comput. complex.

https://doi.org/10.1007/s00037-019-00190-7

computational complexity

SIMULATION THEOREMS VIA PSEUDO-RANDOM PROPERTIES

Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay

Abstract.
We generalize the deterministic simulation theorem of Raz & McKenzie (Combinatorica 19(3):403–435, 1999), to any gadget which satisfies a certain hitting property. We prove that inner product and gap-Hamming satisfy this property, and as a corollary, we obtain a deterministic simulation theorem for these gadgets, where the gadget's input size is logarithmic in the input size of the outer function. This yields the first deterministic simulation theorem with a logarithmic gadget size, answering an open question posed by Göös, Pitassi & Watson (in: Proceedings of the 56th FOCS, 2015).
Our result also implies the previous results for the indexing gadget, with better parameters than was previously known. Moreover, a simulation theorem with logarithmic-sized gadget implies a quadratic separation in the deterministic communication complexity and the logarithm of the 1-partition number, no matter how high the 1-partition number is with respect to the input size—something which is not achievable by previous results of Göös, Pitassi & Watson (2015).

Keywords. Communication complexity, lifting theorem, simulation theorem, Inner-product, gap-Hamming

Subject classification. Theory of computation — Communication complexity

1. Introduction

A very basic problem in computational complexity is to understand the *complexity* of a composed function $f \circ g$ in terms of the com-

Published online: 18 July 2019

plexities of the two functions f and g used for the composition. For concreteness, we consider $f : \{0,1\}^p \to \mathcal{Z}$ and $g : \{0,1\}^m \to \{0,1\}$ and denote the composed function as $f \circ g^p : \{0,1\}^{mp} \to \mathcal{Z}$; then, f is called the *outer function* and g is called the *inner function*, or *gadget*. The special case of $\mathcal{Z}$ being $\{0,1\}$ and f the XOR function has been the focus of several works (Impagliazzo 1995; Lee, Shraibman & Spalek 2008; Levin 1987; Shaltiel 2003; Sherstov 2012b; Viola & Wigderson 2008; Yao 1982), commonly known as XOR lemmas. Another special case is when f is the trivial function that maps each point to itself. This case has also been widely studied in various parts of complexity theory under the names of 'direct-sum' and 'direct-product' problems, depending on the quality of the desired solution (Barak, Braverman, Chen & Rao 2013; Beame, Pitassi, Segerlind & Wigderson 2005; Braverman & Rao 2014; Braverman, Rao, Weinstein & Yehudayoff 2013a,b; Brody, Buhrman, Koucký, Loff, Speelman & Vereshchagin 2013; Drucker 2012; Harsha, Jain, McAllester & Radhakrishnan 2007; Jain 2015; Jain, Klauck & Nayak 2008; Jain, Pereszlényi & Yao 2012; Jain, Radhakrishnan & Sen 2003; Jain & Yao 2012; Kerenidis, Laplante, Lerays, Roland & Xiao 2015; Pankratov 2012). Making progress on even these special cases of the general problem in various models of computation is an outstanding open problem.

In the last few years, there has been some progress toward understanding the complexity of $f \circ g^p$, in the setting of communication complexity. In this setting, each input for g is split between two parties, Alice and Bob. A particular instance of progress from a few years ago is the development of the pattern matrix method by Sherstov (2011) and the closely related block-composition method of Shi & Zhu (2009), which led to a series of interesting developments (Chattopadhyay 2007; Chattopadhyay & Ada 2008; Lee, Shraibman & Spalek 2008; Rao & Yehudayoff 2015; Sherstov 2012a, 2013), resolving several open problems along the way. In both these methods, the relevant analytic property of the outer function is the approximate degree. While the pattern-matrix method entailed the use of a special inner function, the block-composition method, further developed by Chattopadhyay (2009), Lee & Zhang (2010) and Sherstov (Sherstov 2012a, 2013), prescribed the inner function to

have small discrepancy. These methods are able to lower bound the randomized communication complexity of $f \circ g^p$ essentially by the product of the approximate degree of f and the logarithm of the inverse of the discrepancy of g.

From the upper-bound perspective, the following simple protocol is suggestive: Alice and Bob try to solve f using a deterministic decision-tree algorithm. Such an algorithm queries the input bits of f frugally. Whenever there is a query, Alice and Bob solve the relevant instance of g by using the best protocol for g. This allows them to progress with the decision-tree computation of f, yielding (informally) an upper bound of $\mathcal{D}^{cc}\big(f \circ g^p\big) \leq \mathcal{D}^{dt}\big(f\big) \cdot \mathcal{D}^{cc}\big(g\big)$, where $\mathcal{D}^{cc}$ and $\mathcal{D}^{dt}$ denote the deterministic communication complexity and deterministic decision-tree complexity, respectively[1]. A natural question is whether the above upper bound is essentially optimal. The case when both f and g are XOR clearly shows that this is not always the case. However, this may be just a pathological example. It is natural to ask: for which inner functions g, is the above naive algorithm optimal?

In a remarkable and celebrated work, Raz & McKenzie (1999) showed that this naive upper bound is always optimal, when g is a *large* indexing function (IND), i.e., the *gadget size, m, is polynomially large* in p. This theorem was the main technical tool used by Raz-McKenzie to famously separate the monotone NC hierarchy. The work of Raz-McKenzie was recently simplified and built upon by Göös, Pitassi & Watson (2015) to solve a long-standing open problem in communication complexity. In line with Göös, Pitassi & Watson (2015), we call such theorems *simulation theorems*, because they explicitly construct a decision tree for f by simulating a given protocol for $f \circ g^p$.

Simulation theorems have numerous applications. To give an example closely related to (Göös, Pitassi & Watson 2015; Raz & McKenzie 1999): Bonet, Esteban, Galesi & Johannsen (2000), and more recently de Rezende, Nordström & Vinyals (2016) port

[1] An analogous result holds in the randomized model, where the upper bound holds with a multiplicative factor of $\log \mathcal{R}^{dt}(f)$— this is because we need to amplify the success probability of solving each instance of g so that we can do an union bound for the overall success probability of solving all instances of g.

the above deterministic simulation theorem to the model of real communication, yielding new trade-offs for the measures of size and space in the cutting planes proof system. Other applications of composition theorems include monotone-circuit lower bounds (Göös & Pitassi 2014; Johannsen 2001; Karchmer & Wigderson 1990; Raz & McKenzie 1999; Robere, Pitassi, Rossman & Cook 2016; Sokolov 2017), small-depth circuit lower bounds (Chattopadhyay 2007; Sherstov 2009), proof-complexity lower bounds (Beame, Huynh & Pitassi 2010; Huynh & Nordstrom 2012), and separations of complexity classes in communication complexity (David, Pitassi & Viola 2009; Göös, Lovett, Meka, Watson & Zuckerman 2015; Göös, Pitassi & Watson 2015).

Many of these developments have happened recently. Since our work has been publicly disseminated, we have seen new simulation theorems, analogous to the above, proven in various settings (Göös, Kamath, Pitassi & Watson 2017a; Göös, Pitassi & Watson 2017b; Watson 2017); indeed, in FOCS 2017, a workshop (Meka & Pitassi 2017) was devoted entirely to such results and their applications.

1.1. Our contributions. The main contributions of this work are the following:

- **Generalization of Raz-McKenzie.** We generalize the simulation theorem of Raz-McKenzie, by singling out a new property (**P**) of a function $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, that we call "having (δ, h)-hitting monochromatic rectangle distributions", and then showing that a simulation theorem will hold for any gadget g with this property.

  Our paper makes a conceptual contribution, by separating the proof of a deterministic simulation theorem into two distinct parts: a generic argument that guarantees simulation theorems whenever g has property (**P**), and a proof that a desired g has property (**P**). Thus, given our work, if one wished to prove a deterministic simulation theorem for a new gadget g', one will only need to show it has property (**P**) and the rest will seamlessly follow.

  The proof of the first part, the simulation theorem for gadgets g having property (**P**), has a similar structure to the proof

by Göös, Pitassi & Watson (2015) of the Raz & McKenzie (1999) simulation theorem. Some modifications are required to make the argument work for "symmetric" gadgets g.

- **Other gadgets.** Furthermore, we prove that property (**P**) holds for the gap-Hamming problem over n bits (GH_n), where the gap may be as large as $\frac{n}{4}$. For proving this, we make an interesting use of Harper's theorem. We also prove that property (**P**) holds for the inner-product mod 2 function over n bits (IP_n). To establish this, we use a probabilistic argument based on the second-moment method.

- **Improvement in gadget size.** The resulting simulation theorems for $f \circ \mathsf{IP}_n^p$ and $f \circ \mathsf{GH}_n^p$ only require the gadget input size n to be logarithmic in p, whereas the input size of the indexing gadget appearing in (Göös, Pitassi & Watson 2015; Raz & McKenzie 1999) is roughly p^{20}. Our results are the first examples of *deterministic* simulation theorems with such log-size gadgets, and the only example of a simulation theorem proven for a gadget having constant discrepancy (such as gap-Hamming with $\frac{n}{4}$ gap).

  Both of the above arguments require novel techniques, which are different than either the original Raz-McKenzie paper (Raz & McKenzie 1999) or its exposition by Göös, Pitassi & Watson (2015).

- **Application.** As an application of our simulation theorem (with a small gadget), we strengthen the separation result between deterministic communication complexity and logarithm of the 1-partition number (see Section 1.3) by Göös, Pitassi & Watson (2015). This results in a family of functions which exhibit a quadratic separation between these two quantities, no matter how high the 1-partition number is with respect to the input size. The result of Göös, Pitassi & Watson (2015) can show this separation only when the partition number is at most $N^{\frac{1}{42}}$ where N is the input size.

1.2. Statement of our results. Informally, a (δ, h)-hitting rectangle distribution (for $\delta \in (0,1)$ and $h \in \mathbb{N}$) is a distribution over

rectangles such that a random rectangle from this distribution will intersect any 2^{-h}-large rectangle with probability $\geq 1-\delta$. It is easy to come up with such a distribution: Consider a distribution where a rectangle of size $2^{n/2}$ is picked uniformly at random from the set of all rectangles of that size. It is not hard to see that such a random rectangle will intersect a large enough fixed rectangle with high probability, i.e., it is a $(o(1), n/2)$-hitting rectangle distribution. This is a considerably random distribution, i.e., the distribution has large entropy. We are interested in the following kind of *monochromatic* hitting distributions: by a function g having (δ, h)-hitting *monochromatic* rectangle distribution, we mean that there are two (δ, h)-hitting rectangle distributions σ_0 and σ_1, such that σ_c only samples rectangles which are c-monochromatic with respect to g. Note that the distributions σ_c may have much smaller entropy compared to a rectangle distribution μ which chooses a uniformly chosen rectangle of the same size. Even then, like μ, a rectangle sampled from σ_c is also required to intersect a large enough fixed rectangle with nonzero probability. Hence we may think of σ_c as being a *pseudo-random* rectangle distribution. Our generalization of Raz-McKenzie is the following:

THEOREM 1.1. *Let $f : \{0,1\}^p \to \mathcal{Z}$ be a (possibly partial) function over p-bit input, and $\mathcal{Z}$ is any domain. If g has (δ, h)-hitting monochromatic rectangle distributions, $\delta < 1/100$, and $p \leq 2^{\frac{h}{2}}$, then*

$$\mathcal{D}^{dt}(f) \leq \frac{8}{h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

We mention here, much like the Raz–McKenzie simulation theorem for the indexing gadget, Theorem 1.1 works even when f is a search problem, i.e., $f \subseteq \{0,1\}^n \times \mathcal{Z}$ and given query access to $x \in \{0,1\}^n$ we wish to find $z \in \mathcal{Z}$ such that $(x,z) \in f$. This kind of simulation theorem is sometimes harder to prove for search problems than it is for total functions. Contrast this with the following two results: (1) When g is a 2-bit XOR, Hatami, Hosseini & Lovett (2018) proved a simulation theorem of the form $\mathcal{D}^{cc}(f \circ g) \geq \mathcal{D}^{dt}_{\oplus}(f)^{1/6}$, where $\mathcal{D}^{dt}_{\oplus}(f)$ is the parity decision-tree complexity of f. This result, as is proven, requires f to be a total Boolean function. We still do not know whether such a result holds

when f is a search problem. (2) When g is the n-bit equality function, Loff & Mukhopadhyay (2019) have shown that a simulation theorem of the form $\mathcal{D}^{cc}(f \circ g) \geq \mathcal{D}^{dt}(f) \cdot n$ is provably not possible if we consider f to be a search problem. The best that can be proven in this case is $\mathcal{D}^{cc}(f \circ g) \geq \mathcal{D}^{dt}_{\text{AND}}(f) \cdot n$ where $\mathcal{D}^{dt}_{\text{AND}}(f)$ is the AND-decision-tree complexity of f. It is not hard to see that the equality gadget does not admit a hitting 1-monochromatic rectangle distribution, even though it does admit a hitting 0-monochromatic rectangle distribution. Surprisingly, if f is a total Boolean function, the following can be proven: $\mathcal{D}^{cc}(f \circ g) = \Omega(\mathcal{D}^{dt}_{\oplus}(f)^{1/3} \cdot n)$.

We show that two well-studied functions—the inner-product function (IP) and the gap-Hamming family of functions (GH)—have the above property. The inner-product function $\mathsf{IP}_n\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as $\mathsf{IP}_n(x,y) = \sum_{i \in [n]} x_i \cdot y_i$, where the summation is taken over field $\mathbb{F}_2$. Problems in the class of the gap-Hamming promise problems, parameterized with γ and denoted by $\mathsf{GH}_{n,\gamma} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, distinguish the case of (x,y) having Hamming distance at least $(\frac{1}{2} + \gamma)n$ from the case of (x,y) having Hamming distance at most $(\frac{1}{2} - \gamma)n$, for $0 \leq \gamma \leq 1/4$.

THEOREM 1.2. *The inner-product function and any function from the gap-Hamming class of promise functions over n bits admit $(o(1), \frac{n}{5})$ hitting monochromatic rectangle distributions.*

Combining Theorem 1.1 and Theorem 1.2 immediately yields the following simulation theorem.

THEOREM 1.3. *Let $p \leq 2^{\frac{n}{200}}$, $f : \{0,1\}^p \to \mathcal{Z}$ be a (possibly partial) function over p-bit input where $\mathcal{Z}$ is any domain, and $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the inner-product function, or any function from the gap-Hamming class of promise problems. Then,*

$$\mathcal{D}^{cc}\big(f \circ g^p\big) = \Theta\Big(\mathcal{D}^{dt}\big(f\big) \cdot n\Big).$$

The above theorem solves a problem raised by both Göös-Pitassi-Watson (Göös, Pitassi & Watson 2015) and Göös *et al.*

(Göös, Lovett, Meka, Watson & Zuckerman 2015) of proving a Raz-McKenzie style deterministic simulation theorem for a different inner function than indexing with a better gadget size. (Although the results presented in Göös, Lovett, Meka, Watson & Zuckerman (2015) do not deal with *deterministic* simulation theorems, the authors did raise the question of whether the proof of the deterministic simulation theorem can be simplified, and whether a simulation theorem can be shown for a larger class of gadgets g —we answer both these questions in this work.) Moreover, it is not hard to verify that any function $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ reduces to the indexing function $\mathsf{IND}_{2^n} : \{0,1\}^n \times \{0,1\}^{2^n} \to \{0,1\}$ (see Section 2), i.e., by exponentially blowing up the input size. This enables us to re-derive the original Raz-McKenzie simulation theorem for the indexing function, even attaining significantly better parameters. This improvement in parameters answers a question posed to us by Jakob Nordström (Nordström 2016). In the next section, we will show how this strong form of simulation theorem helps us prove a strong complexity separation result.

It is well known that the inner-product function has strong pseudo-random properties. In particular, it has vanishing discrepancy under the uniform distribution which makes it a good 2-source extractor. In fact, such strong properties of inner product were recently used to prove simulation theorems for more exotic models of communication by Göös *et al.* (Göös, Lovett, Meka, Watson & Zuckerman 2015) and also by the authors and Dvořák (Chattopadhyay, Dvorák, Koucký, Loff & Mukhopadhyay 2017a) to resolve a problem with a direct-sum flavor. By comparison, the pseudo-random property we abstract for proving our simulation theorem seems milder. This intuition is corroborated by the fact that we can show that the gap-Hamming problems also possess our property, even though we know that these problems have large $\Omega(1)$ discrepancy under all distributions. Interestingly, any technique that relies on the inner function having small discrepancy, such as the block-composition method, will not succeed in proving simulation theorems for such inner gadgets.

1.3. An application. If $F : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ is a two-player function, the 1-partition number of $F : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$, denoted

by $\chi_1(F)$, is the smallest number of rectangles needed to form a partition of $F^{-1}(1)$. It was known since Yannakakis (1991) that the deterministic communication complexity of F is $O(\log \chi_1(F))$, and Göös, Pitassi & Watson (2015) used a simulation theorem to show a matching separation. At this point, it is interesting to note the relation between input size and the 1-partition number of the functions for which they are able to show this separation. For an input of size $N = p^{21}$, Göös, Pitassi & Watson (2015) exhibit a function that has $\log(\chi_1) = \tilde{O}(\sqrt{p}) = \tilde{O}(N^{1/42})$, whereas the deterministic communication complexity is $\tilde{\Omega}(p) = \tilde{\Omega}(N^{1/21})$. This is shown by first constructing a function f witnessing an analogous separation for query complexity and then using a lifting theorem to establish the above separation for $F = f \circ g^p$. The input size N is large because Göös, Pitassi & Watson (2015) use a gadget g with a large input.

This raises the question whether such a separation is possible when χ_1 is closer to $\sqrt{N}$. The results of Göös, Pitassi & Watson (2015) do not rule out the possibility that for all F such that $\log \chi_1(F)$ is, say, $\omega(N^{\frac{1}{42}})$, the deterministic communication complexity of F is actually linear in $\log \chi_1(F)$. Our lifting theorem, with the improved gadget size, rules out this possibility—our simulation theorem can be used, in the same way as in (Göös, Pitassi & Watson 2015), to construct a function F^* for which $\log \chi_1(F^*)$ is $\tilde{\Theta}(\sqrt{N})$ and for which the deterministic communication complexity is $\tilde{\Omega}(N)$. We are thus able to obtain a quadratic separation in all regimes:

THEOREM 1.4. *For any function $s : \mathbb{Z} \to \mathbb{Z}$ such that $s(N) \leq \frac{\sqrt{N}}{\log N}$, there is a family of functions $\{F_N\}_{N \in \mathbb{Z}}$ such that $F_N : \{0,1\}^N \times \{0,1\}^N \to \mathcal{Z}$ has 1-partition number $\log \chi_1(F_N) = \tilde{O}(s(N))$ and deterministic communication complexity $\mathcal{D}^{cc}(F_N) \geq s(N)^2$.*

1.4. Our techniques. Our main tool for proving a tight deterministic simulation theorem is to use the general framework of the Raz-McKenzie theorem as expounded by Göös, Pitassi & Watson (2015). Here we provide a high-level sketch of our techniques.

Suppose we know a protocol for $f \circ g^p$. We are now given an input $z \in \{0,1\}^p$ for f and wish to compute $f(z)$ using a decision

tree. To do this, we will query the bits of z while simulating (in our head) the communication protocol for $f \circ g^p$ on inputs that are consistent with the queries to z we have made thus far. Namely, we maintain a rectangle $A \times B \subseteq \{0,1\}^{np} \times \{0,1\}^{np}$ so that for any $(x, y) \in A \times B$, $g^p(x, y)$ is *consistent* with z, meaning it $g^p(x, y)$ equals z on all the coordinates that were queried by the decision tree thus far. We will progress through the protocol with our rectangle $A \times B$ from the root to a leaf. As the protocol progresses, $A \times B$ shrinks according to the protocol, and our goal is to maintain the consistency requirement. For that, we need that inputs in $A \times B$ allow for all possible answers of g on those coordinates which we did not yet query. Hence, $A \times B$ needs to be rich enough, and we are choosing a path through the protocol that affects this richness the least. If the protocol forces us to shrink the rectangle $A \times B$ so that we may not be able to maintain the richness condition, we query another coordinate of z to restore the richness. Once we reach a leaf of the protocol we learn a correct answer for $f(z)$, because there is an input $(x, y) \in A \times B$ on which $g^p(x, y) = z$ (since we preserved consistency) and all inputs in $A \times B$ give the same answer for $f \circ g^p$,

The technical property of $A \times B$ that we will maintain is called *thickness*. $A \times B$ is thick on the i-th coordinate if for each input pair $(x, y) \in A \times B$, even after one gets to see all the coordinates of x and y except for x_i and y_i, the *uncertainty* of what appears in the ith coordinate remains large enough so that $g(x_i, y_i)$ can be arbitrary. For a given $x = (x_1, \ldots, x_p) \in \{0,1\}^{np}$, let us denote by $x_{\neq i}$ the tuple $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$ and by $\mathsf{Ext}^i_A(x_{\neq i})$ the set of possible extensions $x' \in \{0,1\}^n$ such that $(x_1, \ldots, x_{i-1}, x', x_{i+1}, \ldots, x_p) \in A$. We define $y_{\neq i}$ and $\mathsf{Ext}^i_B(y_{\neq i})$ similarly. If for a given x and y we know that both $\mathsf{Ext}^i_A(x_{\neq i})$ and $\mathsf{Ext}^i_B(y_{\neq i})$ are of size at least $2^{(\frac{1}{2}+\epsilon)n}$ then for $g = \mathsf{IP}_n$ there are extensions $x' \in \mathsf{Ext}^i_A(x_{\neq i})$ and $y' \in \mathsf{Ext}^i_B(y_{\neq i})$ such that $\mathsf{IP}_n(x_i, y_i) = z_i$. Hence, we say that $A \times B$ is τ-thick if $\mathsf{Ext}^i_A(x_{\neq i})$ and $\mathsf{Ext}^i_B(y_{\neq i})$ are of size at least $\tau \cdot 2^n$, for every choice of i and $x = (x_1, \ldots, x_p) \in A$, $y = (y_1, \ldots, y_p) \in B$.

So if we can maintain the thickness of $A \times B$ at a coordinate i which is not queried yet, then no matter which value z_i takes, there is some $(x, y) \in A \times B$ with $g(x_i, y_i) = z_i$. It turns out that

it is indeed possible to maintain thickness using the technique of Raz-McKenzie and Göös-Pitassi-Watson. Hence, as we progress through the protocol, we maintain a large rectangle $A \times B$ which is reasonably thick on the coordinates not queried so far. Once the size of either A or B drops below certain level, we are forced to make a query to another coordinate z_i and choose a sub-rectangle $A' \times B'$ of $A \times B$, so that $g(x_i, y_i)$ is fixed to z_i for all $(x, y) \in A' \times B'$. This can be done in such a way that the thickness of $A' \times B'$ on the unqueried coordinates is restored.

We give a sufficient condition for the inner function g that allows this type of argument to work, as follows. For $\delta \in (0, 1)$ and integer $h \geq 1$, we say that g has (δ, h)-*hitting monochromatic rectangle distributions* if there are two distributions σ_0 and σ_1 where for each $c \in \{0, 1\}$, σ_c is a distribution over c-monochromatic rectangles $R = U \times V \subset \{0, 1\}^n \times \{0, 1\}^n$ (i.e., $g(u, v) = c$ on every pair $(u, v) \in U \times V$), such that for any set $X \times Y \subset \{0, 1\}^n \times \{0, 1\}^n$ of sufficient size, a rectangle randomly chosen according to σ_c will intersect $X \times Y$ with large probability. More precisely, for any $c \in \{0, 1\}$ and for any $X \times Y$ with $|X|/2^n, |Y|/2^n \geq 2^{-h}$,

$$\Pr_{R \sim \sigma_c} [R \cap (X \times Y) \neq \varnothing] \geq 1 - \delta.$$

If such distributions σ_0 and σ_1 exist, we say that g has (δ, h)-hitting monochromatic rectangle distributions.

The distribution σ_0 for $\mathsf{GH}_{n,\frac{1}{4}}$ is sampled as follows: we first sample a random string x of Hamming weight $\frac{n}{2}$, and we look at the set of all strings which are at Hamming distance at most $\frac{n}{8}$ from x. Let's call this set U_x. The output of σ_0 will be the rectangle $U_x \times U_x$. The output of σ_1 is $U_x \times U_{\bar{x}}$, where $\bar{x}$ is the bitwise complement of x. For any such x, $U_x \times U_x$ will be a 0-monochromatic rectangle and $U_x \times U_{\bar{x}}$ will be a 1-monochromatic rectangle. Note that if U_x does not hit a subset A of $\{0, 1\}^n$, then it means that x is at least $\frac{n}{8}$ Hamming distance away from the set A. By an application of Harper's theorem, we can show that for a sufficiently large set A, the number of strings which are at least $\frac{n}{8}$ Hamming distance away from A is exponentially small. This will imply that both σ_0 and σ_1 will hit a sufficiently large rectangle with probability exponentially close to 1, which is our required hitting property.

The σ_0 distribution for IP_n is picked as follows: To produce a rectangle $U \times V$ we sample uniformly at random a linear subspace $V \subseteq F_2^n$ of dimension $n/2$ and we set $U = V^\perp$ to be the orthogonal complement of V. Since a random vector space of size $2^{n/2}$ hits a fixed subset of $\{0,1\}^n$ of size $2^{(\frac{1}{2}+\epsilon)n}$ with probability $1 - O(2^{-\epsilon n})$, and both U and V are random vector spaces of that size, $U \times V$ intersects a given rectangle $X \times Y$ with probability $1 - O(2^{-\epsilon n})$. Hence, we obtain $(O(2^{-\epsilon n}), (\frac{1}{2}+\epsilon)n)$-hitting distribution for IP. For the 1-monochromatic case, we first pick a random $a \in F_2^n$ of odd Hamming weight and then pick random V and $U = V^\perp$ inside of the orthogonal complement of a. The distribution σ_1 outputs the 1-monochromatic rectangle $(a + V) \times (a + U)$, and will have the required hitting property.

1.5. Organization. Section 2 consists of basic definitions and preliminaries. In Section 3, we prove a deterministic simulation theorem for any gadget admitting (δ, h)-hitting monochromatic rectangle distribution: Section 3.1 provides some supporting lemmas for the proof, and Section 3.2 holds the proof itself. In Section 4, we show that IND_n on n bits has $(\frac{1}{10}, \frac{3}{20}\log n)$-hitting rectangle distribution, in Section 5 we show that $\mathsf{GH}_{n,\frac{1}{4}}$ on n bits has $(o(1), \frac{n}{100})$-hitting rectangle distribution, and in Section 6 we show that IP on n bits has $(o(1), n/5)$-hitting rectangle distribution.

1.6. Further remarks. We remark here that Wu, Yao & Yuen (2017) have independently reported a proof of the simulation theorem for the inner-product function, while a draft of this manuscript was already in circulation. Implicit in their proof is the construction of hitting rectangle distributions for IP, and their construction of these distributions is similar to our own.

We would also like to point out to the readers that a preliminary version of the results obtained in this paper appeared in (Chattopadhyay *et al.* 2017b).

2. Basic definitions and preliminaries

A *combinatorial rectangle*, or just a *rectangle* for short, is any product $A \times B$, where both A and B are finite sets. If $A' \subseteq A$ and

$B' \subseteq B$, then $A' \times B'$ is called a *sub-rectangle* of $A \times B$. We will often be in a scenario where we wish to measure the size of a set A' which is contained in another set A; in this scenario, we will call *density* to the fraction $|A'|/|A|$. For two sets denoted using capital A, such as $A' \subseteq A$, we will use the Greek letter α to denote the density; for two sets denoted using capital B, such as $B' \subseteq B$, we will use β instead.

Consider a product set $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_p$, for some natural number $p \geq 1$, where each $\mathcal{A}_i$ is a subset of $\{0,1\}^n$. Let $A \subseteq \mathcal{A}$ and $I \subseteq [p] \stackrel{\text{def}}{=} \{1, \ldots, p\}$. Let $I = \{i_1 < i_2 < \cdots < i_k\}$, and $J = [p] \backslash I$. For any $a \in (\{0,1\}^n)^p$, we let $a_I = (\langle a_{i_1}, a_{i_2}, \ldots, a_{i_k})$ be the projection of a onto the coordinates in I. Correspondingly, $A_I = \{a_I \mid a \in A\}$ is the projection of the entire set A onto I. For any $a' \in (\{0,1\}^n)^k$ and $a'' \in (\{0,1\}^n)^{p-k}$, we denote by $a' \times_I a''$ the p-tuple a such that $a_I = a'$ and $a_J = a''$. If I is clear from the context, we may omit the set I and write only $a' \times a''$. For $i \in [p]$ and a p-tuple a, $a_{\neq i}$ denotes $a_{[p]\backslash\{i\}}$, and similarly, $A_{\neq i}$ denotes $A_{[p]\backslash\{i\}}$. For $a' \in (\{0,1\}^n)^k$, we define the set of extensions $\mathsf{Ext}_A^J(a') = \{a'' \in (\{0,1\}^n)^{p-k} \mid a' \times_I a'' \in A\}$; we call those a'' *extensions* of a'. Again, if A and I are clear from the context, we may omit them and write only $\mathsf{Ext}(a')$.

Suppose $n \geq 1$ is an integer and $\mathcal{A} = \{0,1\}^n$. For an integer p, a set $A \subseteq \mathcal{A}^p$, and a subset $S \subseteq \mathcal{A}$, the restriction of A to S at coordinate i is the set $A^{i,S} = \{a \in A \mid a_i \in S\}$. We write $A_I^{i,S}$ for the set $(A^{i,S})_I$ (i.e., we first restrict the i-th coordinate and then project onto the coordinates in I). Clearly $A_{\neq i}^{i,S}$ is non-empty if and only if S and A_i intersect.

The density of a set $A \subseteq \mathcal{A}^p$ will be denoted by $\alpha = \frac{|A|}{|\mathcal{A}|^p}$, and $\alpha_I^{i,S} = \frac{|A_I^{i,S}|}{|\mathcal{A}|^{|I|}}$.

Interval algebra. We will use the following notation to denote closed intervals of the real line:

- If δ is a nonnegative real, $1 \pm \delta$ denotes the interval $[1-\delta, 1+\delta]$.
- For two intervals $I = [a,b]$ and $J = [c,d]$, $IJ = \{xy \mid x \in I, y \in J\}$, $I + J = \{x + y \mid x \in I, y \in J\}$, and if $0 \notin J$, then $\frac{I}{J} = \{\frac{x}{y} \mid x \in I, y \in J\}$.

- For an interval $J = [a, b]$ and $x \in \mathbb{R}$, $xJ = \{xy \mid y \in J\}$, $x + J = \{x + y \mid y \in J\}$ and (if $0 \notin J$) $\frac{x}{J} = \{\frac{x}{y} \mid y \in J\}$.

The following is easy to verify:

PROPOSITION 2.1. *Let $0 \leq \delta < 1/2$ and x, y be reals.*

- *(Monotonicity) $1 \pm \delta \subseteq 1 \pm \delta'$ whenever $\delta \leq \delta'$.*
- *(Product rule) $(1 \pm \delta)^2 \subseteq 1 \pm 3 \cdot \delta$.*
- *(Weak inverse) $\frac{1}{1 \pm \delta} \subseteq 1 \pm 2\delta$.*
- *(Weak symmetry) If $x \in (1 \pm \delta) \cdot y$ then $y \in (1 \pm 2\delta) \cdot x$.*

Deterministic communication complexity. See Kushilevitz & Nisan (1997) for an excellent exposition on this topic, which we cover here only very briefly. In the two-party communication model introduced by Yao (1979), two computationally unbounded players, Alice and Bob, are required to jointly compute a function $F : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ where Alice is given $a \in \mathcal{A}$ and Bob is given $b \in \mathcal{B}$. To compute F, Alice and Bob communicate messages to each other, and they are charged for the total number of bits exchanged.

Formally, a *deterministic protocol* $\pi : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ is a binary tree where each internal node v is associated with one of the players; Alice's nodes are labeled by a function $a_v : \mathcal{A} \to \{0, 1\}$, and Bob's nodes by $b_v : \mathcal{B} \to \{0, 1\}$. Each leaf node is labeled by an element of $\mathcal{Z}$. For each internal node v, the two outgoing edges are labeled by 0 and 1, respectively. The *execution* of π on the input $(a, b) \in \mathcal{A} \times \mathcal{B}$ follows a path in this tree: starting from the root, in each internal node v belonging to Alice, she communicates $a_v(a)$, which advances the execution to the corresponding child of v; Bob does likewise on his nodes, and once the path reaches a leaf node, this node's label is the output of the execution. We say that π *correctly computes* F on (a, b) if this label equals $F(a, b)$.

To each node v of a deterministic protocol π, we associate a set $R_v \subseteq \mathcal{A} \times \mathcal{B}$ comprising those inputs (a, b) which cause π to reach node v. It is easy see that this set R_v is a combinatorial rectangle, i.e., $R_v = A_v \times B_v$ for some $A_v \subseteq \mathcal{A}$ and $B_v \subseteq \mathcal{B}$.

The *communication complexity of* π is the height of the tree. The *deterministic communication complexity of* F, denoted $\mathcal{D}^{cc}(F)$, is defined as the smallest communication complexity of any deterministic protocol which correctly computes F on every input.

Decision-tree complexity. In the (Boolean) decision-tree model, we wish to compute a function $f : \{0,1\}^p \to \mathcal{Z}$ when given query access to the input, and are charged for the total number of queries we make.

Formally, a *deterministic decision tree* $T : \{0,1\}^p \to \mathcal{Z}$ is a rooted binary tree where each internal node v is labeled with a variable number $i \in [p]$, each edge is labeled 0 or 1, and each leaf is labeled with an element of $\mathcal{Z}$. The execution of T on an input $z \in \{0,1\}^p$ traces a path in this tree: at each internal node v it queries the corresponding coordinate z_i and follows the edge labeled z_i. Whenever the algorithm reaches a leaf, it outputs the associated label and terminates. We say that T *correctly computes* f on z if this label equals $f(z)$.

The *query complexity* of T is the height of the tree. The *deterministic query complexity* of f, denoted $\mathcal{D}^{dt}(F)$, is defined as the smallest query complexity of any deterministic decision tree which correctly computes f on every input.

Functions of interest. The *Inner-product* function on n bits, denoted IP_n, is defined on $\{0,1\}^n \times \{0,1\}^n$ to be:

$$\mathsf{IP}_n(x,y) = \sum_{i \in [n]} x_i \cdot y_i \mod 2.$$

Whenever n is a power of 2, the *Indexing* function on n bits, IND_n, is defined on $\{0,1\}^{\log n} \times \{0,1\}^n$ to be:

$$\mathsf{IND}_n(x,y) = y_x \quad (\text{the } x\text{'th bit of } y).$$

Let n be a natural number and $\gamma = \frac{k}{n}$ where k is an integer in the interval $[1, n/2 - 1]$ (This implies $\gamma \in (0, 1/2)$.) For two n-bit strings x and y, let $d_H(x,y) = \sum_i x_i \oplus y_i$ be their Hamming

distance. The *gap-Hamming problem* on n bits, denoted $\mathsf{GH}_{n,\gamma}$, is a promise problem defined on $\{0,1\}^n \times \{0,1\}^n$, by the condition

$$\mathsf{GH}_{n,\gamma}(x,y) = \begin{cases} 1 & \text{if } d_H(x,y) \geq (\frac{1}{2}+\gamma)\, n, \\ 0 & \text{if } d_H(x,y) \leq (\frac{1}{2}-\gamma)\, n. \end{cases}$$

3. Deterministic simulation theorem

A *simulation theorem* shows how to construct a decision tree for a function f from a communication protocol for a composition problem $f \circ g^p$. Such a theorem can also be called a *lifting* theorem, if one wishes to emphasize that lower bounds for the decision-tree complexity of f can be *lifted* to lower bounds for the communication complexity of $f \circ g^p$. As mentioned in Section 1, the deterministic lifting theorem proved in (Raz & McKenzie 1999), and subsequently simplified in (Göös, Pitassi & Watson 2015), uses IND_n as inner function g with n being polynomially larger than p. In this section, we will show a deterministic simulation theorem for any function which possesses a certain pseudo-random property, which we will now define. Later, we will show that the inner product and any function of the gap-Hamming family have this property.

Definition 3.1 (Hitting rectangle distributions). *Let $0 \leq \delta < 1$ be a real, $h \geq 1$ be an integer, and $\mathcal{A}, \mathcal{B}$ be some sets. A distribution σ over rectangles within $\mathcal{A} \times \mathcal{B}$ is called a* (δ, h)-hitting rectangle distribution *if, for any rectangle $A \times B$ with $|A|/|\mathcal{A}|, |B|/|\mathcal{B}| \geq 2^{-h}$,*

$$\Pr_{R \sim \sigma}[R \cap (A \times B) \neq \varnothing] \geq 1 - \delta.$$

Let $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be a (possibly partial) function. A rectangle $A \times B$ is c-monochromatic with respect to g if $g(a,b) = c$ for every $(a,b) \in A \times B$.

Definition 3.2. *For a real $\delta \geq 0$ and an integer $h \geq 1$, we say that a (possibly partial) function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ has* (δ, h)-hitting monochromatic rectangle distributions *if there are two (δ, h)-hitting rectangle distributions σ_0 and σ_1, where each σ_c is a*

distribution over rectangles within $\mathcal{A}\times\mathcal{B}$ *that are c-monochromatic with respect to* g.

The theorem we will prove in Section 3.2 is the following:

THEOREM 3.3 (Theorem 1.1 restated). *Let* $\varepsilon \in (0,1)$ *and* $\delta \in (0, \frac{1}{100})$ *be real numbers, and let* $h \geq 6/\varepsilon$ *and* $1 \leq p \leq 2^{h(1-\varepsilon)}$ *be integers. Let* $f : \{0,1\}^p \to \mathcal{Z}$ *be a function and* $g : \mathcal{A}\times\mathcal{B} \to \{0,1\}$ *be a (possibly partial) function. If* g *has* (δ, h)*-hitting monochromatic rectangle distributions, then*

$$\mathcal{D}^{dt}(f) \leq \frac{4}{\varepsilon \cdot h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

In Section 5, we will show that $\mathsf{GH}_{n,\frac{1}{4}}$ has $(o(1), \frac{n}{100})$-hitting monochromatic rectangle distributions. From this, we obtain a simulation theorem for $\mathsf{GH}_{n,\frac{1}{4}}$:

COROLLARY 3.4. *Let* n *be a large enough even integer,* $\varepsilon \in (0,1)$, *and* $p \leq 2^{\frac{n}{100}(1-\varepsilon)}$ *be an integer. For any function* $f : \{0,1\}^p \to \mathcal{Z}$, $\mathcal{D}^{dt}(f) \leq \frac{400}{n\varepsilon} \cdot \mathcal{D}^{cc}(f \circ \mathsf{GH}^p_{n,\frac{1}{4}})$.

In Section 6, we will show that IP_n has $(o(1), n(\frac{1}{2} - \varepsilon))$-hitting monochromatic rectangle distributions, for any constant $\varepsilon \in (0, 1/2)$. This allows us to derive after some simple calculations:

COROLLARY 3.5. *Let* n *be large enough integer,* $\varepsilon \in (0, 1/2)$ *be a constant real, and* $p \leq 2^{(\frac{1}{2}-\varepsilon)n}$ *be an integer. For any function* $f : \{0,1\}^p \to \mathcal{Z}$, $\mathcal{D}^{dt}(f) \leq \frac{36}{n\varepsilon} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IP}^p_n)$.

These two corollaries together imply[2] Theorem 1.3. This allows us to significantly improve the gadget size known for simulation theorem of (Göös, Pitassi & Watson 2015; Raz & McKenzie 1999), that uses the indexing function instead of inner product. Indeed, Jakob Nordström (Nordström 2016) recently posed to us the challenge of

[2]The constant $\frac{1}{4}$ for $\mathsf{GH}^p_{n,\frac{1}{4}}$ in Corollary 3.4 is arbitrary. For any gap $\zeta \leq \frac{1}{2}$, we can show for $\mathsf{GH}^p_{n,\zeta}$ a $(2^{-n(1-H(\frac{1}{2}-\frac{\zeta}{4}))}, (1 - H(\frac{1}{2} - \frac{\zeta}{4}))n)$-hitting monochromatic distribution, where $H(\cdot)$ is the binary entropy function.

proving a simulation theorem for $f \circ \mathsf{IND}_n^p$, with a gadget size n smaller than p^3; note that p^3 is already a significant improvement over (Göös, Pitassi & Watson 2015; Raz & McKenzie 1999).

This follows from the above corollary, because of the following reduction: Given an instance $(a, b) \in \{0,1\}^{mp} \times \{0,1\}^{mp}$ of $f \circ \mathsf{IP}_m^p$ where $p \leq 2^{m(\frac{1}{2}-\varepsilon)}$, Alice and Bob can construct an instance of $f \circ \mathsf{IND}_n^p$ where $n = 2^m$. Bob converts his input $b \in \{0,1\}^{mp}$ to $b' \in \{0,1\}^{np}$, so that each $b'_i = [\mathsf{IP}_n(x_1, b_i)\rangle, \ldots, \mathsf{IP}_n(x_n, b_i)\rangle]$ where $\{x_1, \ldots, x_n\} = \{0,1\}^m$ is an ordering of all m-bit strings. It is easy to see that $\mathsf{IP}_m(a_i, b_i) = \mathsf{IND}_n(a_i, b'_i)$. Hence, it follows as a corollary to our result for IP:

Corollary 3.6. *Let $\varepsilon \in (0, 1/2)$ be a constant real number, and n and p be sufficiently large natural numbers, such that $p \leq n^{\frac{1}{2}-\varepsilon}$. Then, for any function $f : \{0,1\}^p \to \mathcal{Z}$, $\mathcal{D}^{dt}(f) = \frac{36}{\varepsilon \cdot \log n} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IND}_n^p)$.*

Also, it is worth noting that the proof of Lemma 7 in (Göös, Pitassi & Watson 2015), which Göös *et al.* call the 'Projection Lemma', implicitly proves that IND_n has $(\frac{1}{150}, \frac{3}{20} \log n)$-hitting rectangle distribution. Here the c-monochromatic rectangle distribution (c is either 1 or 0) is sampled as follows: Alice samples a subset of indices $U \subset [n]$ of size $n^{7/20}$, and Bob picks $V \subset \{0,1\}^n$ where $V = \{b \mid b_j = c \text{ for all } j \in U\}$.[3] Hence, we can also apply Theorem 3.3 directly to obtain a corollary similar to Corollary 3.6 (albeit with much larger gadget size n). See Section 4 for a detailed derivation.

3.1. Thickness and its properties. In this section, we list several properties related to *'thickness'*, a combinatorial property which will be needed in Section 3.2 to prove a simulation theorem. Readers may also refer to (Göös, Pitassi & Watson 2015).

[3]Readers may note that δ in the proof of Claim 9 of (Göös, Pitassi & Watson 2015) is 1/4, where as we need $\delta < 1/100$. This is not a problem, as we can make δ as small a constant as we wish for by the same calculation as that in the proof of Claim 9.

Definition 3.7 (Aux graph, average and min degrees). *Let $p \geq 2$. For $i \in [p]$ and $A \subseteq \mathcal{A}^p$, the aux graph $G(A, i)$ is the bipartite graph with left side vertices A_i, right side vertices $A_{\neq i}$ and edges corresponding to the set A, i.e., (a', a'') is an edge iff $a' \times_{\{i\}} a'' \in A$.*

We define the average degree of $G(A, i)$ to be the average right degree:

$$d_{\text{avg}}(A, i) = \frac{|A|}{|A_{\neq i}|},$$

and the min-degree of $G(A, i)$, to be the minimum right degree:

$$d_{\min}(A, i) = \min_{a' \in A_{\neq i}} |\mathsf{Ext}(a')|.$$

Definition 3.8 (Thickness and average thickness). *For $p \geq 2$ and $\tau, \varphi \in (0, 1)$, a set $A \subseteq \mathcal{A}^p$ is called τ-*thick *if $d_{\min}(A, i) \geq \tau \cdot |\mathcal{A}|$ for all $i \in [p]$. (Note, an empty set A is τ-thick.) Similarly, A is called φ-*average-thick *if $d_{\text{avg}}(A, i) \geq \varphi \cdot |\mathcal{A}|$ for all $i \in [p]$. For a rectangle $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$, we say that the rectangle $A \times B$ is τ-thick if both A and B are τ-thick. For $p = 1$, set $A \subseteq \mathcal{A}$ is τ-thick if $|A| \geq \tau \cdot |\mathcal{A}|$.*

The following property is from (Göös, Pitassi & Watson 2015, Lemma 6). Informally it says that if we can maintain high average-thickness of a set, then there is a large enough subset of it which has high thickness. Looking ahead, this will be useful while traveling down the protocol tree where we only have to worry about maintaining high average-thickness. For completeness, we also include the proof.

Lemma 3.9 (Average thickness implies thickness). *For any $p \geq 2$, if $A \subseteq \mathcal{A}^p$ is φ-average-thick, then for every $\delta \in (0, 1)$ there is a $\frac{\delta}{p}\varphi$-thick subset $A' \subseteq A$ with $|A'| \geq (1 - \delta)|A|$.*

Proof. The set A' is obtained by running Algorithm 1.

Algorithm 1

1: Set $A^0 = A$, $j = 0$.
2: **while** $d_{\min}(A^j, i) < \frac{\delta}{p}\varphi \cdot |\mathcal{A}|$ for some $i \in [p]$ **do**
3: Let a' be a right node of $G(A^j, i)$ with nonzero degree less than $\frac{\delta}{p}\varphi \cdot |\mathcal{A}|$.
4: Set $A^{j+1} = A^j \backslash \{a'\} \times_i \mathsf{Ext}(a')$, i.e., remove every extension of a'. Increment j.
5: Set $A' = A^j$.

The total number of iteration of the algorithm is at most $\sum_{i\in[p]} |A_{\neq i}|$. (We remove at least one node in some $G(A^j, i)$ in each iteration which was a node also in the original $G(A, i)$.) So the number of iterations is at most

$$\sum_{i\in[p]} |A_{\neq i}| = \sum_{i\in[p]} \frac{|A|}{d_{\text{avg}}(A, i)} \leq \frac{p|A|}{\varphi \cdot |\mathcal{A}|}.$$

As the algorithm removes at most $\frac{\delta}{p}\varphi \cdot |\mathcal{A}|$ elements of A in each iteration, the total number of elements removed from A is at most $\delta|A|$, so $|A'| \geq (1-\delta)|A|$. Hence, the algorithm always terminates with a non-empty set A' that must be $\frac{\delta}{p}\varphi$-thick. □

LEMMA 3.10. *Let $p \geq 2$ be an integer, $i \in [p]$, $A \subseteq \mathcal{A}^p$ be a τ-thick set, and $S \subseteq \mathcal{A}$. The set $A^{i,S}_{\neq i}$ is τ-thick. $A^{i,S}_{\neq i}$ is empty iff $S \cap A_i$ is empty.*

PROOF. Notice that $A^{i,S}_{\neq i}$ is non-empty iff $S \cap A_i$ is non-empty. Consider the case of $p \geq 3$. Let $a \in A$, where $a_i \in S$. Set $a' = a_{\neq i}$. For $j' \in [p-1]$, let $j = j'+1$ if $j' \geq i$, and $j = j'$ otherwise. Clearly, $\mathsf{Ext}^{\{j\}}_A(a_{\neq j}) \subseteq \mathsf{Ext}^{\{j'\}}_{A^{i,S}_{\neq i}}(a'_{\neq j'})$; hence, the degree of a' in $G(A^{i,S}_{\neq i}, j')$ is at least the degree of a in $G(A, j)$ which is at least $\tau \cdot |\mathcal{A}|$. Hence, $A^{i,S}_{\neq i}$ is τ-thick.

To see the case $p = 2$, assume there is some string $a' \in A_{\neq i}$ which has some extension $a'' \in S$, but A itself is τ-thick, so there have to be at least $\tau \cdot |\mathcal{A}|$ many such a', which will then all be in $A^{i,S}_{\neq i}$. □

The next lemma is the heart of the proof of the simulation theorem. To provide context, recall from Section 1.4, we will traverse down the protocol tree maintaining high average-thickness over the coordinates which are not queried yet which, in turn, will guarantee high thickness over those coordinates, thanks to Lemma 3.9. We may end up in a situation where we do not have high average-thickness anymore, and we have to issue a query. The following lemma provides a way to gain back thickness in the unqueried coordinates irrespective of the value of the query issued.

LEMMA 3.11. *Let $h \geq 1$, $p \geq 2$ and $i \in [p]$ be integers and $\delta, \tau, \varphi \in (0,1)$ be reals, where $\tau \geq 2^{-h}$. Consider a function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ which has (δ, h)-hitting monochromatic rectangle distributions. Suppose $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ is a non-empty rectangle which is τ-thick, and suppose also that $d_{\text{avg}}(A, i) \leq \varphi \cdot |\mathcal{A}|$. Then for any $c \in \{0,1\}$, there is a c-monochromatic rectangle $U \times V \subseteq \mathcal{A} \times \mathcal{B}$ such that*

(i) $A_{\neq i}^{i,U}$ and $B_{\neq i}^{i,V}$ is τ-thick,

(ii) $\alpha_{\neq i}^{i,U} \geq \frac{1}{\varphi}(1 - 3\delta)\alpha$,

(iii) $\beta_{\neq i}^{i,V} \geq (1 - 3\delta)\beta$,

where $\alpha = |A|/|\mathcal{A}|^p$, $\beta = |B|/|\mathcal{B}|^p$, $\alpha_{\neq i}^{i,U} = |A_{\neq i}^{i,U}|/|\mathcal{A}|^{p-1}$ and $\beta = |B_{\neq i}^{i,V}|/|\mathcal{B}|^{p-1}$.

The constant 3 in the statement may be replaced by any value greater than 2, so the lemma is still meaningful for δ arbitrarily close to $1/2$.

PROOF. Fix $c \in \{0,1\}$. Consider a matrix M where rows correspond to strings $a \in A_{\neq i}$, and columns correspond to rectangles $R = U \times V$ in the support of σ_c. Set each entry $M(a, R)$ to 1 if $U \cap \mathsf{Ext}_A^{\{i\}}(a) \neq \emptyset$, and set it to 0 otherwise.

For each $a \in A_{\neq i}$, $|\mathsf{Ext}_A^{\{i\}}(a)| \geq \tau|\mathcal{A}|$, and because σ_c is a (δ, h)-hitting rectangle distribution and $\tau \geq 2^{-h}$, we know that if we pick a column R according to σ_c, then $M(a, R) = 1$ with probability

$\geq 1-\delta$. So the probability that $M(a,R)=1$ over uniform a and σ_c-chosen R is $\geq 1-\delta$.

Call a column of M *A-good* if $M(a,R)=1$ for at least $1-3\delta$ fraction of the rows a. Now it must be the case that the A-good columns have strictly more than $1/2$ of the σ_c-mass. Suppose not. The expected number of 0's in each column is at most δ. So, by Markov's inequality, the fraction of columns which has at least 3δ fraction of 0's is at most $1/3$. This means that at least $2/3 > 1/2$ fraction of columns will have at least $1-3\delta$ fraction of 1's.

A similar argument also holds for Bob's set $B_{\neq i}$. Hence, there is a c-monochromatic rectangle $R = U \times V$ whose column is both A-good and B-good in their respective matrices. This is our desired rectangle R.

We know: $|A^{i,U}_{\neq i}| \geq (1-3\delta)|A_{\neq i}|$ and $|B^{i,V}_{\neq i}| \geq (1-3\delta)|B_{\neq i}|$. Since $|B_{\neq i}| \geq \frac{|B|}{|\mathcal{B}|}$, we obtain $|B^{i,V}_{\neq i}|/|\mathcal{B}|^{p-1} \geq (1-3\delta)|B_{\neq i}|/|\mathcal{B}|^{p-1}$ which is at least $(1-3\delta)\beta$. Because $|A|/|A_{\neq i}| \leq \varphi|\mathcal{A}|$, we get

$$\frac{|A_{\neq i}|}{|\mathcal{A}|^{(p-1)}} \geq \frac{1}{\varphi} \cdot \frac{|A|}{|\mathcal{A}|^p} = \frac{\alpha}{\varphi}.$$

Combined with the lower bound on $|A^{i,U}_{\neq i}|$, we obtain $|A^{i,U}_{\neq i}|/|\mathcal{A}|^{p-1} \geq (1-3\delta)\alpha/\varphi$. The thickness of $A^{i,U}_{\neq i}$ and $B^{i,V}_{\neq i}$ follows from Lemma 3.10. □

The next lemma will be used as a closing argument for the proof of the simulation theorem. At the end of our traversal down the protocol tree, when we land on a leaf, we will be left with a rectangle which is thick on all unqueried coordinates. The next lemma says that, for any instantiation of these coordinates, there is an input pair inside the rectangle which, when g applied on it, will have those values in the corresponding coordinates.

LEMMA 3.12. *Let $p, h \geq 1$ be integers and $\delta, \tau \in (0,1)$ be reals, where $\tau \geq 2^{-h}$. Consider a function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ which has (δ, h)-hitting monochromatic rectangle distributions. Let $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ be a τ-thick non-empty rectangle. Then for every $z \in \{0,1\}^p$ there is some $(a,b) \in A \times B$ with $g^p(a,b) = z$.*

PROOF. This follows from repeated use of Lemma 3.10. Fix arbitrary $z \in \{0,1\}^p$. Set $A^{(1)} = A$ and $B^{(1)} = B$. We proceed in rounds $i = 1, \ldots, p-1$ maintaining a τ-thick rectangle $A^{(i)} \times B^{(i)} \subseteq \mathcal{A}^{p-i+1} \times \mathcal{B}^{p-i+1}$. If we pick $U_i \times V_i$ from σ_{z_i}, then the rectangle $(A^{(i)})_{\{i\}} \cap U_i \times (B^{(i)})_{\{i\}} \cap V_i$ will be non-empty with probability $\geq 1 - \delta > 0$ (because σ_{z_i} is a (δ, h)-hitting rectangle distribution and $\tau \geq 2^{-h}$). Fix such U_i and V_i. Set a_i to an arbitrary string in $(A^{(i)})_{\{i\}} \cap U_i$, and b_i to an arbitrary string in $(B^{(i)})_{\{i\}} \cap V_i$. Set $A^{(i+1)} = (A^{(i)})^{i,\{a_i\}}_{\neq i}$, $B^{(i+1)} = (B^{(i)})^{i,\{b_i\}}_{\neq i}$, and proceed for the next round. By Lemma 3.10, $A^{(i+1)} \times B^{(i+1)}$ is τ-thick.

Eventually, we are left with a rectangle $A^{(p)} \times B^{(p)} \subseteq \mathcal{A} \times \mathcal{B}$ where both $A^{(p)}$ and $B^{(p)}$ are τ-thick (and non-empty). Again with probability $1 - \delta > 0$, the z_p-monochromatic rectangle $U_p \times V_p$ chosen from σ_{z_p} will intersect $A^{(p)} \times B^{(p)}$. We again set a_p and b_p to come from the intersection, and set $a = \langle a_1, a_2, \ldots, a_p \rangle$ and $b = \langle b_1, b_2, \ldots, b_p \rangle$. □

3.2. Proof of the simulation theorem. Now we are ready to present the proof of the simulation theorem (Theorem 3.3). Let $\varepsilon \in (0, 1/2)$ and $\delta \in (0, 1/100)$ be real numbers, and $h \geq 6/\varepsilon$ and $1 \leq p \leq 2^{h(1-\varepsilon)}$ be integers. Let $f : \{0,1\}^p \to \mathcal{Z}$ be a function and $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be a (possibly partial) function. Assume that g has (δ, h)-hitting monochromatic rectangle distributions. We assume we have a communication protocol Π for solving $f \circ g^p$, and we will use Π to construct a decision tree (procedure) for f. Let C be the communication cost of the protocol Π. If $p \leq 5C/h$, the theorem is true trivially. So assume $p > 5C/h$. Set $\varphi = 4 \cdot 2^{-\varepsilon h}$ and $\tau = 2^{-h}$. The decision-tree procedure is presented in Algorithm 2 (page 25). On an input $z \in \{0,1\}^p$, it uses the protocol Π to decide which bits of z to query.

An informal description of simulation algorithm. Given an input $z \in \{0,1\}^p$, the algorithm starts traversing a path from the root of the protocol tree of Π. The variable v indicates the node of the protocol tree which is the current node during the ongoing simulation. Associated with v, the algorithm maintains a rectangle $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ and a set $I \subseteq [p]$ of indices. I corresponds to coordinates of the input z that were not queried, yet.

Throughout the execution of the algorithm, the following invariants are maintained: The set $A \times B$ is thick in the coordinates I, and every pair of inputs $(x, y) \in A \times B$ is consistent with the answer to the queries made so far. To start off, I is $[p]$, and $A \times B = \mathcal{A}^p \times \mathcal{B}^p$. So the invariants are trivially maintained at the beginning.

In each iteration of the simulation, the algorithm checks the following condition: Are both A_I and B_I φ-average-thick? Depending on the answer to this check, the algorithm does one of the following two things:

If both A_I and B_I are φ-average-thick, the algorithm proceeds to that child of v whose corresponding rectangle R_v has at least half the mass of $A_I \times B_I$ and applies Lemma 3.9 to prune the rectangle $(A \times B) \cap R_v$ to ensure the thickness condition. Note that the working set $A \times B$ loses a constant fraction of density in doing so.

Otherwise, if there is a coordinate i in I, where A_I or B_I has low average degree, then the algorithm queries z_i and, depending on the value of z_i, applies Lemma 3.11 accordingly. Lemma 3.11 crucially exploits the fact that A_I and B_I are thick in i-th coordinate and outputs a sub-rectangle of $A \times B$ which, in the i-th coordinate, is restricted to a z_i-monochromatic rectangle $U \times V$, while maintaining the thickness invariant in the coordinates $I \backslash \{i\}$. This also results in a boost in density of $A \times B$ in the current working universe $\mathcal{A}^{I \backslash i} \times \mathcal{B}^{I \backslash i}$. The algorithm updates I to be $I \backslash \{i\}$ and reiterates (i.e., does the average thickness check again on $A \times B$ in the coordinate of the new I). We describe the parameters of the algorithm next in more detail.

Correctness. The algorithm maintains an invariant that $A_I \times B_I$ is τ-thick. This invariant is trivially true at the beginning.

If both A_I and B_I are φ-average-thick, the algorithm finds sets A' and B' on lines 5–7 as follows. Consider the case that Alice communicates at node v. She is sending one bit. Let A_0 be inputs from A on which Alice sends 0 at node v and $A_1 = A \backslash A_0$. We can pick $c \in \{0, 1\}$ such that $|(A_c)_I| \geq |A_I|/2$. Set $A'' = A_c$. Since A_I is φ-average-thick, A''_I is $\varphi/2$-average-thick. So using Lemma 3.9 on A''_I with δ set to $1/2$, we can find a subset A' of A'' such that A'_I is $\frac{\varphi}{4 \cdot |I|}$-thick and $|A'_I| \geq |A''_I|/2$. ($A' \subseteq A''$ will be the pre-

Algorithm 2 Decision-tree procedure

Input: $z \in \{0,1\}^p$

Output: $f(z)$

Set v to be the root of the protocol tree for Π, $I = [p]$, $A = \mathcal{A}^p$ and $B = \mathcal{B}^p$.
while v is not a leaf **do**
  if A_I and B_I are both φ-average-thick **then**
    Let v_0, v_1 be the children of v.
    Choose $c \in \{0,1\}$ for which there is $A' \times B' \subseteq (A \times B) \cap R_{v_c}$ such that
    (1) $|A'_I \times B'_I| \geq \frac{1}{4}|A_I \times B_I|$
    (2) $A'_I \times B'_I$ is τ-thick. ▷ Using Lemma 3.9
    Update $A = A'$, $B = B'$ and $v = v_c$.
  else if $d_{\text{avg}}(A_I, j) < \varphi|\mathcal{A}|$ for some $j \in [|I|]$ **then**
    Query z_i, where i is the j-th (smallest) element of I.
    Let $U \times V$ be a z_i-monochromatic rectangle of g such that
    (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
    (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\alpha$,
    (3) $\beta^{i,V}_{I\setminus\{i\}} \geq (1-3\delta)\beta$, ▷ Using Lemma 3.11
    Update $A = A^{i,U}, B = B^{i,V}$ and $I = I\setminus\{i\}$.
  else if $d_{\text{avg}}(B_I, j) < \varphi|\mathcal{B}|$ for some $j \in [|I|]$ **then**
    Query z_i, where i is the j-th (smallest) element of I.
    Let $U \times V$ be a z_i-monochromatic rectangle of g such that
    (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
    (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq (1-3\delta)\alpha$,
    (3) $\beta^{i,V}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\beta$, ▷ Using Lemma 3.11
    Update $A = A^{i,U}, B = B^{i,V}$ and $I = I\setminus\{i\}$.
Output $f \circ g^p(A \times B)$.

image of A'_I obtained from the lemma.) Since $\varphi = 4 \cdot 2^{-\varepsilon h}$ and $|I| \leq p \leq 2^{h(1-\varepsilon)}$, the set A'_I will be 2^{-h}-thick, i.e., τ-thick. Setting $B' = B$, the rectangle $A' \times B'$ satisfies properties from lines 6–7. A similar argument holds when Bob communicates at node v.

If A_I is not φ-average-thick, the existence of $U \times V$ at line 11 is guaranteed by Lemma 3.11. Similarly in the case when B_I is not φ-average-thick.

Next we argue that the number of queries made by Algorithm 2 is at most $5C/\varepsilon h$. In the first part of the **while** loop (lines 3–8), the density of the current $A_I \times B_I$ drops by a factor 4 in each iteration. There are at most C such iterations; hence, this density can drop by a factor of at most $4^{-C} = 2^{-2C}$. For each query that the algorithm makes, the density of the current $A_I \times B_I$ increases by a factor of at least $(1-3\delta)^2/\varphi \geq \frac{1}{2\varphi} \geq 2^{\varepsilon h - 3}$. (Here we use the fact that $\delta \leq 1/100$.) Since the density can be at most one, the number of queries is upper bounded by

$$\frac{2C}{\varepsilon h - 3} \leq \frac{4C}{\varepsilon h}, \qquad \text{when } h \geq 6/\varepsilon.$$

Finally, we argue that $f(A \times B)$ at the termination of Algorithm 2 is the correct output. Given an input $z \in \{0,1\}^p$, whenever the algorithm queries any z_i, the algorithm makes sure that all the input pairs (x, y) in the rectangle $A \times B$ are such that $g(x_i, y_i) = z_i$ — because $U \times V$ is always a z_i-monochromatic rectangle of g. At the termination of the algorithm, I is the set of i such that z_i was not queried by the algorithm. As $p > 4C/\varepsilon h$, I is non-empty. Since $A_I \times B_I$ is τ-thick, it follows from Lemma 3.12 that $A \times B$ contains some input pair (x, y) such that $g^{|I|}(x_I, y_I) = z_I$, and so $g^p(x, y) = z$. Since Π is correct, it must follow that $f(z) = f \circ g^p(A \times B)$. This concludes the proof of correctness. □

With greater care the same argument allows for δ to be close to $\frac{1}{2}$. This would require also tightening the $1 - 3\delta$ factors appearing in Lemma 3.11 to something close to $1 - 2\delta$ and make the calculations (only) slightly longer. Although we noticed this improvement, we found no use for it, so we opted to keep the above presentation.

4. Hitting rectangle distribution for IND

Here we derive the $(\frac{1}{150}, \frac{3}{20}\log n)$-hitting monochromatic rectangle distribution for IND_n. Consider the following distribution σ_c over c-monochromatic rectangles: Alice samples a subset of indices $U \subset [n]$ of size $n^{7/20}$, and Bob picks $V \subset \{0,1\}^n$ where $V = \{b \mid b_j = c \text{ for all } j \in U\}$. We next show the following lemma.

LEMMA 4.1. *The distribution σ_c, for $c \in \{0,1\}$, is a $(\frac{1}{150}, \frac{3}{20}\log n)$-hitting c-monochromatic distribution for IND_n.*

The proof of this lemma is implicit in the proof of Lemma 7 (Projection lemma) of (Göös, Pitassi & Watson 2015). They show the following properties of σ_c in the course of proving their Lemma 7.

LEMMA 4.2. *If $U \times V$ is sampled from σ_c, then*

(i) For any set $A' \subseteq [n]$ that has size at least $n^{17/20}$, $\Pr_U[A' \cap U \neq \varnothing] \geq 1 - e^{-n^{1/5}}$,

(ii) For any set $B' \subseteq \{0,1\}^n$ with $\frac{|B'|}{2^n} \geq 2^{-n^{11/20}}$, $\Pr_U[B' \cap V \neq \varnothing] \geq exp(-14(n^{-2/20} + n^{-6/20}))$.

The inverse exponential term on RHS is lower bounded by 3/4 in (Göös, Pitassi & Watson 2015). We can bound this term by 199/200 as well. Hence, for this distribution, $\delta \leq 1/200 + e^{-n^{1/5}} \leq 1/150$.

Now we bound h. We have $\frac{|A'|}{n} \geq n^{-3/20} = 2^{-\frac{3}{20}\log n}$ from property (1). The bound on the size of B' comes from property (2), which is much smaller compared to $\frac{|A'|}{n}$. Hence we have $h = \frac{3}{20}\log n$.

5. Hitting rectangle distributions for GH

We construct a hitting monochromatic rectangle distribution for $\mathsf{GH}_{n,\frac{1}{4}}$. Subsequently, we will show that the distribution is $(2^{-\frac{n}{100}}, \frac{n}{100})$ hitting rectangle distribution which will show a deterministic simulation result when the inner function is $\mathsf{GH}_{n,\frac{1}{4}}$, i.e.,

$$\mathcal{D}^{cc}(f \circ \mathsf{GH}^p_{n,\frac{1}{4}}) \geq \mathcal{D}^{dt}(f) \cdot \Omega(n).$$

Let $d_H(x, y)$ denotes the Hamming distance between the strings x and y. Let $B_r(x)$ be the Hamming ball of radius r around x, i.e., $B_r(x) = \{y \in \{0,1\}^n \mid d_H(x, y) \leq r\}$; for a set $A \subset \{0,1\}^n$, $B_r(A) = \cup_{a \in A} B_r(a)$.

Let $\varepsilon = \frac{1}{8}$ and $\mathcal{H}$ be the set of all strings in $\{0,1\}^n$ with Hamming weight $n/2$. Now consider the rectangle distributions σ_0 and σ_1 obtained from the following sampling procedure:

Sampling from σ_0: Choose a random string $x \in \mathcal{H}$. Now let $U_x = B_{\varepsilon n}(x)$; output $U_x \times U_x$.

Sampling from σ_1: Let $\bar{x} \in \mathcal{H}$ be the bitwise complement of x and $V_x = B_{\varepsilon n}(\bar{x})$. Output $U_x \times V_x$.

For the chosen value of ε, $U_x \times V_x$ is a 1-monochromatic rectangle, since for any $u \in U_x, v \in V_x$,

$$d_H(u, v) \geq n - 2\varepsilon n \geq \frac{3}{4}n.$$

On the other hand, $U_x \times U_x$ is 0-monochromatic, since for any $u, u' \in U_x$,

$$d_H(u, u') \leq 2\varepsilon n \leq \frac{1}{4}n.$$

Both inequalities are obtained by a straightforward application of triangle inequality.

LEMMA 5.1. *The distributions σ_0 and σ_1 are $(2^{-\frac{n}{100}}, \frac{n}{100})$-hitting monochromatic rectangle distributions for $\mathsf{GH}_{n,\frac{1}{4}}$.*

To prove Lemma 5.1, we need the following theorem due to Harper. We will call $S \subset \{0,1\}^n$ a *Hamming ball with center* $c \in \{0,1\}^n$ if $B_r(c) \subseteq S \subset B_{r+1}(c)$ for some nonnegative integer r. For sets $S, T \subset \{0,1\}^n$, we define the *distance* between S and T as $d(S, T) = \min\{d_H(s, t) \mid s \in S, t \in T\}$.

THEOREM 5.2. *Harper's theorem, (Frankl & Füredi 1981), (Harper 1966) Given any non-empty subsets S and T of $\{0,1\}^n$, there exist a Hamming ball S_0 with center $\bar{1}$ and Hamming ball T_0 with center $\bar{0}$ such that $|S| = |S_0|, |T| = |T_0|$ and $d(S_0, T_0) \geq d(S, T)$.*

Note that Theorem 5.2 also tells us when $B_r(S)$ is smallest for a set $S \subset \{0,1\}^n$ in the following way:

LEMMA 5.3. *Let $r \in [n]$ be any nonnegative integer and let $\mathcal{S}_k = \{S \subset \{0,1\}^n \mid |S| = k\}$ for any k. If S is a Hamming ball centered around either $\bar{1}$ or $\bar{0}$, then $|B_r(S)| \leq |B_r(S')|$ for any $S' \in \mathcal{S}_k$.*

PROOF. Fix any $k \leq n$. The cases when $k = 0$ and $k = n$ are trivial. Given a set $S \in \mathcal{S}_k$, let $T_S = \{0,1\}^n \backslash B_r(S)$. It is immediate that $d(S, T_S) = r + 1$. Now let us suppose that S is such that it achieves the smallest $B_r(S')$ among all $S' \in \mathcal{S}_k$. This also means that T_S is the biggest such set. Using Harper's theorem, we can find set S_0 and T_0 such that $d(S_0, T_0) \geq r + 1$ where S_0 is centered around $\bar{1}$ and T_0 is centered around $\bar{0}$ with $|S_0| = |S|$ and $|T_0| = |T_S|$. Now it is easy to see that $T_0 \subseteq \{0,1\}^n \backslash B_r(S_0)$, i.e., $|T_S| = |T_0| \leq |T_{S_0}|$, which is a contradiction. This means that $|B_r(S)|$ will be the smallest if S is a Hamming ball centered around $\bar{1}$. This proves the lemma. □

Now we state the proof of Lemma 5.1.

PROOF (Proof of Lemma 5.1). We will show that any set $A \subset \{0,1\}^n$ of size $|A| \geq 2^{\frac{99}{100}n}$ will be hit by U_x with probability $\geq 1 - 2^{-\frac{n}{100}}$. The lemma now follows since U_x and V_x have the same marginal distribution.

Let us first suppose that x is chosen uniformly at random from the entire Hamming cube. We first show that, for such an x, U_x does not intersect A with extremely small probability. Then it follows immediately that, when conditioned on the event that x is chosen uniformly at random from $\mathcal{H}$, the same result holds. To

this end, note that the event $U_x \cap A = \emptyset$ happens exactly when $x \notin B_{\varepsilon n}(A)$:

$$\Pr[U_x \cap A = \emptyset] = \Pr[x \notin B_{\varepsilon n}(A)] \leq \frac{2^n - |B_{\varepsilon n}(A)|}{2^n}.$$

From Lemma 5.3, we know that $|B_{\varepsilon n}(A)|$ is smallest when A is itself a Hamming ball around 0 of the same density as A, i.e., if, for some $\gamma \leq 1 - \varepsilon$, $|B_{\gamma n}(0)| \leq |A|$, then

$$|B_{\varepsilon n}(A)| \geq |B_{\varepsilon n}(B_{\gamma n}(0))| = |B_{(\gamma+\varepsilon)n}(0)|.$$

Next we argue that if A is large enough, then the smallest such $B_{\varepsilon n}(A)$ set is big enough to include a good fraction of x. Now we estimate the value of γ. For $\gamma = \frac{1}{2} - \frac{\varepsilon}{2} = \frac{1}{2} - \frac{1}{16}$, and since $H(\gamma) < \frac{98}{99}$, we have

$$|B_{\gamma n}(0)| \leq 2^{H(\gamma)n} \leq 2^{\frac{98}{99}n} \leq |A|.$$

And so $|B_{\varepsilon n}(A)| \geq |B_{(\gamma+\varepsilon)n}(0)| = |B_{\frac{n}{2}+\frac{n}{16}}(0)| \geq 2^n - |B_{\frac{n}{2}-\frac{n}{16}}(1)| \geq 2^n - 2^{\frac{98}{99}n}$. As promised, this is a large set. It now follows

$$\Pr[U_x \cap A = \emptyset] \leq \frac{2^{\frac{98}{99}n}}{2^n} \leq 2^{-\frac{n}{99}}.$$

Now if we condition on $x \in \mathcal{H}$, then we get

$$\begin{aligned}\Pr[U_x \cap A = \emptyset \mid x \in \mathcal{H}] &\leq \frac{\Pr[U_x \cap A = \emptyset]}{\Pr[x \in \mathcal{H}]} \\ &\leq 2^{-\frac{n}{99}} \cdot \sqrt{\pi \cdot \frac{n}{2}} \leq 2^{-\frac{n}{100}}. \qquad \square\end{aligned}$$

6. Hitting rectangle distributions for IP

In this section, we first construct the hitting monochromatic rectangle distribution for IP_n. We will then show that IP_n has $(4 \cdot 2^{-n/20}, n/5)$-hitting monochromatic rectangle distributions. This will show a deterministic simulation result when the inner function is IP_n, i.e.,

$$\mathcal{D}^{cc}(f \circ \mathsf{IP}_n^p) \geq \mathcal{D}^{dt}(f) \cdot \Omega(n).$$

We define the distributions σ_0 and σ_1 by the following sampling methods:

Sampling from σ_0: We choose a uniformly random $\frac{n}{2}$-dimensional subspace V of $\mathbb{F}_2^n$, and let $V^\perp$ be its orthogonal complement; output $V \times V^\perp$.

Sampling from σ_1: First we pick $a \in \{0,1\}^n$ uniformly at random conditioned on the fact that a has odd Hamming weight; then, we pick random subspace W of dimension $(n-1)/2$ from $a^\perp$, and let $W^\perp$ be the orthogonal complement of W *inside* $a^\perp$. We output $V \times V^{\|}$, where $V = a + W$ and $V^{\|} = a + W^\perp$.

LEMMA 6.1. *For all $0 < \varepsilon < 1/2$ and every sufficiently large n, the distributions σ_0 and σ_1 are IP_n has $(2 \cdot 2^{-\frac{\varepsilon}{4}n}, (\frac{1}{2} - \varepsilon)n)$-hitting monochromatic rectangle distributions.*

To prove this, we use the well-known second-moment method. The idea is the following: Let us consider the distribution σ_0 and consider a large enough rectangle $A \times B$. We show that a random $\frac{n}{2}$-dimensional subspace V intersects A with very high probability. Moreover, the intersection size is concentrated around its mean—this follows from pairwise independence of the indicator variables $[x \in V]$, for different x, using which we may use concentration bounds to complete the argument. The orthogonal complement of V has the same marginal distribution as that of V, and hence, a similar argument will follow for the intersection of $V^\perp$ and B. For σ_1, we have use a similar but slightly more delicate argument. Below we will formalize both of the arguments. We start with the following well-known variant of Chebyshev's inequality. Readers can choose to skip to Lemmas 6.4 and 6.5 if needed.

PROPOSITION 6.2 (Second-moment method). *Suppose that $X_i \in [0,1]$ and $X = \sum_i X_i$ are random variables. Suppose also that for all i and j, X_i and X_j are* anti-correlated, *in the sense that*

$$\mathbf{E}[X_i X_j] \leq \mathbf{E}[X_i] \cdot \mathbf{E}[X_j].$$

Then X is well-concentrated around its mean, namely for every ε:

$$\Pr[X \in \mu(1 \pm \varepsilon)] \geq 1 - \frac{1}{\varepsilon^2 \mu}.$$

All of the rectangle distributions rely on the following fundamental anti-correlation property:

LEMMA 6.3 (Hitting probabilities of random subspaces). *Let $0 \leq d \leq n$ be natural numbers. Fix any $v \neq w$ in $\mathbb{F}_2^n$, and pick a random subspace V of dimension d. Then the probability that $v \in V$ is exactly*

$$p_v = \begin{cases} \frac{2^d - 1}{2^n - 1} & \text{if } v \neq 0 \\ 1 & \text{if } v = 0. \end{cases}$$

And the probability that both $v, w \in V$ is exactly

$$p_{v,w} = \begin{cases} \binom{2^d-1}{2} \Big/ \binom{2^n-1}{2} & \text{if } v, w \neq 0 \\ p_v & \text{if } w = 0, \text{ and} \\ p_w & \text{if } v = 0. \end{cases}$$

Hence, it always holds that $p_{v,w} \leq p_v p_w$.

PROOF. The case when v or w are 0 is trivial. The value $p_v = \Pr[v \in V]$ for a random subspace V of dimension d equals $\Pr[Mv = 0]$ for a random non-singular $(n-d) \times n$ matrix M, letting $V = \ker M$. For any $v \neq 0, v' \neq 0$, M will have the same distribution as MN, where N is some fixed linear bijection of F_2^n mapping v to v'; it then follows that $p_v = p_{v'}$ always. But then

$$\sum_{v \neq 0} p_v = \mathbf{E}\left[\sum_{v \neq 0} [v \in V]\right] = 2^d - 1,$$

and since all p_v's are equal, then $p_v = \frac{2^d - 1}{2^n - 1}$.

Now let $p_{v,w} = \Pr[v \in V, w \in V]$. In the same way, we can show that $p_{v,w} = p_{v',w'}$ for all two such pairs, since a linear bijection will

exist mapping v to v' and w to w' (because every $v \neq w$ is linearly independent in $\mathbb{F}_2^n$). And now

$$\sum_{v,w\neq 0} p_{v,w} = \mathbf{E}\left[\sum_{v,w\neq 0} [v \in V][w \in V]\right] = \binom{2^d-1}{2}.$$

The value of $p_{v,w}$ is then as claimed. We conclude by estimating

$$\frac{p_{v,w}}{p_v p_w} = \frac{\binom{2^d-1}{2}}{\binom{2^n-1}{2}} \cdot \frac{1}{p_v p_w} = \frac{2^d-2}{2^d-1} \cdot \frac{2^n-1}{2^n-2} < 1. \qquad \square$$

It can now be shown that a random subspace of high dimension will hit a large set w.h.p.:

LEMMA 6.4. *Let $\varepsilon < \frac{1}{2}$ be a positive real number and consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 2^{-(\frac{1}{2}-\varepsilon)n}$. Pick V to be a random linear subspace of $\{0,1\}^n$ of dimension d, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$. Then*

$$\Pr_V\left[\frac{|B \cap V|}{|V|} \in (1 \pm 2^{-\frac{\varepsilon}{4}n}) \cdot \beta\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

PROOF. Let $b_1, \ldots, b_N$ be the elements of B and define the random variables $X_i = [b_i \in V]$ and $X = |B \cap V| = \sum_i X_i$. The $\mathbf{E}[X_i]$ were computed in the proof of Lemma 6.3, which gives us

$$\mu = \mathbf{E}[X] = \sum_i \mathbf{E}[X_i] = \begin{cases} \beta 2^n \frac{2^d-1}{2^n-1} & \text{if } \bar{0} \notin B \\ \beta 2^n \frac{2^d-1}{2^n-1} + (1 - \frac{2^d-1}{2^n-1}) & \text{otherwise.} \end{cases}$$

Let's look at the case where $\bar{0} \notin B$. We can estimate μ as follows:

$$\begin{aligned} \mu &= \left(1 + \frac{1}{2^n-1}\right)(1-2^{-d})\beta|V| \\ &\in (1 \pm 2^{-(\frac{1}{2}-\frac{\varepsilon}{2})n})^2 \beta|V| \\ &\subseteq \left(1 \pm \frac{1}{3} \cdot 2^{-\frac{\varepsilon}{2}n}\right)\beta|V|. \end{aligned}$$

When $\bar{0} \in B$ we still have $\mu \in (1 \pm 2^{-\frac{\varepsilon}{2}n})\beta|V|$, because $1 - \frac{2^d-1}{2^n-1} \leq 1 \ll \frac{1}{3} \cdot 2^{-\frac{\varepsilon}{2}n}\beta|V|$. So this holds in both cases.

Lemma 6.3 also says that $\mathbf{E}[X_iX_j] \leq \mathbf{E}[X_i]\mathbf{E}[X_j]$ for all $i \neq j$. And so by the second-moment method (Lemma 6.2):

$$\Pr\left[X \in \mu\,(1 \pm \delta)\right] \geq 1 - \frac{1}{\delta^2\mu},$$

which means

$$\Pr\left[X \in (1 \pm 2^{-\frac{\varepsilon}{2}n})(1 \pm \delta)\beta|V|\right] \geq 1 - \frac{1}{\delta^2 \cdot \beta \cdot 2^d \cdot (1 - 2^{-\frac{\varepsilon}{2}n})}.$$

Taking $\delta = \frac{1}{3}2^{-\frac{\varepsilon}{4}n}$, we get

$$\begin{aligned}\Pr\left[X \in (1 \pm 2^{-\frac{\varepsilon}{4}n})\beta|V|\right] &\geq 1 - \frac{9}{2^{-\frac{\varepsilon}{2}n} \cdot 2^{-(\frac{1}{2}-\varepsilon)n} \cdot 64 \cdot 2^{(\frac{1}{2}-\frac{\varepsilon}{4})n}} \\ &\geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.\end{aligned}$$

□

We will show a similar result when we pick the set V in the following manner: First we pick a uniformly random odd Hamming weight vector $a \in \{0,1\}^n$, and then, we pick W to be a random subspace of dimension d within $a^\perp$, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$; then $V = a + W$.

Lemma 6.5. *Consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 2^{-(\frac{1}{2}-\varepsilon)n}$. Pick V as described above. Then*

$$\Pr_V\left[\frac{|B \cap V|}{|V|} \in \beta(1 \pm 2^{-\frac{\varepsilon}{4}n})\right] \geq 1 - 2^{-\frac{\varepsilon}{4}n}.$$

Proof. Let $B' = (-a+B) \cap a^\perp$ where $-a+B$ denotes the affine subspace which is obtained by adding the vector $-a$ to all vectors in B and let $\beta' = \frac{|B'|}{|a^\perp|}$. A vector $a \in \{0,1\}^n$ is called *good* when

$$\beta' \stackrel{\text{def}}{=} \frac{|(-a+B) \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}).$$

We will later show that if a is a uniformly random string of odd Hamming weight, then

$$(*) \qquad \Pr_a\left[a \text{ is good}\right] \geq 1 - \frac{2}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

For every good a, Lemma 6.4 gives us:

$$\Pr_W\left[\frac{|B' \cap W|}{|W|} \in \beta'(1 \pm 2^{-\frac{\varepsilon}{4}n}) \;\middle|\; a\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Our result then follows by Bayes' rule.

To prove (∗), suppose that a is chosen to be a uniformly random nonzero string (i.e., with either even or odd Hamming weight). Then $a^\perp$ is a uniformly random subspace of dimension $n - 1 \gg (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$. Hence by Lemma 6.4,

$$(**) \qquad \Pr_a\left[\frac{|B \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n})\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Now $|a^\perp| = 2^{n-1}$, so if $a^\parallel$ denotes the complement of $a^\perp$ (in $\{0,1\}^n$), then $|a^\parallel| = 2^{n-1}$ also, and

$$\frac{|B \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}) \iff |B \cap a^\perp| \in \frac{1}{2}|B| \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n})$$
$$\iff \frac{|B \cap a^\parallel|}{|a^\parallel|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}).$$

So (∗∗) also holds with respect to the rightmost (equivalent) event. Since a uniformly random nonzero a has odd Hamming weight with probability $> \frac{1}{2}$, it must then follow that if we pick a uniformly random a with odd Hamming weight, then:

$$\Pr_a\left[\frac{|B \cap a^\parallel|}{|a^\parallel|} \in \beta \cdot (1 \pm 2^{-n/20})\right] \geq 1 - \frac{2}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Now notice that $|a^\parallel| = |a^\perp|$ and that for odd Hamming weight a, $B \cap a^\parallel = (-a + B) \cap a^\perp$; this establishes (∗). □

The lemmas above are the key to constructing rectangle distributions for IP.

Proof (Proof of Lemma 6.1). The rectangles produced in σ_0 and σ_1 are monochromatic as required. Also, V and $V^\perp$ of σ_0 are both

random subspaces of dimension $\geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$ — as required by Lemma 6.4 — and V and $V^{\|}$ of σ_1 are both obtained by the kind of procedure required in Lemma 6.5. It then follows by a union bound that if R is chosen by either σ_0 or σ_1 that, if A, B are subsets of $\{0,1\}^n$ of densities $\alpha, \beta \geq 2^{-(\frac{1}{2}-\varepsilon)n}$, then

$$\Pr_R \left[\frac{|A \times B \cap R|}{|R|} = (1 \pm 9 \cdot 2^{-\frac{\varepsilon}{4}n}) \cdot \alpha\beta \right] \geq 1 - 2 \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Hence, the same probability lower-bounds the event that $A \times B \cap R \neq \varnothing$. □

7. Conclusion and follow-up work

We have shown deterministic simulation theorems for two choices of the inner function g for which such theorems were hitherto unknown. The input size for our chosen gadgets is exponentially smaller than for the indexing function.

A recent follow-up paper of Alexander Kozachinskiy (Kozachinskiy 2018) also makes use of our technique, by proving that certain gadgets constructed from expander graphs have monochromatic rectangle distributions with good hitting parameters. In particular, he constructs such distributions for the gadget $\mathsf{SQR}^q(a, b)$ which decides whether the difference $a-b$, of two elements of the field $\mathbb{F}_{q^2}$, is a perfect square. Simulation theorems then follow from our result.

Kozachinskiy also shows that our thickness lemma (Lemma 3.9) cannot be improved. This lemma is the bottleneck which prevents the technique from working with even smaller gadgets, and Kozachinskiy's result suggests that any further improvement in the gadget size of deterministic simulation theorems may well require a new approach.

After our paper, a randomized simulation theorem was proven by Göös, Pitassi & Watson (2017b) which uses the indexing function as a gadget. A follow-up work by Loff & Mukhopadhyay (2019) shows a deterministic simulation theorem for the equality gadget, using techniques similar to our own. Very recently, a randomized simulation theorem was proven for the inner-product function by Chattopadhyay, Filmus, Koroth, Meir & Pitassi (2019).

The most important open problem in this topic is currently to prove a simulation theorem for a constant-size gadget. This would lead to significant improvements to known lower bounds on monotone circuits, propositional proof systems, and possibly more.

Acknowledgements

Part of the research for this work was done at the Institut Henri Poincaré, as part of the workshop *Nexus of Information and Computation Theories*.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 616787. The first author was partially supported by a Ramanujan Fellowship of the DST, India, and the last author is partially supported by a TCS fellowship and by European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme under grant agreement No. 715672.

The research leading to these results has also received funding from the Foundation for Science and Technology (FCT), Portugal, grant number SFRH/BPD/116010/2016. This work is partially funded by the ERDF through the COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT as part of project UID/EEA/50014/2013.

We are thankful to the anonymous referees for their thorough reading and numerous suggestions for improving the paper.

References

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

Paul Beame, Trinh Huynh & Toniann Pitassi (2010). Hardness amplification in proof complexity. In *Proceedings of the 42nd STOC*, 87–96.

Paul Beame, Toniann Pitassi, Nathan Segerlind & Avi Wigderson (2005). A direct sum theorem for corruption and the multiparty NOF communication complexity of set disjointness. In *Proceedings of the 20th CCC*, 52–66.

Maria Luisa Bonet, Juan Luis Esteban, Nicola Galesi & Jan Johannsen (2000). On the Relative Complexity of Resolution Refinements and Cutting Planes Proof Systems. *SIAM Journal on Computing* **30**(5), 1462–1484. ISSN 0097-5397. URL https://doi.org/10.1137/S0097539799352474.

Mark Braverman & Anup Rao (2014). Information Equals Amortized Communication. *IEEE Transactions on Information Theory* **60**(10), 6058–6069.

Mark Braverman, Anup Rao, Omri Weinstein & Amir Yehudayoff (2013a). Direct Product via Round-Preserving Compression. In *Proceedings of the 40th ICALP*, 232–243.

Mark Braverman, Anup Rao, Omri Weinstein & Amir Yehudayoff (2013b). Direct Products in Communication Complexity. In *Proceedings of the 54th FOCS*, 746–755.

Joshua Brody, Harry Buhrman, Michal Koucký, Bruno Loff, Florian Speelman & Nikolay Vereshchagin (2013). Towards a reverse newman's theorem in interactive information complexity. In *Proceedings of the 28th CCC*, 24–33.

Arkadev Chattopadhyay (2007). Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits. In *Proceedings of the 48th FOCS*, 449–458.

Arkadev Chattopadhyay (2009). *Circuits, Communication and Polynomials*. Ph.D. thesis, McGill University.

Arkadev Chattopadhyay & Anil Ada (2008). Multiparty Communication Complexity of Disjointness. Technical Report TR08-002, Electronic Colloquium on Computational Complexity (ECCC). URL http://eccc.hpi-web.de/eccc-reports/2008/TR08-002/index.html.

Arkadev Chattopadhyay, Pavel Dvorák, Michal Koucký, Bruno Loff & Sagnik Mukhopadhyay (2017a). Lower Bounds for Elimination via Weak Regularity. In *Proceedings of the 34th STACS*, 21:1–21:14.

Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir & Toniann Pitassi (2019). Query-to-communication lifting for BPP using inner product. arXiv:1904.13056.

Arkadev Chattopadhyay, Michal Koucký, Bruno Loff & Sagnik Mukhopadhyay (2017b). Composition and Simulation Theorems via Pseudo-random Properties. *Electronic Colloquium on Computational Complexity (ECCC)* **24**, 14.

Matei David, Toniann Pitassi & Emanuele Viola (2009). Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory* **1**(2).

Andrew Drucker (2012). Improved direct product theorems for randomized query complexity. *Computational Complexity* **21**(2), 197–244.

Peter Frankl & Zoltán Füredi (1981). A short proof for a theorem of Harper about Hamming-spheres. *Discrete Mathematics* **34**(3), 311–313.

Mika Göös, Pritish Kamath, Toniann Pitassi & Thomas Watson (2017a). Query-to-communication Lifting for P^{NP}. In *Proceedings of the 32nd CCC*.

Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson & David Zuckerman (2015). Rectangles are nonnegative juntas. In *Proceedings of the 47th STOC*, 257–266. ACM.

Mika Göös & Toniann Pitassi (2014). Communication lower bounds via critical block sensitivity. In *Proceedings of the 46th STOC*, 847–856.

Mika Göös, Toniann Pitassi & Thomas Watson (2015). Deterministic communication vs. partition number. In *Proceedings of the 56th FOCS*.

Mika Göös, Toniann Pitassi & Thomas Watson (2017b). Query-to-Communication Lifting for BPP. In *Proceedings of the 58th FOCS*.

L.H. Harper (1966). Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory* **1**(3), 385 – 393.

Prahladh Harsha, Rahul Jain, David McAllester & Jaikumar Radhakrishnan (2007). The communication complexity of correlation. In *Proceedings of the 22nd CCC*, 10–23.

Hamed Hatami, Kaave Hosseini & Shachar Lovett (2018). Structure of Protocols for XOR Functions. *SIAM J. Comput.* **47**(1), 208–217. URL https://doi.org/10.1137/17M1136869.

Trinh Huynh & Jakob Nordstrom (2012). On the virtue of succinct proofs: Amplifying communication complexity hardness to time-space trade-offs in proof complexity. In *Proceedings of the 44th STOC*, 233–248.

Russell Impagliazzo (1995). Hard-Core Distributions for Somewhat Hard Problems. In *Proceedings of the 36th FOCS*, 538–545.

Rahul Jain (2015). New strong direct product results in communication complexity. *Journal of the ACM* **62**(3), 20.

Rahul Jain, Hartmut Klauck & Ashwin Nayak (2008). Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the 40th STOC*, 599–608.

Rahul Jain, Attila Pereszlényi & Penghui Yao (2012). A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53rd FOCS*, 167–176.

Rahul Jain, Jaikumar Radhakrishnan & Pranab Sen (2003). A direct sum theorem in communication complexity via message compression. In *Proceedings of the 20th ICALP*, 300–315.

Rahul Jain & Penghui Yao (2012). A strong direct product theorem in terms of the smooth rectangle bound. Technical report, arXiv:1209.0263.

Jan Johannsen (2001). Depth Lower Bounds for Monotone Semi-Unbounded Fan-in Circuits. *ITA* **35**(3), 277–286. URL https://doi.org/10.1051/ita:2001120.

Mauricio Karchmer & Avi Wigderson (1990). Monotone Circuits for Connectivity Require Super-Logarithmic Depth. *SIAM Journal on Discrete Mathematics* **3**(2), 255–265.

Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland & David Xiao (2015). Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing* **44**(5), 1550–1572.

Alexander Kozachinskiy (2018). Raz-McKenzie simulation: new gadget and unimprovability of Thickness Lemma. In *Proceedings of the 43rd MFCS*.

Eyal Kushilevitz & Noam Nisan (1997). *Communication complexity*. Cambridge University Press. ISBN 978-0-521-56067-2.

Troy Lee, Adi Shraibman & Robert Spalek (2008). A Direct Product Theorem for Discrepancy. In *Proceedings of the 23rd CCC*, 71–80.

Troy Lee & Shengyu Zhang (2010). Composition theorems in communication complexity. In *Proceedings of the 27th ICALP*, 475–489. Springer.

Leonid A. Levin (1987). One-way functions and pseudorandom generators. *Combinatorica* **7**(4), 357–363.

Bruno Loff & Sagnik Mukhopadhyay (2019). Lifting Theorems for Equality. In *Proceedings of the 36th STACS*, 50:1–50:19. URL https://doi.org/10.4230/LIPIcs.STACS.2019.50.

Raghu Meka & Toniann Pitassi (editors) (2017). *Hardness Escalation in Communication Complexity and Query Complexity, Workshop at 58th FOCS*. URL https://raghumeka.github.io/workshop.html.

Jakob Nordström (2016). Private communication.

Denis Pankratov (2012). *Direct sum questions in classical communication complexity*. Ph.D. thesis, Masters thesis, University of Chicago.

Anup Rao & Amir Yehudayoff (2015). Simplified Lower Bounds on the Multiparty Communication Complexity of Disjointness. In *Proceedings of the 30th CCC*, 88–101.

Ran Raz & Pierre McKenzie (1999). Separation of the Monotone NC Hierarchy. *Combinatorica* **19**(3), 403–435.

Susanna F. de Rezende, Jakob Nordström & Marc Vinyals (2016). How Limited Interaction Hinders Real Communication. In *Proceedings of the 56th FOCS*.

Robert Robere, Toniann Pitassi, Benjamin Rossman & Stephen A Cook (2016). Exponential lower bounds for monotone span programs. In *Proceedings of the 57th FOCS*, 406–415.

Ronen Shaltiel (2003). Towards proving strong direct product theorems. *Computational Complexity* **12**(1–2), 1–22.

Alexander A. Sherstov (2009). Separating AC0 from Depth-2 Majority Circuits. *SIAM Journal on Computing* **38**(6), 2113–2129. URL http://dx.doi.org/10.1137/08071421X.

Alexander A Sherstov (2011). The pattern matrix method. *SIAM Journal on Computing* **40**(6), 1969–2000.

Alexander A. Sherstov (2012a). The multiparty communication complexity of set disjointness. In *Proceedings of the 44th STOC*, 525–548.

Alexander A Sherstov (2012b). Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing* **41**(5), 1122–1165.

Alexander A. Sherstov (2013). Communication lower bounds using directional derivatives. In *Proceedings of the 45th STOC*, 921–930.

Yaoyun Shi & Yufan Zhu (2009). Quantum communication complexity of block-composed functions. *Quantum Information & Computation* **9**(5), 444–460.

Dmitry Sokolov (2017). Dag-like communication and its applications. In *International Computer Science Symposium in Russia*, 294–307.

Emanuele Viola & Avi Wigderson (2008). Norms, XOR Lemmas, and Lower Bounds for Polynomials and Protocols. *Theory of Computing* **4**(1), 137–168.

Thomas Watson (2017). A $\mathsf{ZPP}^{\mathsf{NP}}$ Lifting Theorem. *Unpublished preprint.*

Xiaodi Wu, Penghui Yao & Henry Yuen (2017). Raz-McKenzie simulation with the inner product gadget. Technical Report TR17-010, Electronic Colloquium on Computational Complexity (ECCC). URL https://eccc.weizmann.ac.il/report/2017/010/.

Mihalis Yannakakis (1991). Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences* **43**(3), 441–466.

Andrew Chi-Chih Yao (1979). Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *Proceedings of the 11h STOC*, 209–213.

Andrew Chi-Chih Yao (1982). Theory and Applications of Trapdoor Functions (Extended Abstract). In *Proceedings of the 23rd FOCS*, 80–91.

Manuscript received 8 July, 2018

Arkadev Chattopadhyay
TIFR, Mumbai, India
arkadev.c@tifr.res.in

Michal Koucký
Charles University, Prague,
Czech Republic
koucky@iuuk.mff.cuni.cz

Bruno Loff
INESC-TEC & U. Porto,
Porto, Portugal
bruno.loff@gmail.com

Sagnik Mukhopadhyay
KTH, Stockholm, Sweden
sagnik@kth.se