additive δ [1]. When the target distribution is Gaussian, the best possible correlation that is attainable is exactly the well-known "maximal correlation coeffcient" which was first introduced by Hirschfeld [2] and Gebelein [3] and then studied by Rényi [4]. However, when the target distribution is not Gaussian, the best correlation is not well-understood and this is the question explored in this paper. Specifically, we study the Boolean version of this question where the extracted correlation is required to be in the form of bits with fixed specified marginals. We give an algorithm that, given $\delta > 0$, computes the best such correlation up to an additive δ.

Questions such as the above are well-studied in the information theory literature under the label of "Non-Interactive Simulation". The roots of this exploration go back to classical works by Gács and Körner [5] and Wyner [6]. In this line of work, the problem is described by a source distribution $P(X, Y)$ and a target distribution $Q(U, V)$ and the goal is to determine the maximum rate at which samples of P can be converted into samples of Q. (So the goal is to start with n samples from P and generate $R \cdot n$ samples from Q, for the largest possible R.) Gács and Körner considered the special case where Q required the output to be a pair of identical uniformly random bits, i.e., $U = V = \mathrm{Ber}(1/2)$ and introduced what is now known as the *Gács-Körner common information* of $P(X, Y)$ to characterize the maximum rate in terms of this quantity. Wyner, on the other hand considered the "inverse" problem where $X = Y = \mathrm{Ber}(1/2)$ and Q was arbitrary. Wyner characterized the best possible conversion rate in this setting in terms of what is now known as the *Wyner common information* of $Q(U, V)$. There is a rich history of subsequent work (see, for instance, [7] and the references within) exploring more general settings where neither P nor Q produces identical copies of some random variable. In such settings, even the question of when can the rate be positive is unknown and this is the question we explore in this paper.

The Non-Interactive Simulation problem is also a generalization of the Non-Interactive Correlation Distillation problem which was studied by [8], [9][1]. Our setup can be thought of as a "positive-rate" version of the setup of Gács and Körner. Namely, for a known

[1] which considered the problem of maximizing agreement on a single bit, in various multi-party settings.

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.65

source distribution $P(X, Y)$, Alice and Bob are given an arbitrary number of i.i.d. samples and wish to generate *one sample* from the distribution $Q(U, V)$ which is given by $U = V = \mathrm{Ber}(1/2)$. (This is possible if and only if the Gács-Körner rate is positive.)

Motivation. Our motivation for studying the best discrete correlation that can be produced is twofold. On the one hand, this question forms part of the landscape of questions arising from a quest to weaken the assumptions about randomness when it is employed in distributed computing. Computational tasks are often solved well if parties have access to a common source of randomness and there has been recent interest in cryptography [10], [11], [12], [13], [14], [15], quantum computing [16], [17], [18] and communication complexity [19], [20], [21] to study how the ability to solve these tasks gets affected by weakening the source of randomness. In this space of investigations, it is a very natural question to ask how well one source of randomness can be tranformed to a different one, and Non-Interactive Simulation studies exactly this question.

On the other hand, from the analysis point of view, the Non-Interactive Simulation problem forms part of "tensor power" questions that have been challenging to analyze computationally. Specifically, in such questions, the quest is to understand how some quantity behaves as a function of the dimensionality of the problem as the dimension tends to infinity. Notable examples of such problems include the *Shannon capacity of a graph* [22], [23] where the goal is to understand how the independence number of the power of a graph behaves as a function of the exponent. Some more closely related examples arise in the problems of local state transformation of quantum entanglement [24], [25] and the problem of computing the entangled value of a game (see for eg, [26] and also the open problem [27]). A more recent example is the problem of computing the amortized communication complexity of a communication problem. Braverman-Rao [28] showed that this equals the information complexity of the communication problem, however the task of approximating the information complexity was only recently shown to be computable [29]. In our case, the best non-interactive simulation to get one pair of correlated bits might require many copies of (x, y) drawn from P and the challenge is to determine how many copies get us close. Convergence results of this type are not obvious. Indeed, the task of approximating the Shannon capacity remains open to this day [30]. Our work is motivated in part by the quest to understand tools that can be used to analyze such questions where rate of convergence to the desired quantity is non-trivial to bound.

Estimating Binary Correlations: Previous Work and our Result. In his work generalizing the results of Gács and Körner, Witsenhausen [1] gave an efficient algorithm that achieves a *quadratic* approximation to the Non-Interactive Simulation problem when $Q(U, V)$ is the distribution where U and V are marginally uniform over ± 1 and U is an ρ-correlated copy of V, i.e. $\mathbb{E}[UV] = \rho$ (henceforth, we refer to this distribution as $\mathrm{DSBS}(\rho)$).[2] Indeed, Witsenhausen introduced the Gaussian correlation problem as an intermediate step to solving this problem and his rounding technique to convert the Gaussian random variables into Boolean ones is essentially the same as that of the Goemans-Williamson algorithm for approximating maximum cut sizes in graphs [31]. Already implicit from the work of Witsenhausen is that "maximum correlation" gives a way to upper bound the best achievable ρ when simulating $\mathrm{DSBS}(\rho)$. Recent works in the information theory community [32], [7], [33] enhance the collection of analytical tools that can be used to show stronger impossibility results. While these works produce stronger bounds, they do not necessarily converge to the optimal limit and indeed basic questions about simulation remain open. For instance, till our work, even the following question was open [34]: If P is the uniform disribution on $\{(0,0), (0,1), (1,0)\}$ and $Q = \mathrm{DSBS}(.49)$ (i.e. U, V are uniformly ± 1, with $\mathbb{E}[UV] = .49$), can P simulate Q arbitrarily well? Our work answers such questions in principle. (Specifically we do give a finite time procedure to approximate the best ρ to within arbitrary accuracy. However, we have not run this algorithm to determine the answer to this specific question.)

Below we state our main theorem informally (see Theorem II.5 for the formal statement).

Theorem I.1 (Informal)**.** *There is an* algorithm *that takes as inputs a source distribution P, a parameter $\rho > 0$ and an error parameter $\delta > 0$, runs in time bounded by some computable function of P, ρ and δ, and either outputs a non-interactive protocol that simulates* $\mathrm{DSBS}(\rho)$ *up to additive δ in total variation distance, or asserts that there is no protocol that gets $O(\delta)$-close to* $\mathrm{DSBS}(\rho)$ *in total variation distance.*

More generally, the proof techniques extend to deciding the non-interactive simulation problem for an arbitrary 2×2 target distribution. In particular, we also show the following (see Theorem II.3 for the formal statement).

Theorem I.2 (Informal)**.** *There is an* algorithm *that takes as inputs a source distribution P, a 2×2 target distribution Q and an error parameter $\delta > 0$, runs in time bounded by some computable function of P, Q*

[2] Henceforth, we assume that bits are in the set $\{\pm 1\}$. By a *quadratic* approximation, we mean an algorithm distinguishing between the cases (i) $\rho \geq 1 - \eta$ and (ii) $\rho < 1 - O(\sqrt{\eta})$ for any given parameter $\eta > 0$.

and δ, and either outputs a non-interactive protocol that simulates Q up to additive δ in total variation distance, or asserts that there is no protocol that gets $O(\delta)$-close to Q in total variation distance.

The crux of Theorems I.1 and I.2 is to prove *computable* bounds on the number of copies of (X, Y) that are needed in order to come δ-close to the target distribution. We now describe the challenges towards achieving such bounds, and the techniques we use.

A. Proof Overview

We start by describing some illustrative special cases of the problem. In the case where $P = \mathrm{DSBS}(\rho)$, maximal correlation based arguments imply that $\mathrm{DSBS}(\rho)$ is the 'best' DSBS distribution that can simulated [1]. Thus, in this case, dictators functions achieve the optimal strategy. Consider now the case where P is a pair of ρ-correlated zero-mean unit-variance Gaussians[3]. Then, Borell's isoperimetric inequality implies that the strategy where each of Alice and Bob outputs the sign of her/his Gaussian achieves the best possible DSBS [35].

Given the above two examples where a *single-copy* strategy is optimal, it is tempting to try to determine the best DSBS that can be simulated using a single copy of P and hope that it would be close to the optimal DSBS (i.e., to the one that can be simulated using an arbitrary number of copies of P). But this approach cannot work as is illustrated by the following example which shows that using many copies of P is in some cases actually *needed*. Consider the source joint distribution corresponding to the bipartite graph in Figure 1 with $\alpha > 0$ being a small parameter (we interpret the distribution as the one obtained by sampling a random edge in the graph). This graph is the union of two components: a low-correlation component which has probability $1 - \alpha$ and a perfect-correlation component which has probability α. If we use a small number of copies of μ, the corresponding samples will most likely fall in the low-correlation component, and hence the best DSBS that can be produced in such a way would have a small correlation. On the other hand, as the number of used copies becomes larger than $1/\alpha$, with high probability at least one of the corresponding samples will fall in the perfect-correlation component, and hence the resulting DSBS would have correlation very close to 1. As another example, consider the distribution that is uniform on triples $\{(0,0),(0,1),(1,0)\}$. It follows from [1] that it is possible to simulate $\mathrm{DSBS}(1/3)$ using many copies of this distribution. However, it can be shown that using only a single copy of this distribution (along with private randomness), Alice and Bob can at best simulate $\mathrm{DSBS}(1/4)$.

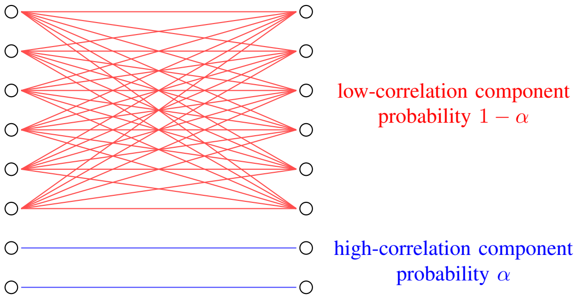

Fig. 1. Example source distribution for which many copies need to be considered.

We now describe at a high level, the main ideas that give us the computable bound on the number of samples of the joint distribution that are sufficient to obtain a δ-approximation to a given $\mathrm{DSBS}(\rho)$. First, we observe that the problem of deciding if one can come δ-close to simulating $\mathrm{DSBS}(\rho)$, is equivalent to checking if Alice and Bob can non-interactively come up with a distribution (X, Y) on $[-1,1] \times [-1,1]$ such that the marginals of X and Y have means close to 0, but $\mathbb{E}[XY]$ is large.

The results on correlation bounds for low-influence functions (obtained using the invariance principle) [36], [37], say that if Alice and Bob are using only low-influential functions, then in fact the correlation that they get cannot be much better than that obtained by taking appropriate threshold functions on correlated gaussians. Moreover, Alice and Bob can in fact simulate correlated gaussians using only a constant number of samples from the joint distribution, by applying the maximal correlation based technique of Witsenhausen [1].

In the general case, we show that we can first convert Alice and Bob's functions to have *low degree*, after which we apply a regularity lemma (inspired from that of [38]) to conclude that after fixing a constant number of coordinates, the restricted function is in fact low-influential. This reduces the general case to the special case of having low-influential functions and which is handled as described in the previous paragraph.

The more general case of simulating arbitrary 2×2 distribution also follows a similar outline. For a more technical overview of the proof, we refer the reader to Section III-A.

B. Roadmap of the paper

In Section II, we give some of the basic definitions, etc. Our main theorems are also presented in this section as Theorems II.3 and II.5. In Section III, we state our main technical lemma (Theorem III.1), which is used to prove Theorem II.5. We also give a proof overview for Theorem III.1. In Sections IV, V, VI and VII, we state and give some proof overview of the technical

[3] allowing here continuous distributions for the sake of intuition

lemmas involved in proving Theorem III.1. Finally, in Section VIII, we put together everything to prove Theorem III.1. We end with some open questions in Section IX. Additional preliminaries and detailed proofs are present in the full version [39].

II. Preliminaries

A. Notation

We use script letters $\mathcal{A}$, $\mathcal{B}$, etc. to denote finite sets, and μ will usually denote a probability distribution. $(\mathcal{A}\times\mathcal{B},\mu)$ is a joint probability space. We use μ_A and μ_B to denote the marginal distributions of μ. We use letters x, y, etc to denote elements of $\mathcal{A}$, and bold letters $\mathbf{x}$, $\mathbf{y}$, etc. to denote elements in $\mathcal{A}^n$. We use x_i, y_i to denote individual coordinates of $\mathbf{x}$, $\mathbf{y}$, respectively.

For a probability space $(\mathcal{A},\mu)$, we will use the following definitions and notations borrowed from [40]. $(\mathcal{A}^n,\mu^{\otimes n})$ denotes the product space $\mathcal{A}\times\mathcal{A}\times\cdots\times\mathcal{A}$ endowed with the product distribution. $\mathrm{Supp}(\mu) \overset{\text{def}}{=} \{x : \mu(x) > 0\}$ is the support of μ. We would generally assume without loss of generality that $\mathrm{Supp}(\mu) = \mathcal{A}$. $\alpha(\mu) \overset{\text{def}}{=} \min\{\mu(x) : x \in \mathrm{Supp}(\mu)\}$ denotes the minimum non-zero probability of any atom in $\mathcal{A}$ under the distribution μ. $L^2(\mathcal{A},\mu)$ denotes the space of functions from $\mathcal{A}$ to $\mathbb{R}$. The inner product on $L^2(\mathcal{A},\mu)$ is denoted by $\langle f,g\rangle_\mu := \mathbb{E}_{x\sim\mu}[f(x)g(x)]$. The ℓ_p-norm by $\|f\|_p := \left[\mathbb{E}_{x\sim\mu}|f(x)|^p\right]^{1/p}$. Also, $\|f\|_\infty := \max_{\mu(x)>0}|f(x)|$. It is easy to verify that $\|f\|_p \le \|f\|_q$ for $1 \le p \le q$. For two distributions μ and ν, $d_{\mathrm{TV}}(\mu,\nu)$ is the total variation distance between μ and ν.

B. The non-interactive simulation problem

The problem of non-interactive simulation is defined as follows,

Definition II.1 (Non-interactive simulation [7]). *Let $(\mathcal{A}\times\mathcal{B},\mu)$ and $(\mathcal{U}\times\mathcal{V},\nu)$ be two probability spaces. We say that the distribution ν can be* non-interactively simulated *using distribution μ, if there exists a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$ such that, $f_n : \mathcal{A}^n \to \mathcal{U}$, $g_n : \mathcal{B}^n \to \mathcal{V}$ and the distribution $\nu_n \sim (f_n(\mathbf{x}), g_n(\mathbf{y}))_{\mu^{\otimes n}}$ over $\mathcal{U}\times\mathcal{V}$ is such that $\lim_{n\to\infty} d_{\mathrm{TV}}(\nu_n,\nu) = 0$.*

The notion of non-interactive simulation is pictorially depicted in Figure 2. We formulate a natural gap-version of the non-interactive simulation problem defined as follows,

Problem II.2 (Gap-NIS$((\mathcal{A}\times\mathcal{B},\mu),(\mathcal{U}\times\mathcal{V},\nu),\delta)$). *Given probability spaces $(\mathcal{A}\times\mathcal{B},\mu)$ and $(\mathcal{U}\times\mathcal{V},\nu)$, and an error parameter $\delta > 0$, distinguish between the following cases:*

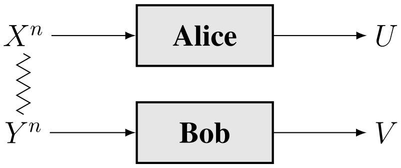

Fig. 2. Non-Interactive simulation as studied in [32], [7]

(i) there exists N, and functions $f : \mathcal{A}^N \to \mathcal{U}$ and $g : \mathcal{B}^N \to \mathcal{V}$, the distribution $\nu' = (f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$ is such that $d_{\mathrm{TV}}(\nu',\nu) \le \delta$.

(ii) for all N and all functions $f : \mathcal{A}^N \to \mathcal{U}$ and $g : \mathcal{B}^N \to \mathcal{V}$, the distribution $\nu' = (f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$ is such that $d_{\mathrm{TV}}(\nu',\nu) > 8\delta$. [4]

The main result in this paper is the following theorem showing that the problem of Gap-NIS is decidable when $|\mathcal{U}| = |\mathcal{V}| = 2$.

Theorem II.3 (Decidability of Gap-NIS for binary targets). *Given probability spaces $(\mathcal{A}\times\mathcal{B},\mu)$ and $(\mathcal{U}\times\mathcal{V},\nu)$ such that $|\mathcal{U}| = |\mathcal{V}| = 2$, and an error parameter δ, there exists an algorithm that runs in time $T((\mathcal{A}\times\mathcal{B},\mu),\delta)$ (which is an explicitly computable function), and decides the problem of* Gap-NIS$((\mathcal{A}\times\mathcal{B},\mu),(\mathcal{U}\times\mathcal{V},\nu),\delta)$. *The run time $T((\mathcal{A}\times\mathcal{B},\mu),\delta)$ is upper bounded by,*

$$\exp\exp\exp\left(\mathrm{poly}\left(\frac{1}{\delta},\ \frac{1}{1-\rho_0},\ \log\left(\frac{1}{\alpha}\right)\right)\right)$$

where $\rho_0 = \rho(\mathcal{A},\mathcal{B};\mu)$ is the maximal correlation of $(\mathcal{A}\times\mathcal{B},\mu)$ (defined in Section II-E) and $\alpha \overset{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ.

Doubly Symmetric Binary Source: In order to ease the presentation of ideas in proving the above theorem, we restrict to a special case, where the distribution $(\mathcal{U}\times\mathcal{V};\nu)$ is a *doubly symmetric binary source* defined below.

Definition II.4 (Doubly Symmetric Binary Source). *The distribution* $\mathrm{DSBS}(\rho)$ *is the joint distribution on ±1 random variables (U,V) given by the following table,*

	$V=+1$	$V=-1$
$U=+1$	$(1+\rho)/4$	$(1-\rho)/4$
$U=-1$	$(1-\rho)/4$	$(1+\rho)/4$

In particular, $\mathbb{E}[U] = \mathbb{E}[V] = 0$ and $\mathbb{E}[UV] = \rho$.

We will prove a special case of Theorem II.3, where the probability space $(\mathcal{U}\times\mathcal{V},\nu)$ is the distribution $\mathrm{DSBS}(\rho)$ for some ρ (see Theorem II.5 below). Even though we are proving only this special case, the main

[4] for sake of definition, the constant 8 could be replaced by any constant greater than 1. For a minor technical reason however our decidability results (Theorems II.3 and II.5) will require this constant to be strictly greater than 2. We choose to go ahead with 8 for convenience.

ideas involved here easily generalize to the proof of Theorem II.3 (proof in the full version [39]).

Theorem II.5 (Decidability of GAP-NIS for DSBS targets)**.** *Given a probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, and parameters ρ and δ, there exists an algorithm that runs in time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ (which is an explicitly computable function), and decides the problem of* $\text{GAP-NIS}((\mathcal{A} \times \mathcal{B}, \mu), \text{DSBS}(\rho), \delta)$.
The run time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ is upper bounded by,

$$\exp \exp \exp \left(\text{poly} \left(\frac{1}{\delta}, \frac{1}{1-\rho_0}, \log \left(\frac{1}{\alpha} \right) \right) \right)$$

where $\rho_0 = \rho(\mathcal{A}, \mathcal{B}; \mu)$ is the maximal correlation of $(\mathcal{A} \times \mathcal{B}, \mu)$ (defined in Section II-E) and $\alpha \overset{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ.

We will use $\text{GAP-NIS}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ as a shorthand for $\text{GAP-NIS}((\mathcal{A} \times \mathcal{B}, \mu), \text{DSBS}(\rho), \delta)$. Theorem II.5 will follow easily from the main technical lemma (Theorem III.1). The proof of Theorem II.5, assuming Theorem III.1 is present in the full version [39].

C. Reformulation of GAP-NIS

With the end goal of proving Theorem II.5, we introduce a new problem of Gap-Balanced-Maximum-Inner-Product, to which we show a reduction from GAP-NIS. This new formulation will be better suited for applying our techniques.

Problem II.6 $(\text{GAP-BAL-MAX-IP}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta))$**.** *Given a probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, and parameters ρ and δ, distinguish between the following cases:*

(i) there exists N, and functions $f : \mathcal{A}^N \to [-1, 1]$ and $g : \mathcal{B}^N \to [-1, 1]$, satisfying $|\mathbb{E}[f(\mathbf{x})]| \le \delta$ and $|\mathbb{E}[g(\mathbf{y})]| \le \delta$, such that the following holds,

$$\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \ge \rho - \delta$$

(ii) for all N and all functions $f : \mathcal{A}^N \to [-1, 1]$ and $g : \mathcal{B}^N \to [-1, 1]$, satisfying $|\mathbb{E}[f(\mathbf{x})]| \le 2\delta$ and $|\mathbb{E}[g(\mathbf{y})]| \le 2\delta$, the following holds,

$$\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] < \rho - 4\delta$$

The following proposition gives a reduction from the problem of GAP-NIS to the problem of GAP-BAL-MAX-IP (proof in full version [39]).

Proposition II.7. *For any probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and $\rho, \delta > 0$, the following reduction holds,*

1) *Case (i) of* $\text{GAP-NIS}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ *holds* $\implies$ *Case (i) of* $\text{GAP-BAL-MAX-IP}((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$ *holds*
2) *Case (ii) of* $\text{GAP-NIS}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ *holds* $\implies$ *Case (ii) of* $\text{GAP-BAL-MAX-IP}((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$ *holds*

D. Fourier analysis and Hypercontractivity

We will use standard notations in Fourier analysis for functions in $L^2(\mathcal{A}^n, \mu^{\otimes n})$, and use standard definitions such as Influence, Variance, etc. We will also use some concentration bounds based on hypercontractivity. Owing to space constraints, we present the requisite preliminaries in the full version [39].

E. Maximal Correlation and Witsenhausen's rounding

The "maximal correlation coeffcient" was first introduced by Hirschfeld [2] and Gebelein [3] and then studied by Rényi [4].

Definition II.8 (Maximal correlation)**.** *Given a joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, we define the maximal correlation of the joint distribution $\rho(\mathcal{A}, \mathcal{B}; \mu)$ as follows,*

$$\rho(\mathcal{A}, \mathcal{B}; \mu) \overset{\text{def}}{=} \sup_{\substack{f:\mathcal{A}\to\mathbb{R} \\ g:\mathcal{B}\to\mathbb{R}}} \mathop{\mathbb{E}}_{(x,y)\sim\mu} \frac{(f(x) - \mathbb{E}f)(g(y) - \mathbb{E}g)}{\sqrt{\text{Var}(f)\,\text{Var}(g)}}$$

Maximal correlation has certain properties which imply necessary conditions for when non-interactive simulation is possible (see full version [39] for more details). In addition, using a result of Witsenhausen [1], we have the following theorem,

Theorem II.9 (Witsenhausen [1])**.** *For any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, with $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$, then the largest ρ^* for which $(\mathcal{A} \times \mathcal{B}, \mu)$ can non-interactively simulate $\text{DSBS}(\rho^*)$ is bounded as follows,*

$$1 - \frac{2\arccos(\rho)}{\pi} \quad \le \quad \rho^* \quad \le \quad \rho$$

Note that, maximal correlation is an easily computable quantity, namely, it is the second largest singular value of the Markov operator[5] corresponding to $(\mathcal{A} \times \mathcal{B}, \mu)$.

Remark II.10. *The astute reader might have noticed a strong resemblance between Theorem II.9 and the random hyperplane rounding of Goemans-Williamson [31] used in the approximation algorithm for MAX-CUT. This is not a coincidence and indeed the bounds in Theorem II.9 come from morally the same technique as in [31].*

III. MAIN TECHNICAL LEMMA AND OVERVIEW

In this section we state the main technical lemma which will be used to solve GAP-BAL-MAX-IP. We also give a high level overview of the proof techniques.

Theorem III.1. *Given any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and any $\delta > 0$, there exists $n_0 = n_0((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ such that for any n and any functions $f : \mathcal{A}^n \to [-1, 1]$ and $g : \mathcal{B}^n \to [-1, 1]$, there exist functions $\widetilde{f} : \mathcal{A}^{n_0} \to$*

[5] The Markov operator corresponding to $(\mathcal{A} \times \mathcal{B}, \mu)$ is a $|\mathcal{A}| \times |\mathcal{B}|$ matrix T which is given by $T(x, y) = \mu(y|X = x)$.

$[-1,1]$ and $\widetilde{g} : \mathcal{B}^{n_0} \to [-1,1]$ such that $\left|\mathbb{E}[\widetilde{f}] - \mathbb{E}[f]\right| \le \delta/3$, $\left|\mathbb{E}[\widetilde{g}] - \mathbb{E}[g]\right| \le \delta/3$ and

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n_0}} \left[\widetilde{f}(\mathbf{x}) \cdot \widetilde{g}(\mathbf{y})\right] \ge \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})] - \delta$$

Most importantly, n_0 is a computable function in the parameters of the problem. In particular, one may take,

$$n_0 = \exp\left(\operatorname{poly}\left(\frac{1}{\delta}, \frac{1}{1-\rho}, \log\left(\frac{1}{\alpha}\right)\right)\right)$$

where $\rho \overset{\text{def}}{=} \rho(\mathcal{A}, \mathcal{B}; \mu)$ is the maximal correlation of $(\mathcal{A} \times \mathcal{B}, \mu)$ and $\alpha \overset{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ.

A. *Proof overview*

The proof of Theorem III.1 goes through a series of intermediate steps, which we describe at a high level here. At each step we lose only a small amount in the correlation. The first three steps preserve the marginals $\mathbb{E}[f]$ and $\mathbb{E}[g]$ exactly, while the fourth step incurs a small additive error in the same. The full proof is presented in Section VIII.

(I) *Smoothing of strategies.* We transform f and g into functions f_1, g_1 such that f_1 and g_1 have 'most' of their Fourier mass concentrated on terms of degree at most d, where d is a constant that depends on the distribution $(\mathcal{A} \times \mathcal{B}, \mu)$ and a tolerance parameter, but is independent of n. This transformation is described in Section IV.

(II) *Regularity lemma for low degree functions.* We first prove a *regularity lemma* (similar to the one in [38]) which roughly shows that for any degree-d polynomial, there exists a h-sized subset of variables, such that under a random restriction of the variables in this subset, the resulting function on the remaining variables has low individual influences (i.e. $\le \tau$). Note that h will be a constant depending on the degree d and τ, but will be independent of n.

We apply this regularity lemma on the degree-d truncated versions of both f_1 and g_1 obtained from Step (I). We take the union of the subsets obtained for f_1 and g_1. We show that with high probability over random restrictions of the variables in this subset, the resulting restriction of f_1 and g_1 on the remaining variables has low individual influences. This step is described in Section V.

Note that this step does not change the functions f_1 and g_1 at all, but we gain some structural knowledge about the same.

(III) *Correlation bounds for low influence functions.* We use results about correlation bounds for low influential functions [36], [37]. Intuitively, these results suggest that if the functions f_1 and g_1 were low influential functions to begin with, then the correlation $\mathbb{E}[f_1(\mathbf{x})g_1(\mathbf{y})]$ will not be 'much' better than the correlation between certain threshold functions applied on correlated gaussians.

We apply the above correlation bounds for the low influential functions obtained by restrictions of the small subset of variables in f_1 and g_1, to obtain functions $f_2 : \mathcal{A}^h \times \mathbb{R} \to [-1,1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \to [-1,1]$, where Alice and Bob together have access to h samples from $(\mathcal{A} \times \mathcal{B}, \mu)$ and a single copy of ρ-correlated gaussians, that is, $\mathcal{G}(\rho)$[6]. Here the correlation ρ is same as the maximal correlation $\rho(\mathcal{A}, \mathcal{B}; \mu)$. This step is described in Section VI.

(IV) *Simulating correlated gaussians.* Finally, Alice and Bob can non-interactively simulate the distribution $\mathcal{G}(\rho)$ using constantly many samples from $(\mathcal{A} \times \mathcal{B}, \mu)$. This is done using the technique of Witsenhausen [1], which primarily uses a 2-dimensional central limit theorem. This step is described in Section VII.

IV. Smoothing of Strategies

The first step in our approach is to obtain smoothed versions of the functions $f : \mathcal{A}^n \to [-1,1]$ and $g : \mathcal{B}^n \to [-1,1]$, which have *small Fourier tails*, without hurting the correlation by much. In particular, we show the following lemma (proof in full version [39]).

Lemma IV.1 (Smoothing of strategies). *Given any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and parameters $\lambda, \eta > 0$, there exists $d = d((\mathcal{A} \times \mathcal{B}, \mu), \lambda, \eta)$ such that for any n and any functions $f : \mathcal{A}^n \to [-1,1]$ and $g : \mathcal{B}^n \to [-1,1]$, there exist functions $f_1 : \mathcal{A}^n \to [-1,1]$ and $g_1 : \mathcal{B}^n \to [-1,1]$ such that $\mathbb{E}[f_1] = \mathbb{E}[f]$ and $\mathbb{E}[g_1] = \mathbb{E}[g]$, and*

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] - \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})]\right| \le \lambda$$

such that f_1 and g_1 have low energy Fourier tails, namely,

$$\sum_{|\boldsymbol{\sigma}|>d} \widehat{f_1}(\boldsymbol{\sigma})^2 \le \eta \quad \text{and} \quad \sum_{|\boldsymbol{\sigma}|>d} \widehat{g_1}(\boldsymbol{\sigma})^2 \le \eta$$

In particular, one may take $d = \frac{\log \eta}{2 \log \gamma}$, where $\gamma = 1 - C\frac{(1-\rho)\lambda}{\log(1/\lambda)}$, and $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$.

V. Joint Regularity Lemma for Fourier Concentrated Functions

The second step in our approach is to apply a *regularity lemma* on the functions $f_1 : \mathcal{A}^n \to [-1,1]$ and $g_1 : \mathcal{B}^n \to [-1,1]$ obtained from the previous step of smoothing. *Regularity lemma* is a loosely referred

[6] $\mathcal{G}(\rho)$ denotes a 2-dimensional gaussian distribution with mean $\begin{bmatrix}0\\0\end{bmatrix}$ and covariance matrix $\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$.

term which shows that for various types of combinatorial objects, an arbitrary object can be approximately decomposed into a constant number of “pseudorandom” sub-objects.

Our version of the regularity lemma draws inspiration from that of [38]; in fact our proofs also closely follow theirs. Formally, we show the following lemma (proof in full version [39]).

Lemma V.1 (Joint regularity lemma for Fourier-concentrated functions)**.** *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space. Let $d \in \mathbb{N}$ and $\tau > 0$ be any given constant parameters. There exists an $\eta \overset{\text{def}}{=} \eta(\tau) > 0$ and $h \overset{\text{def}}{=} h((\mathcal{A} \times \mathcal{B}, \mu), d, \tau)$ such that the following holds:*

For all $P \in L^2(\mathcal{A}^n, \mu_A^{\otimes n})$ and $Q \in L^2(\mathcal{B}^n, \mu_B^{\otimes n})$ satisfying $\sum_{|\boldsymbol{\sigma}|>d} \widehat{P}(\boldsymbol{\sigma})^2 \leq \eta$, $\sum_{|\boldsymbol{\sigma}|>d} \widehat{Q}(\boldsymbol{\sigma})^2 \leq \eta$, and $\mathrm{Var}[P] \leq 1$ and $\mathrm{Var}[Q] \leq 1$: there exists a subset of indices $H \subseteq [n]$ with $|H| \leq h$, such that the restrictions of the functions P and Q obtained by evaluating the coordinates in H according to distribution μ, satisfy the following (where we denote $T = [n] \setminus H$),

- (i) *With probability at least $1 - \tau$ over $\xi \sim \mu_A^{\otimes h}$, the restriction $P_\xi(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\mathrm{Inf}_i(P_\xi(\mathbf{x}_T)) \leq \tau$*
- (ii) *With probability at least $1 - \tau$ over $\xi \sim \mu_B^{\otimes h}$, the restriction $Q_\xi(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\mathrm{Inf}_i(Q_\xi(\mathbf{x}_T)) \leq \tau$*

In particular, one may take $\eta = \tau^2/16$ and $h = \frac{d}{\tau^2} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha)}{\alpha \cdot d \cdot \tau}\right)^{O(d)}$ which is a constant that depends on d, τ and $\alpha \overset{\text{def}}{=} \alpha(\mu)$, which is the minimum non-zero probability in μ. See the full version [39] for the definition of $C_4(\alpha)$, which is the hypercontractivity parameter.

Our regularity lemma draws inspiration from the one in [38]. In fact, our proof of the above regularity lemma also closely follows the proof steps in [38]. However their regularity lemma was much more involved as they were dealing with low-degree polynomial threshold functions, whereas we are directly dealing with low-degree polynomials. In particular, a major difference in our regularity lemmas is that [38] obtain a (potentially) *adaptive* decision tree, whereas we obtain just a single subset H. Also, our notion of ‘regularity’ is much simpler in that we only need all influences to be small. Another aspect of our regularity lemma is that it is robust enough to also work for Fourier concentrated functions, as opposed to only low-degree functions (potentially, [38] could also be modified to have this feature, although it was not required for their application). Another minor difference is that our Fourier analysis is for functions in $L^2(\mathcal{A}^n, \mu_A^{\otimes n})$, as opposed to functions on the boolean hypercube. But this is not really a significant difference and the proof steps go through as it is, albeit with slightly different parameters which depend on the hypercontractivity parameters of the distribution $(\mathcal{A}, \mu_A)$.

VI. Applying correlation bounds for low-influence functions

The third step in our approach is to use *correlation bounds for low-influence functions* obtained from the invariance principle [36], [37], to convert the functions $f_1 : \mathcal{A}^n \to [-1, 1]$ and $g_1 : \mathcal{B}^n \to [-1, 1]$ into functions $f_2 : \mathcal{A}^h \times \mathbb{R} \to [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \to [-1, 1]$ using the following lemma (proof in full version [39]).

Lemma VI.1 (Applying correlation bounds for low-influence functions)**.** *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space. Let $\gamma > 0$ be any given constant parameter. There exists a $\tau \overset{\text{def}}{=} \tau((\mathcal{A} \times \mathcal{B}, \mu), \gamma) > 0$ such that the following holds:*

For all functions $f_1 : \mathcal{A}^n \to [-1, 1]$ and $g_1 : \mathcal{B}^n \to [-1, 1]$, and a subset $H \subseteq [n]$ with $|H| = h$, such that the restrictions of the functions f_1 and g_1 obtained by evaluating the coordinates in H according to distribution μ, satisfy (i) and (ii) as in Lemma V.1 (replacing P and Q by f_1 and g_1 respectively).
There exist functions $f_2 : \mathcal{A}^h \times \mathbb{R} \to [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \to [-1, 1]$, such that,

$$\underset{\substack{\mathbf{x} \sim \mu_A^{\otimes h} \\ r_A \sim \mathcal{N}(0,1)}}{\mathbb{E}} f_2(\mathbf{x}, r_A) = \underset{\mathbf{x} \sim \mu_A^{\otimes n}}{\mathbb{E}} f_1(\mathbf{x})$$

$$\underset{\substack{\mathbf{y} \sim \mu_B^{\otimes h} \\ r_B \sim \mathcal{N}(0,1)}}{\mathbb{E}} g_2(\mathbf{y}, r_B) = \underset{\mathbf{y} \sim \mu_B^{\otimes n}}{\mathbb{E}} g_1(\mathbf{y})$$

and,

$$\begin{aligned} &\underset{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}}{\mathbb{E}} [f_2(\mathbf{x}, r_A) \cdot g_2(\mathbf{y}, r_B)] \\ \geq &\underset{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}}{\mathbb{E}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] - \gamma \end{aligned}$$

Additionally, f_2 and g_2 will have the following special form: there exist functions $f_2' : \mathcal{A}^h \to \mathbb{R}$ and $g_2' : \mathcal{B}^h \to \mathbb{R}$ such that,

$$f_2(\mathbf{x}, r) = \begin{cases} 1 & r \geq f_2'(\mathbf{x}) \\ -1 & r < f_2'(\mathbf{x}) \end{cases}$$

$$g_2(\mathbf{y}, r) = \begin{cases} 1 & r \geq g_2'(\mathbf{y}) \\ -1 & r < g_2'(\mathbf{y}) \end{cases}$$

Also, one may take $\tau = \gamma^{O\left(\frac{\log(1/\gamma)\log(1/\alpha)}{(1-\rho)\gamma}\right)}$, where $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$ and $\alpha \overset{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ.

The main technical tool in proving Lemma VI.1 is a result about correlation bounds for low influence functions (which are generalizations of the ‘Majority is Stablest’ theorem), which is obtained from the invariance principle [36], [37].

VII. Simulating Correlated Gaussians

In this section, we use the technique due to Witsenhausen [1] which shows that for any joint probability space $(\mathcal{A}\times\mathcal{B},\mu)$ with maximal correlation ρ, Alice and Bob can non-interactively simulate ρ-correlated gaussians upto arbitrarily small *2-dimensional Kolmogorov distance*. We obtain the following lemma (proof in full version [39]).

Lemma VII.1 (Witsenhausen's rounding)**.** *Let $(\mathcal{A}\times\mathcal{B},\mu)$ be a joint probability space, and let $\rho=\rho(\mathcal{A},\mathcal{B};\mu)$ be its maximal correlation. Let $\zeta>0$ be any given parameter. Then, there exists $w\stackrel{\text{def}}{=}w((\mathcal{A}\times\mathcal{B},\mu),\zeta)\in\mathbb{N}$, such that the following holds:*

For all functions $f_2:\mathcal{A}^h\times\mathbb{R}\to[-1,1]$ and $g_2:\mathcal{B}^h\times\mathbb{R}\to[-1,1]$ having the special form as in Lemma VI.1, there exist functions $f_3:\mathcal{A}^{h+w}\to[-1,1]$ and $g_3:\mathcal{B}^{h+w}\to[-1,1]$, such that,

$$\left|\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_A^{\otimes(h+w)}} f_3(\mathbf{x})-\mathop{\mathbb{E}}_{\substack{\mathbf{x}\sim\mu_A^{\otimes h}\\ r_A\sim\mathcal{N}(0,1)}}[f_2(\mathbf{x},r_A)]\right|\le\zeta$$

$$\left|\mathop{\mathbb{E}}_{\mathbf{y}\sim\mu_B^{\otimes(h+w)}} g_3(\mathbf{y})-\mathop{\mathbb{E}}_{\substack{\mathbf{x}\sim\mu_B^{\otimes h}\\ r_B\sim\mathcal{N}(0,1)}}[g_2(\mathbf{y},r_B)]\right|\le\zeta$$

and,

$$\left|\begin{array}{l}\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes(h+w)}}[f_3(\mathbf{x})\cdot g_3(\mathbf{y})]\\ -\mathop{\mathbb{E}}_{\substack{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes h}\\ (r_A,r_B)\sim\mathcal{G}(\rho)}}[f_2(\mathbf{x},r_A)\cdot g_2(\mathbf{y},r_B)]\end{array}\right|\quad\le\quad\zeta$$

In particular, one may take $w=O\left(\frac{1+\rho}{\alpha\cdot(1-\rho)^3\cdot\zeta^2}\right)$, where $\alpha\stackrel{\text{def}}{=}\alpha(\mu)$ is the minimum non-zero probability in μ.

VIII. Putting it all together!

In this section we finally use all the lemmas we have developed to prove Theorem III.1.

Proof of Theorem III.1. Given $(\mathcal{A}\times\mathcal{B},\mu)$ and $\delta>0$ and functions $f:\mathcal{A}^n\to[-1,1]$ and $g:\mathcal{B}^n\to[-1,1]$, we wish to apply Lemma VI.1 with parameter $\gamma=\delta/3$ followed by Lemma VII.1 with parameter $\zeta=\delta/3$. Lemma VI.1 will dictate a value $\tau=\tau((\mathcal{A}\times\mathcal{B},\mu),\gamma)$. We wish to apply the Joint regularity lemma (Lemma V.1), with this parameter τ, which will dictate a value of $\eta=\eta(\tau)$. Using this value of η, and $\lambda=\delta/3$, we apply the Smoothing lemma (Lemma IV.1), which will dictate a value of $d=d((\mathcal{A}\times\mathcal{B},\mu),\lambda,\eta)$. We use this d to feed into the joint regularity lemma (Lemma V.1), to obtain a value of h. The final value of n_0 is the sum of $h((\mathcal{A}\times\mathcal{B},\mu),d,\tau)$ given by the joint regularity lemma (Lemma V.1) and $w((\mathcal{A}\times\mathcal{B},\mu),\zeta)$ given by Witsenhausen's rounding procedure (Lemma VII.1). This dependency of parameters is pictorially described in Figure 3 (the dependencies on $(\mathcal{A}\times\mathcal{B},\mu)$ are suppressed, for sake of clarity). It can be shown by putting everything together that $n_0=\exp\left(\text{poly}\left(\frac{1}{\delta},\ \frac{1}{1-\rho},\ \log\left(\frac{1}{\alpha}\right)\right)\right)$.

Once we have all the parameters set, we are now able to apply them to any pair of functions $f:\mathcal{A}^n\to[-1,1]$ and $g:\mathcal{B}^n\to[-1,1]$. In particular, we proceed as described in the overview (Section III).

- **(I)** We apply Lemma IV.1 to functions f and g with parameters λ and η as obtained above. This gives us a degree d and functions f_1 and g_1, such that, $\sum_{|\boldsymbol{\sigma}|>d}\widehat{f}(\boldsymbol{\sigma})^2<\eta$ and $\sum_{|\boldsymbol{\sigma}|>d}\widehat{g}(\boldsymbol{\sigma})^2<\eta$.
- **(II)** We apply the joint regularity lemma (Lemma V.1) on functions f_1 and g_1, with parameters d and τ as obtained above (note that, the conditions involving η are satisfied, because we chose precisely this η to be given to the Smoothing lemma). This gives us a subset $H\subseteq[n]$ such that $|H|\le h$ and with high probability over restrictions to this subset H, the restricted versions of both f_1 and g_1 have all individual influences to be at most τ.
- **(III)** We apply the correlation bounds result (Lemma VI.1) to functions f_1 and g_1 (note that all the conditions involving τ are satisfied already because we chose precisely this τ to be given to the joint regularity lemma).
  This gives us functions $f_2:\mathcal{A}^h\times\mathbb{R}\to[-1,1]$ and $g_2:\mathcal{B}^h\times\mathbb{R}\to[-1,1]$ of the special form as in Lemma VI.1.
- **(IV)** Functions f_2 and g_2 are exactly in the form for which Lemma VII.1 is applicable, which we use with parameters ζ as obtained above. This gives us functions $f_3:\mathcal{A}^{h+w}\to[-1,1]$ and $g_3:\mathcal{B}^{h+w}\to[-1,1]$.

Note that, $\mathbb{E}f=\mathbb{E}f_1=\mathbb{E}f_2$ and $\left|\mathbb{E}f_3-\mathbb{E}f_2\right|\le\zeta=\delta/3$ and similarly $\mathbb{E}g=\mathbb{E}g_1=\mathbb{E}g_2$ and $\left|\mathbb{E}g_3-\mathbb{E}g_2\right|\le\zeta=\delta/3$. Moreover, we have from Lemmas VII.1, VI.1 and IV.1 that,

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes(h+w)}}[f_3(\mathbf{x})\cdot g_3(\mathbf{y})]\\ \ge\ &\mathop{\mathbb{E}}_{\substack{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes h}\\ (r_A,r_B)\sim\mathcal{G}(\rho)}}[f_2(\mathbf{x})\cdot g_2(\mathbf{y})]-\zeta\\ \ge\ &\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}[f_1(\mathbf{x})\cdot g_1(\mathbf{y})]-\gamma-\zeta\\ \ge\ &\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}[f(\mathbf{x})\cdot g(\mathbf{y})]-\lambda-\gamma-\zeta\\ =\ &\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}[f(\mathbf{x})\cdot g(\mathbf{y})]-\delta\end{aligned}$$

Hence, taking $\widetilde{f}=f_3$ and $\widetilde{g}=g_3$, proves Theorem III.1. □

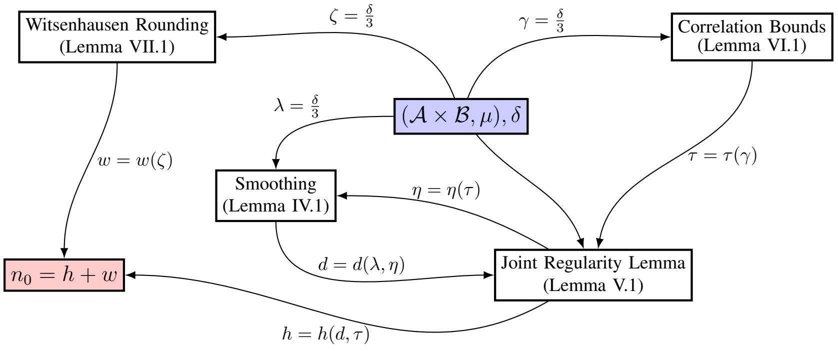

Fig. 3. Dependency of parameters in the proof of Theorem III.1

IX. Open Questions

In this work, we proved computable bounds on the non-interactive simulation of any 2×2 distribution. We now conclude with some interesting open questions.

The running time of our algorithm is at least doubly-exponential in the input size[7]. It would be very interesting to understand the computational complexity of the non-interactive simulation problem. We point out that the question of generating the best DSBS can be thought of as a tensored version of the following "Min-Bipartite-Bisection" problem: We are given a weighted bipartite graph $G = (L \cup R, E)$, and we wish to find a subset S of $L \cup R$ such that $S \cap L$ roughly contains half the vertices of L, and $S \cap R$ roughly contains half the vertices of R, while minimizing the total weight of edges crossing the cut $(S, \overline{S})$. While it follows from [41] that Min-Bipartite-Bisection is hard to approximate, the same is not necessarily true about its tensored version.

Another interesting open question is to generalize our decidability results to larger alphabets, which seems to require new technical ideas. Indeed, our proof of Theorems I.1 and I.2 relied on the fact that for (X, Y) being correlated random Gaussians, the maximum possible agreement of any pair of ± 1-valued functions $f(X)$ and $g(Y)$ is at most that of two appropriate dictator threshold functions $F(X_1)$ and $G(Y_1)$ where F only depends on the marginals of f (i.e., the probability that f takes the values -1 and $+1$), and similarly G only depends on the marginals of g. The analogous statement for the ternary case is not true. Namely, let $f(X), g(Y) \in \{0, 1, 2\}$, and assume that the marginals of f are $(1/3, 1/3, 1/3)$. Then, depending on whether the marginals of g are $(1/3, 1/3, 1/3)$ or $(1/2, 1/2, 0)$, the largest agreement of (f, g) would be achieved by very different functions f, assuming the "Standard Simplex Conjecture" (see [42] and Proposition 2.10 of [43]). This example shows that in the ternary case Alice cannot replace f by a function of a very small number of copies without taking the marginals of Bob's function g into account, and this is a major obstacle in generalizing our approach for proving Theorems I.1 and I.2 to larger alphabets.

Yet another interesting open question is to generalize our computability results to more than two players, which also seems to require new technical ideas.

Finally, it will be very interesting to see if these techniques could apply to other 'tensored' problems. The most relevant problems seem to be (i) deciding a quantum version of our problem, namely that of local state transformation of quantum entanglement [24], [25] and (ii) approximately computing the entangled value of a 2-prover 1-round game ([26]; also see the open problem [27]).

X. Acknowledgments

We thank Sudeep Kamath for explaining to us the state-of-the-art results in the information theory community, with regards to the problem of non-interactive simulation. We also thank Boaz Barak, Mohammad Bavarian and Mohsen Ghaffari for helpful discussions. We thank Matthew Coudron and Robin Kothari for pointing us to the related problems in the quantum literature.

References

[1] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[2] H. O. Hirschfeld, "A connection between correlation and contingency," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, no. 04. Cambridge Univ Press, 1935, pp. 520–524.

[3] H. Gebelein, "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*, vol. 21, no. 6, pp. 364–379, 1941.

[7] For constant values of δ and ρ, the running time is doubly-exponential in $2^{\mathrm{poly}(\log m)}$. Here we think of the input as a bipartite graph with m edges. This follows because $\alpha \sim 1/m$.

[4] A. Rényi, "On measures of dependence," *Acta mathematica hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.

[5] 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.

[6] A. D. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975. [Online]. Available: http://dx.doi.org/10.1109/TIT.1975.1055346

[7] S. Kamath and V. Anantharam, "On non-interactive simulation of joint distributions," *arXiv preprint arXiv:1505.00769*, 2015.

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

[9] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality," *Israel Journal of Mathematics*, vol. 154, no. 1, pp. 299–336, 2006.

[10] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. part i: secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, 1993.

[11] ——, "Common randomness in information theory and cryptography. ii. cr capacity," *Information Theory, IEEE Transactions on*, vol. 44, no. 1, pp. 225–240, 1998.

[12] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion," in *advances in Cryptology—EUROCRYPT'93*. Springer, 1994, pp. 410–423.

[13] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 344–366, 2000.

[14] U. M. Maurer, "Secret key agreement by public discussion from common information," *Information Theory, IEEE Transactions on*, vol. 39, no. 3, pp. 733–742, 1993.

[15] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in cryptology-ASIACRYPT 2005*. Springer, 2005, pp. 199–216.

[16] M. A. Nielsen, "Conditions for a class of entanglement transformations," *Physical Review Letters*, vol. 83, no. 2, p. 436, 1999.

[17] E. Chitambar, R. Duan, and Y. Shi, "Tripartite entanglement transformations and tensor rank," *Physical review letters*, vol. 101, no. 14, p. 140502, 2008.

[18] P. Delgosha and S. Beigi, "Impossibility of local state transformation via hypercontractivity," *Communications in Mathematical Physics*, vol. 332, no. 1, pp. 449–476, 2014.

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

[20] C. Cannonne, V. Guruswami, R. Meka, and M. Sudan, "Communication with imperfectly shared randomness," *ITCS*, 2014.

[21] B. Ghazi, P. Kamath, and M. Sudan, "Communication complexity of permutation-invariant functions," in *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, 2016, pp. 1902–1921. [Online]. Available: http://dx.doi.org/10.1137/1.9781611974331.ch134

[22] C. E. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, 1956.

[23] L. Lovász, "On the shannon capacity of a graph," *Information Theory, IEEE Transactions on*, vol. 25, no. 1, pp. 1–7, 1979.

[24] S. Beigi, "A new quantum data processing inequality," *CoRR*, vol. abs/1210.1689, 2012. [Online]. Available: http://arxiv.org/abs/1210.1689

[25] P. Delgosha and S. Beigi, "Impossibility of local state transformation via hypercontractivity," *CoRR*, vol. abs/1307.2747, 2013. [Online]. Available: http://arxiv.org/abs/1307.2747

[26] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick, "Entangled games are hard to approximate," *SIAM J. Comput.*, vol. 40, no. 3, pp. 848–877, 2011. [Online]. Available: http://dx.doi.org/10.1137/090751293

[27] "OpenQIProblemsWiki - All the Bell Inequalities," http://qig.itp.uni-hannover.de/qiproblems/1, accessed: 2016-07-12.

[28] M. Braverman and A. Rao, "Information equals amortized communication," in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*. IEEE, 2011, pp. 748–757.

[29] M. Braverman and J. Schneider, "Information complexity is computable," *arXiv preprint arXiv:1502.02971*, 2015.

[30] N. Alon and E. Lubetzky, "The shannon capacity of a graph and the independence numbers of its powers," *Information Theory, IEEE Transactions on*, vol. 52, no. 5, pp. 2172–2176, 2006.

[31] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995. [Online]. Available: http://doi.acm.org/10.1145/227683.227684

[32] S. Kamath and V. Anantharam, "Non-interactive simulation of joint distributions: The hirschfeld-gebelein-rényi maximal correlation and the hypercontractivity ribbon," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1057–1064.

[33] S. Beigi and A. Gohari, "On the duality of additivity and tensorization," *arXiv preprint arXiv:1502.00827*, 2015.

[34] S. Kamath, "Personal communication," 2015.

[35] C. Borell, "Geometric bounds on the ornstein-uhlenbeck velocity process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 70, no. 1, pp. 1–13, 1985.

[36] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: invariance and optimality," in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*. IEEE, 2005, pp. 21–30.

[37] E. Mossel, "Gaussian bounds for noise correlation of functions," *Geometric and Functional Analysis*, vol. 19, no. 6, pp. 1713–1756, 2010.

[38] I. Diakonikolas, R. A. Servedio, L.-Y. Tan, and A. Wan, "A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions," in *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*. IEEE, 2010, pp. 211–222.

[39] B. Ghazi, P. Kamath, and M. Sudan, "Decidability of non-interactive simulation of joint distributions," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 23, p. 104, 2016. [Online]. Available: http://eccc.hpi-web.de/report/2016/104

[40] P. Austrin and J. Håstad, "Randomly supported independence and resistance," *SIAM Journal on Computing*, vol. 40, no. 1, pp. 1–27, 2011.

[41] P. Raghavendra, D. Steurer, and M. Tulsiani, "Reductions between expansion problems," in *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*. IEEE, 2012, pp. 64–73.

[42] M. Isaksson and E. Mossel, "Maximally stable gaussian partitions with discrete applications," *Israel Journal of Mathematics*, vol. 189, no. 1, pp. 347–396, 2012.

[43] S. Heilman, E. Mossel, and J. Neeman, "Standard simplices and pluralities are not the most noise stable," in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, 2015, p. 255. [Online]. Available: http://doi.acm.org/10.1145/2688073.2688076