Electronic Colloquium on Computational Complexity, Report No. 14 (2017)

Composition and Simulation Theorems via Pseudo-random Properties

Arkadev Chattopadhyay [1] Michal Koucký[2] Bruno Loff[3] Sagnik Mukhopadhyay [4]

Abstract

We prove a randomized communication-complexity lower bound for a composed function — $\mathsf{OrderedSearch} \circ \mathsf{IP}$ — by lifting the randomized query-complexity lower-bound of $\mathsf{OrderedSearch}$ to the communication-complexity setting. We do this by extending ideas from a paper of Raz and Wigderson [RW89]. We think that the techniques we develop will be useful in proving a randomized simulation theorem.

We also generalize the deterministic simulation theorem of Raz and McKenzie [RM99], to any inner-function which satisfies certain hitting property. We prove that IP satisfies this property, and as a corollary we obtain deterministic simulation theorem for an inner-function gadget with logarithmic block-size with respect to the arity of the outer function. This answers an open question posed by Göös, Pitassi and Watson [GPW15]. Our result also implies the previous results for the Indexing inner-function.

[1]Tata Institute of Fundamental Research, Mumbai, arkadev.c@tifr.res.in
[2]Charles University, Prague, koucky@iuuk.mff.cuni.cz
[3]Charles University, Prague, bruno.loff@gmail.com
[4]Tata Institute of Fundamental Research, Mumbai, sagnik@tifr.res.in

ISSN 1433-8092

Contents

1 Introduction

A very basic problem in computational complexity is to understand the *complexity* of a composed function in terms of the complexities of the two simpler 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 $f \circ g^p : \{0,1\}^{mp} \to \mathcal{Z}$. The special case of $\mathcal{Z}$ being $\{0,1\}$ and f the XOR function has been the focus of several works [Yao82, Lev87, Imp95, Sha03, LSS08, VW08, She12b], 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 [JRS03, BPSW05, HJMR07, JKN08, Dru12, Pan12, JPY12, JY12, BBCR13, BRWY13a, BRWY13b, BBK$^+$13, BR14, KLL$^+$15, Jai15]. Making progress on even these special cases of the general problem in various models of computation are outstandingly open.

While no such general theorems are known, there has been some progress in communication complexity setting. In this setting the input for g is split between two parties, Alice and Bob. A particular case of progress from a few years ago is the development of the pattern matrix method by Sherstov [She11] and the closely related block-composition method of Shi and Zhu [SZ09], which led to a series of interesting developments [Cha07, LSS08, CA08, She12a, She13, RY15] resolving several open problems. In both these methods, the relevant analytic property of the outer function is approximate degree. While the pattern-matrix method entailed the use of a special inner function, the block-composition method, further developed by Chattopadhyay [Cha09], Lee and Zhang [LZ10] and Sherstov [She12a, She13], 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 discrepancy of g.

The following simple protocol is suggestive: Alice and Bob try to solve f using a decision tree (randomized/deterministic) 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 $\mathcal{M}^{cc}\big(f \circ g^p\big) = O(\mathcal{M}^{dt}\big(f\big) \cdot \mathcal{M}^{cc}\big(g\big))$, where $\mathcal{M}$ could be the deterministic or randomized model and $\mathcal{M}^{dt}$ denotes the decision tree complexity. A natural question is if 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.

In a remarkable work, Raz and McKenzie [RM99] showed that this naïve upper bound is always optimal for *deterministic protocols*, when g is the Indexing function (IND), provided the *gadget size is polynomially large* in p. This theorem was the main technical workhorse of Raz and McKenzie to famously separate the monotone NC hierarchy. The work of Raz and McKenzie was recently simplified and built upon by Göös, Pitassi and Watson [GPW15] to solve a longstanding open problem in communication complexity. In line with [GPW15], we call such theorems *simulation theorems*, because they explicitly construct a decision-tree for f by simulating a given protocol for $f \circ g^p$. More recently, de Rezende, Nordström and Vinyals [dRNV16] port 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.

In another recent work, Göös et.al. [GLM$^+$15] proved, using different techniques, that whenever g satisfies a certain 2-source extractor property, which the Inner-product function (IP) does, simulation theorems in other models of communication (e.g. non-deterministic) can be proven. This also has found several applications (see [GJ16, ABB$^+$16] for example). Despite this progress, no simulation theorem is known for the fundamental model of bounded-error randomized communication complexity. While we do not attain this goal here, we make interesting progress by developing new techniques and identifying some key natural properties of the inner function g that we believe should enable proving such randomized simulation theorems in the future.

To make progress, we let $f = \mathsf{OS}_p$ be a natural partial outer-function that we call 'ordered search', and which is defined as follows: consider p-bit inputs that are promised to be of the form 1^i0^{p-i} for some $1 \le i < p$; then $\mathsf{OS}_p(1^i0^{p-i}) = i$. It is not hard to show that a decision tree implementing binary search, of cost $\log p$, is optimal even in the randomized case. Hence $\mathsf{OS}_p \circ g^p$ may also be solved by simulating binary search — but it is not at all clear if this is the best possible strategy for a communication protocol. In particular, the approximate-degree-based composition-theorems can be currently made to yield only a lower bound of $\Omega(n \cdot \sqrt{\log p})$ on the randomized communication complexity. This uses the fact that OS_p is known to have approximate degree $\Omega\big(\sqrt{\log p}\big)$ [Buh16] and IP_n (the Inner-product on n bits) has discrepancy $2^{-\Omega(n)}$ [CG88]. The techniques of Göös et.al. [GLM+15] also do not seem to give any non-trivial bound for the following reason: all decision-tree models considered in [GLM+15] are at least as powerful as non-deterministic decision trees with unique witnesses. It is simple to verify that the query-complexity of OS for such non-deterministic decision-trees is $O(1)$, preventing the application of main result in [GLM+15] to get a tight bound for $\mathsf{OS}_p \circ \mathsf{IP}_n^p$.

Exploiting the fact that small rectangular discrepancy implies a certain equi-distribution property that we call *regularity*, and a certain other structural property of IP which we will describe later, our main result shows the following:

Theorem 1.1. The bounded-error randomized communication complexity of $\mathsf{OS}_p \circ \mathsf{IP}_n^p$ is $\Theta\left(n \log p\right)$, when $n = \Omega(\log p)$.

The proof of this theorem heavily builds upon a relatively less known work of Raz and Wigderson [RW89] which proved lower bounds on the randomized communication complexity of a certain communication game that is derived from the monotone Karchmer-Wigderson game corresponding to s-t connectivity problem in graphs. As we will provide more details later, our proof uses properties of IP to directly "lift" the simple proof of the randomized decision-tree complexity of Ordered Search to the more sophisticated model of 2-party communication. This is one of the main reasons for us to believe that the proof technique and the concerned properties of IP are likely to enable proving a randomized simulation theorem. This intuition seems to be further corroborated by the following fact: we show that a natural weakening of the properties of IP used to prove Theorem 1.1 is sufficient to yield the following deterministic simulation theorem.

Theorem 1.2. Let $p \leq 2^{n/10}$ and $f : \{0,1\}^p \to \mathcal{Z}$, where $\mathcal{Z}$ is any domain. Then,

$$\mathcal{D}^{cc}\big(f_p \circ \mathrm{IP}_n^p\big) = \Theta\bigg(\mathcal{D}^{dt}\big(f\big) \cdot n\bigg).$$

This is the first deterministic simulation theorem with logarithmic gadget size, while that in Raz-McKenzie needed a polynomial size gadget. This answers a problem raised by both Göös-Pittasi-Watson [GPW15] and Göös et.al. [GLM+15] of proving a Raz-McKenzie style deterministic simulation theorem for a different inner function than Indexing with a better gadget size. Moreover, it is not hard to verify that an IP instance easily embeds in Indexing by exponentially blowing up the size. This enables us to also re-derive the original Raz-McKenzie simulation theorem for the Indexing function, even attaining significantly better parameters. That answers a question posed to us recently by Jakob Nordström [Nor16].

One aspect of the previous work of [GLM+15] is that they consider a protocol of cost C as a partition of the universe into at most 2^C rectangles, and a decision tree of height C as a partition of the universe into at most 2^C sub-cubes. But deterministic (and randomized) protocols and decision trees induce very special partitions. To extract a more restricted object like a deterministic (or randomized) decision tree for f from a protocol for $f \circ g^p$, it thus seems very important to use the special structure of a deterministic (or randomized) communication protocol. This is what we focus on in our work. We discuss the special properties of IP that allow us to do this in the follow-up sections.

1.1 Our techniques

1.1.1 Deterministic simulation theorem from a hitting property

It will be convenient for us to begin the discussion in the deterministic setting, because it is technically simpler. Here the main tool for us is to use the general framework of the Raz-McKenzie theorem as used by Göös-Pittasi-Watson [GPW15]. On input $z \in \{0,1\}^p$ for f we will simulate (in our head) the communication protocol for $f \circ g^p$ on inputs that are consistent with queries to z made so far. Namely, we will maintain a rectangle $A \times B \subseteq \{0,1\}^{n\times p} \times \{0,1\}^{n\times p}$ so that for any $(x,y) \in A \times B$, $g^p(x,y)$ is consistent with z on coordinates that were queried. 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 while 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 coordinates not queried, yet. 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 and which guarantees the necessary richness 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. Let us denote by $\mathsf{Ext}^i_A(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p)$ the set of possible extensions x_i such that $\langle x_1,\ldots,x_p\rangle \in A$. We define $\mathsf{Ext}^i_B(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p)$ similarly. If for a given $x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p$ and $y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p$ we know that $\mathsf{Ext}^i_A(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p)$ and $\mathsf{Ext}^i_B(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p)$ are of size at least $2^{(\frac{1}{2}+\epsilon)n}$ then for $g = \mathsf{IP}_n$ we know that there are extensions $x_i \in \mathsf{Ext}^i_A(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p)$ and $y_i \in \mathsf{Ext}^i_B(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p)$ such that $\mathsf{IP}_n(x_i,y_i) = z_i$. Hence, we say that $A \times B$ is τ-thick if $\mathsf{Ext}^i_A(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p)$ and $\mathsf{Ext}^i_B(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p)$ are of size at least $\tau \cdot 2^n$ (or empty), for every choice of i and $x_1,\ldots,x_p,y_1,\ldots,y_p$.

So if we can maintain the thickness of $A \times B$, we maintain the necessary richness of $A \times B$. It turns out that this is indeed possible using the technique of Raz-McKenzie and Göös-Pittasi-Watson. Hence as we progress through the protocol we maintain $A \times B$ to be τ-thick and dense. Once the density of either A or B drops below certain level we are forced to make a query to another coordinate of z. Magically, that restores the density (and thus thickness) of $A \times B$ on coordinates not queried. (An intuitive reason is that if the density of extensions in some coordinate is low then the density in the remaining coordinates must be large.)

We capture the property of IP_n that allows this type of argument to work for other functions g 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 $U \times V \subset \{0,1\}^n \times \{0,1\}^n$ (i.e., $g(u,v) = c$ on every pair $(x,y) \in U \times V$), 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 $|X|/2^n, |Y|/2^n \geq 2^{-h}$,

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

If such distributions σ_0 and σ_1 exist, we say that g has (δ,h)-hitting monochromatic rectangle-distributions. We then prove the following:

Theorem 1.3. If g has (δ,h)-hitting monochromatic rectangle-distributions, $\delta < 1/6$, and $p \leq 2^{\frac{h}{2}}$, then

$$\mathcal{D}^{dt}(f) \leq \frac{5}{h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

We prove this general theorem and then establish that IP over n-bits has $(o(1), \frac{n}{5})$-hitting rectangle-distributions. This immediately yields Theorem 1.2.

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 sub-space $V \subseteq F_2^n$ 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 them 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.

1.1.2 A randomized simulation theorem from a pseudo-random property?

How could we extend our result for deterministic communication complexity and decision trees to randomized communication and randomized decision trees? A natural way to prove the equivalent of Theorem 1.3 in randomized setting would be to use Yao's principle and prove Theorem 1.3 in distributional setting. For that a single input in $A \times B$ consistent with z would not be enough and we would need a more robust property. Interestingly, for IP, the (δ, h)-hitting monochromatic rectangle-distribution property can be strengthened quantitatively in the following way: For our distribution σ_0, if $X \times Y$ is a large enough rectangle ($|X|, |Y| \geq 2^{(\frac{1}{2}+\epsilon)n}$) and we sample $U \times V$ according to σ_0, then the intersection of $X \times Y$ and $U \times V$ has nearly its expected size with high probability. Namely:

$$\Pr_{U\times V\sim\sigma_0}\left[(1-\delta)\frac{|X\times Y|}{2^{2n}} \leq \frac{|(U\times V)\cap(X\times Y)|}{|U\times V|} \leq (1+\delta)\frac{|X\times Y|}{2^{2n}}\right] \geq 1-\delta.$$

This follows by the second moment method. Similarly for σ_1.

At this point, it is natural to ask whether this stronger property suffixes to show a randomized simulation theorem. The point being that we will be able to preserve many strings in $A \times B$ consistent with z, hence hopefully preserve the distributional success probability of our protocol. Perhaps using some additional properties of IP such as its low discrepancy could lead to such a result. We feel that such a proof should be possible, but unfortunately we do not know how to prove it. However we are able to use this stronger property, together with ideas from the work of Raz and Wigderson [RW89], to show a randomized communication-complexity lower bound for $\mathsf{OS} \circ \mathsf{IP}$, that is we solve the case when $f = \mathsf{OS}$.

1.1.3 A decision-tree lower-bound on OS

Our goal is to prove a lower bound on the randomized communication complexity of $\mathsf{OS} \circ \mathsf{IP}$. Let us first look at what is the natural upper bound on this complexity. We can again obtain a protocol for this problem by simulating a decision tree for Ordered Search, and whenever the decision tree queries its (fictitious) input we solve the corresponding instance of IP_n using $n+1$ bits of communication. The Ordered Search on instances of the form 1^i0^{p-i} can be solved by a binary search: query a bit close to the middle of the input, if it is 0 then continue the search on the prefix of the string, otherwise continue with the suffix. The complexity of this decision tree is $O(\log p)$ giving the upper bound $O(n \log p)$ on the communication complexity of $\mathsf{OS} \circ \mathsf{IP}_\mathsf{n}^\mathsf{p}$.

Before going to the argument that the randomized communication complexity of $\mathsf{OS} \circ \mathsf{IP}_n^p$ is $\Omega(n \log p)$, let us first present a brief argument that the randomized decision tree complexity of Ordered Search on inputs from $\mathcal{F}_{1,p}$ is $\Omega(\log p)$ where $\mathcal{F}_{\ell,r} = \{1^i0^{p-i} \mid \ell \leq i \leq r\}$. We fix a uniform distribution $\mu_{\ell,r}$ on $\mathcal{F}_{\ell,r}$. Define $\mathsf{OS}_p : \mathcal{F}_{1,p} \to [p]$ by $\mathsf{OS}_p(1^i0^{p-i}) = i$. We prove that no deterministic decision tree with average success probability $2/3$ over $\mu_{1,p}$ can solve the problem using less than $\frac{1}{100}\log p$ queries. By Yao's principle this implies $\Omega(\log p)$ lower bound on the *randomized* decision tree of OS_p.

Proposition 1.4. The randomized query-complexity of OS_p is $\Omega(\log p)$.

We will use the following lemma — which will be proven in the context of communication complexity in a later section.

Lemma 1.5. Let $\mathcal{T}$ be a deterministic decision-tree which, when given input z drawn from $\mu_{1,p}$, outputs $\mathsf{OS}_p(z)$ with probability γ. For a natural number $1 \le p_1 \le p$, let $p_2 = p - p_1$, let γ_1 be the success probability of $\mathcal{T}$ over μ_{1,p_1}, and let γ_2 be the success probability of $\mathcal{T}$ over $\mu_{p_1+1,p}$. Then there exists a choice of p_1 such that

1. $p_1, p_2 \ge p/200$, and,
2. $\gamma_1, \gamma_2 \ge \gamma/2$

Now suppose we have a deterministic decision-tree $\mathcal{T}$ which, when given input z drawn from $\mu_{1,p}$, outputs $\mathsf{OS}_p(z)$ with probability γ by making no more than t queries. From $\mathcal{T}$ we are going to construct another deterministic decision-tree $\mathcal{T}'$ which, when given input z drawn from $\mu_{1,p'}$, outputs $\mathsf{OS}_{p'}(z)$ with probability γ' within $t-1$ queries, and where $p' \ge p/200$ and $\gamma' \ge \gamma/2$.

We may view $\mathcal{T}$ as a binary tree with each node having a coordinate to be queried and having two children — one for each value of the query. Now suppose that $\mathcal{T}$ makes the first query in the j-th coordinate, where $j > p_1$. Let $\mathcal{T}'$ be obtained from $\mathcal{T}$ by contracting every query to any coordinate $> p_1$, by answering 0 to that query. This eliminates the first query, hence the height of $\mathcal{T}'$ is $\le t-1$. It is not hard to see that $\mathcal{T}'$ is a deterministic decision-tree with success probability γ_2 over μ_{1,p_1}. The case when $j \le p_1$ can be proven similarly.

We repeat this height-reduction procedure until we exhaust all the queries. In the end, we have a deterministic decision-tree $\mathcal{T}^*$ which solves OS_{p^*} on the set $\mathcal{F}_{1,p^*}$ with success probability γ^* where $p^* \ge 200^{-t}p$ and $\gamma^* \ge \gamma 2^{-t}$. If we set $\gamma \ge 2/3$ and $t \le \frac{1}{100}\log p$, we get that $\mathcal{T}^*$, solves OS on $p^* \ge p^{9/10}$ coordinates, under the distribution μ_{0,p^*}, with success probability $\gamma^* \ge p^{1/10}$, *without making any query* — which is impossible. Hence any $\mathcal{T}$ solving OS on p coordinates with success probability $2/3$ must have query-complexity $t \ge \frac{1}{100}\log p$.

1.1.4 Overview of the $\mathsf{OS} \circ \mathsf{IP}$ lower-bound

The structure of the lower-bound proof on the randomized communication complexity of $\mathsf{OS} \circ \mathsf{IP}$ resembles the structure of our deterministic simulation lemma for $f \circ \mathsf{IP}_n^p$. We progress through a deterministic protocol for $\mathsf{OS} \circ \mathsf{IP}$ as before, but this time mimicking the binary search procedure implicit in the above randomized decision-tree lower-bound for Ordered Search. However, there will be various technical differences and challenges from the previous argument, and we will not be explicitly constructing a decision tree for Ordered Search.

We will fix a distribution on the inputs $\{0,1\}^{n\times p} \times \{0,1\}^{n\times p}$ which is a *lifted* distribution of $\mu_{1,p}$ from the randomized decision tree lower bound (i.e., our distribution is uniform on pre-images of IP_n^p for z sampled according to $\mu_{1,p}$). We will go through a deterministic protocol for $\mathsf{OS} \circ \mathsf{IP}_n^p$ that has large success probability γ over our distribution, and low cost $\epsilon n \log p$, and we will maintain a rich set of inputs $A \times B$ as before. This time the richness will be controlled by the density of each A and B (instead of thickness), and we must also keep track of the success probability of the protocol within $A \times B$.

After communicating $\epsilon' n$ bits our rectangle $A \times B$ shrinks according to the protocol. We use our *Sub-rectangle lemma* to show that some path in the protocol tree will cause the rectangle not to shrink too much, while simultaneously preserving most of the success probability.

After this shrinking, we will think of the coordinates of Alice's and Bob's inputs as being split into two parts. If we have p coordinates of n bits each, the *prefix* will be a string $\ell \in \{0,1\}^{n\times p_1}$, and the suffix a string $r \in \{0,1\}^{n\times p_2}$, for some $p_1 + p_2 = p$. We will then *zoom-in* on either prefixes or suffixes of the remaining inputs in $A \times B$. This effectively corresponds to querying the other coordinates as would be done in the deterministic lower bound.

Each (say) prefix ℓ on Alice's side can be extended by some number of suffixes r, so that $\ell \times r \in A$. Likewise for Bob. For most prefixes ℓ of Alice we will find an extension $r = r(\ell)$, and for most prefixes ℓ' of Bob we will find an extension $r' = r'(\ell')$, such that every r and r' have inner-product 0 on all p_2 coordinates. On these inputs the Ordered Search function must now output a coordinate in the prefix: this is why we say that we are *zooming-in* on the prefix. The ℓ and ℓ' for which we cannot find a suitable r (resp. r') are simply discarded. We will do this zooming-in in such a way that the density of the surviving prefixes within $\{0,1\}^{n\times p_1}$ is substantially greater than the density of A in $\{0,1\}^{n\times p}$. Once this happens we may communicate another batch of $\epsilon' n$ bits. We also do this in a way such that the success probability of our protocol on the surviving strings is sufficiently preserved.

Achieving simultaneously both objectives of boosting the density and preserving the success probability is substantially harder than in the deterministic case. The main obstacle is that we do not have apriori control over distribution of the protocol error on the inputs. For example, it could be that we have relatively few prefixes with many extensions that carry most of the success probability while having vast majority of prefixes with few extensions with low success probability of the protocol. Fixing a single extension for each prefix would dramatically reduce our success probability. Hence, the process of zooming-in involves an iterative application of our *Amplification lemma*. Depending on the structure of the inputs and the distribution of the success probability each such iteration either boosts the density of prefixes while not loosing much of success probability (and hence achieving our objective) or it increases the success probability while preserving the density. This increase in success probability guarantees that after a limited number of steps we must achieve our objective.

By alternating the application of the Sub-rectangle lemma and the Amplification lemma, we exhaust all the communication of the protocol. We now get a contradiction by having a constant protocol with fairly large success probability, which is successful on a dense set of inputs. This will be a contradiction similar to the case of randomized decision trees.

So how do we find the promised suffix extensions r and r'? We will use our (δ, h)-hitting monochromatic rectangle-distribution σ_0 for IP, to obtain p_2 0-monochromatic rectangles. Then every $r(\ell)$ will come from Alice's side of these rectangles, and every $r'(\ell')$ will come from Bob's side. The hitting property will not be enough, however. We will need to use a more elaborate result, which we call *Extension lemma*, which in addition to finding extensions for most existing prefixes is also able to find extensions which preserve the overall success probability. This latter requirement is highly non-trivial to obtain, it is the main obstacle that needs to be overcome.

Below we briefly describe a property of IP which is used throughout the lower-bound proof, to enforce good behavior of the rectangle $A \times B$ we are keeping track of. Regularity will ensure, for example, that the density of $A \times B$ is approximately equal to its mass under our lifted distribution; It is also the property from which we ultimately derive a contradiction (the non-existence of a zero-communication protocol with sufficient success probability). The regularity property, together with the extension lemma (which is a non-trivial strengthening of the hitting property), are the main driving forces behind the lower-bound, and so the lower-bound will follow for any function g other than IP for which these two properties can be proven.

1.1.5 The regularity property

Let $g : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Any rectangle $A \times B \subset \{0,1\}^{np} \times \{0,1\}^{np}$ is partitioned into sets $O^z_{AB} = (A \times B) \cap (g^p)^{-1}(z)$, one for each $z \in \{0,1\}^p$. If all of these sets are roughly the same size — $(1 \pm \delta)2^{-p}|A \times B|$ — then we say that $A \times B$ is *δ-regular*. We will say that a function $g : \{0,1\}^n \to \{0,1\}$ is δ-regular if $A \times B$ is δ-regular whenever $|A \times B| \geq \delta \cdot 2^{2np}$. We will show in Section 4 that IP_n is $2^{-n/10}$-regular.

In some sense the regularity property generalizes the notion of discrepancy to non-Boolean outputs. This connection is explained in greater detail in [CrK+16], where the regularity property was used by the authors for showing lower-bounds for the *elimination problem* (which is itself also a composition problem in communication complexity, but where f is a relation and not a function. See [CrK+16] for more details). We think it is a very useful and powerful property, which will find further applications.

1.2 Organization

The sections containing more technical exposition of our results are Sections 3 and 3.2 for the deterministic lower bound, and Sections 5 and 5.1 for the randomized lower bound. An interested reader might want to visit these sections directly.

In Section 2 we recall the notion of communication complexity and decision tree complexity — the two complexity measures that we try to connect in the rest of the paper in various settings. We also state a few combinatorial lemmas that will come handy in subsequent sections. The proof of deterministic simulation theorem with IP gadget is given in Section 3. This section is organized in the following way: in Section 3.1 we provide some supporting lemmas for the proof. In Section 3.2 we prove the deterministic simulation theorem for gadget g which has (δ, h)-hitting rectangle distribution and in Section 3.3 we show that IP on n-bits has $(o(1), n/5)$-hitting rectangle distribution.

In Section 4, we introduce the notion of regularity, lifted distributions and quality. In Section 4.4, we provide the proof of $2^{-n/10}$-regularity property of IP on n-bits. In Section 5, we delve into the proof of communication lower-bound of $\mathsf{OS} \circ \mathsf{IP}$. This section is organized as follows: we first provide the main argument of the proof in Section 5.1 assuming two lemmas - *Sub-rectangle* lemma and *Amplification* lemma - which constitute the meat of the proof. These two lemmas are proved subsequently. In Section 5.2, we provide the proof of the Sub-rectangle lemma and the proof of Amplification lemma is provided in Section 5.3. The proof of Amplification lemma is re-factored into following three lemmas – each of which appears in its own subsection. In Section 5.4, we provide the proof of *extension* lemma. The proof of *Zooming-in* lemma appears in Section 5.5 and lastly, Section 5.6 contains the proof of *Min-quality* lemma.

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 $A \times B$. The density of A' in A is $\alpha = |A'|/|A|$.

Consider a product set $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_p$, 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 $I = \{i_1 < i_2 < \cdot < i_k\}$, and $J = [p] \setminus I$. For any $a \in (\{0,1\}^n)^p$, we let $a_I = \langle a_{i_1}, a_{i_2}, \ldots, a_{i_k} \rangle$ 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 a special case $I = [p']$ where $p' \leq p$, we denote A_I as $A_{\leq p'}$. Similarly, for $I = [p] \setminus [p']$, we denote A_I as $A_{>p'}$. 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 $I = [k]$ for some $k \leq p$, 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 $a_{[p]\setminus\{i\}}$, and similarly, $A_{\neq i}$ denotes $A_{[p]\setminus\{i\}}$. 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 $\mathsf{Ext}(a')$.

Notation for intervals and approximation

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

- If δ is a non-negative 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 $0 \notin J$) $\frac{x}{J} = \{\frac{x}{y} \mid y \in J\}$.
- For $x, y \in \mathbb{R}$, we use the notation $x \stackrel{\delta}{\approx} y$ to mean that both $x \in (1 \pm \delta)y$ and $y \in (1 \pm \delta)x$.

The following claim is easy to verify:

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

- (Weak symmetry) If $x \in (1 \pm \delta)y$ then $x \stackrel{2\delta}{\approx} y$ (since $\frac{1}{1\pm\delta} \subseteq 1 \pm 2\delta$).
- (Weak transitivity) If $x \stackrel{\delta}{\approx} y \stackrel{\delta}{\approx} z$, then $x \stackrel{3\delta}{\approx} z$.

Communication complexity

See [KN97] for an excellent exposition on this topic, which we cover here only very briefly. In the two-party communication model introduced by Yao [Yao79], 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.

A *randomized protocol* is a distribution Π over deterministic protocols $\pi : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$. We say that Π computes F with *success probability* γ if for every input (a,b), a random π chosen according to Π will correctly compute F on (a,b) with probability $\geq \gamma$. The communication complexity of Π is the maximum over all π in its support. The *randomized communication complexity of* F *for error* ε, denoted $\mathcal{R}^{cc}_{\varepsilon}(F)$, is the smallest communication complexity of any randomized protocol which correctly computes F with success probability $\geq 1 - \varepsilon$.

Suppose λ is a distribution over $\mathcal{A} \times \mathcal{B}$ and π is a deterministic protocol as above; then the *success probability of* π *on* λ is the probability that it will correctly compute F on inputs (a,b) drawn from λ. We may then define the *distributional communication complexity of* F*, with respect to* λ *and error* ε, denoted $\mathcal{D}^{cc}_{\lambda,\varepsilon}(F)$, to be the smallest communication complexity of any protocol having success probability $\geq 1 - \varepsilon$ on λ. It is then well-known that:

Theorem 2.2 (Yao's principle for communication complexity). $\mathcal{R}^{cc}_{\varepsilon}(F) = \max_{\lambda} \mathcal{D}^{cc}_{\lambda,\varepsilon}(F)$

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 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.

We now define the notion of *randomized* and *distributional* query complexities, in exactly the same way as above. A *randomized decision-tree* is a distribution T over deterministic decision-trees $t : \{0,1\}^p \to \mathcal{Z}$. We say that T computes f with *success probability* γ if for every input z, a random t chosen according to T will correctly compute f on z with probability $\geq \gamma$. The query complexity of T is the maximum query complexity over all t in its support. The *randomized query complexity of f for error ε*, denoted $\mathcal{R}^{dt}_{\varepsilon}(f)$, is the smallest query complexity of any randomized decision-tree which correctly computes f with success probability $\geq 1-\varepsilon$.

Suppose μ is a distribution over $\{0,1\}^p$ and t is a deterministic decision-tree as above; then the *success probability of t on μ* is the probability that it will correctly compute f on inputs z drawn from μ. We may then define the *distributional query complexity of f, with respect to μ and error ε*, denoted $\mathcal{D}^{dt}_{\mu,\varepsilon}(f)$, to be the smallest query complexity of any decision-tree having success probability $\geq 1-\varepsilon$ on μ. It is then well-known that:

Theorem 2.3 (Yao's principle for query complexity)**.** $\mathcal{R}^{dt}_{\varepsilon}(F) = \max_{\mu} \mathcal{D}^{dt}_{\mu,\varepsilon}(F)$

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:

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

For $N = 2^n$, the *Indexing function on N-bits*, IND_N, is defined on $\{0,1\}^{\log N} \times \{0,1\}^N$ to be:

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

Let $\mathcal{F}_{1,p} = \{1^i 0^{p-i} \mid 1 \leq i \leq p\} \subseteq \{0,1\}^p$. The *Ordered Search function on p bits* is defined on $\mathcal{F}_{1,p}$ to be:

$$\mathsf{OS}_p(1^i 0^{p-i}) = i.$$

The second moment method

We will use the well-known second moment method. We use the following variant of Chebyshev's inequality.

Proposition 2.4 (Chebyshev's inequality)**.** 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:

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

Proof. First compute

$$\mathbf{E}[X^2] = \sum_i \mathbf{E}[X_i^2] + 2\sum_{i\neq j} \mathbf{E}[X_i X_j];$$

since $X_i \in [0,1]$, and from the anti-correlation property, this is at most

$$\sum_i \mathbf{E}[X_i] + 2\sum_{i\neq j} \mathbf{E}[X_i]\mathbf{E}[X_j] \leq \mu + \mu^2.$$

From Markov's inequality we now have

$$\Pr[|X-\mu| \geq \varepsilon\mu] = \Pr[(X-\mu)^2 \geq \varepsilon^2\mu^2] \leq \frac{\mathbf{E}[(X-\mu)^2]}{\varepsilon^2\mu^2}.$$

Since $\mathbf{E}[(X-\mu)^2] = \mathbf{E}[X^2 - 2X\mu + \mu^2] = \mathbf{E}[X^2 - \mu^2] \leq \mu$,

$$\Pr[|X-\mu| \geq \varepsilon\mu] \leq \frac{\mu}{\varepsilon^2\mu^2} = \frac{1}{\varepsilon^2\mu}.$$ □

Remark 2.5. In Section 3, we will use Proposition 2.4 where X_i's are independent Bernoulli random variables. In Section 5, however, we will use the full power of Proposition 2.4.

Boosting the density of projections

Let $\mathcal{A} = \mathcal{L} \times \mathcal{R}$ for some finite sets $\mathcal{L}$ and $\mathcal{R}$; if $\ell \in \mathcal{L}$, then denote by $\mathsf{Ext}(\ell)$ the set of $r \in \mathcal{R}$ with $\ell \times r \in A$; if $r \in \mathcal{R}$, then denote by $\mathsf{Ext}(r)$ the set of $\ell \in \mathcal{L}$ with $\ell \times r \in A$.

Proposition 2.6. Suppose $A \subseteq \mathcal{A}$ has density $\alpha = \frac{|A|}{|\mathcal{A}|}$. Consider the two sets

$$A_{\mathrm{L}} = \left\{ \ell \in \mathcal{L} \;\middle|\; \frac{|\mathsf{Ext}(\ell)|}{|\mathcal{R}|} \geq \frac{1}{4}\alpha \right\} \quad \text{and} \quad A_{\mathrm{R}} = \left\{ r \in \mathcal{R} \;\middle|\; \frac{|\mathsf{Ext}(r)|}{|\mathcal{L}|} \geq \frac{1}{4}\alpha \right\}.$$

Then either $\frac{|A_{\mathrm{L}}|}{|\mathcal{L}|} \geq \frac{1}{4}\sqrt{\alpha}$ or $\frac{|A_{\mathrm{R}}|}{|\mathcal{R}|} \geq \frac{1}{4}\sqrt{\alpha}$ (or both).

Proof. Consider a Boolean matrix $M = \mathcal{L} \times \mathcal{R}$ such that $M_{\ell,r} = 1$ iff $\ell \times r \in A$. From the premise, we know that the fraction of 1's in M is $\geq \alpha$. How many 1's can we fit into a matrix M if $|A_{\mathrm{L}}| < \frac{\alpha}{4}|\mathcal{L}|$ and $|A_{\mathrm{R}}| < \frac{\alpha}{4}|\mathcal{R}|$? Clearly $A_{\mathrm{L}} \times A_{\mathrm{R}}$ could well be the all 1-matrix. But in each column of $\mathcal{L} \times (\mathcal{R} \setminus A_{\mathrm{R}})$ we can only fit $\frac{\alpha}{4}|\mathcal{L}|$ many 1's, and in each row of $(\mathcal{L} \setminus A_{\mathrm{L}}) \times \mathcal{R}$ we can fit at most $\frac{\alpha}{4}|\mathcal{R}|$ many 1's. Hence the total number of 1's that we can fit in M is at most:

$$|A_{\mathrm{L}} \times A_{\mathrm{R}}| + 2 \cdot \frac{\alpha}{4} \cdot |\mathcal{L}| \cdot |\mathcal{R}| < \left(\frac{\alpha}{16} + \frac{\alpha}{2}\right) \cdot |\mathcal{L}| \cdot |\mathcal{R}| < \alpha \cdot |\mathcal{L}| \cdot |\mathcal{R}|.$$

Hence, either $|A_{\mathrm{L}}| \geq \frac{\alpha}{4}|\mathcal{L}|$ or $|A_{\mathrm{R}}| \geq \frac{\alpha}{4}|\mathcal{R}|$. $\square$

Weighted average to uniform average

Lemma 2.7 (Weighted average to uniform average)**.** Let $A \subseteq \mathcal{L}\times\mathcal{R}$ be sets, and $\alpha = |A|/(|\mathcal{L}|\cdot|\mathcal{R}|)$ be a real. Suppose that to each $a \in A$ corresponds a non-negative real number $q(a)$, and that

$$\frac{1}{|A|}\sum_{a \in A} q(a) \geq x.$$

Let A_{L} be the projection of A onto $\mathcal{L}$. For $\ell \in A_{\mathrm{L}}$, let $q(\ell) = \frac{1}{|\mathsf{Ext}_A(\ell)|}\sum_{r \in \mathsf{Ext}_A(\ell)} q(\ell r)$.

Then there exists a subset $A' \subseteq A_{\mathrm{L}} \subseteq \mathcal{L}$ with $|A'| \geq \lfloor \alpha \cdot |\mathcal{L}| \rfloor$ and

$$\frac{1}{|A'|}\sum_{\ell \in A'} q(\ell) \geq x.$$

Proof. Set $k = \lfloor \alpha|\mathcal{L}| \rfloor$. Clearly, $|A_{\mathrm{L}}| \geq k$. Let $A_{\mathrm{L}} = \{\ell_1, \ldots, \ell_{|A_{\mathrm{L}}|}\}$ be an ordering of A_{L} by decreasing value of $q(\ell)$. Set $A' = \{\ell_1, \ldots, \ell_k\}$. It remains to show $\sum_{i=1}^{k} q(\ell_i)/k \geq x$. Denote $\mu_i = \frac{|\mathsf{Ext}_A(\ell_i)|}{|\mathcal{R}|}$. We have

$$\sum_{i=1}^{|A_{\mathrm{L}}|} \mu_i = \alpha|\mathcal{L}| \leq k = \sum_{i=1}^{k} 1.$$

It must then hold that

$$\sum_{i=1}^{k} 1 - \mu_i \geq \sum_{i=k+1}^{|A_{\mathrm{L}}|} \mu_i.$$

For any $i \leq k < j$, $q(\ell_i) \geq q(\ell_k) \geq q(\ell_j)$ and $\mu_j, 1 - \mu_i \geq 0$. So

$$\sum_{i=1}^{k} q(\ell_i)(1-\mu_i) \geq \sum_{i=1}^{k} q(\ell_k)(1-\mu_i) \geq \sum_{j=k+1}^{|A_{\mathrm{L}}|} q(\ell_k)\mu_j \geq \sum_{j=k+1}^{|A_{\mathrm{L}}|} q(\ell_j)\mu_j,$$

which simplifies to

$$\sum_{i=1}^{k} q(\ell_i) \geq \sum_{j=1}^{|A_{\mathrm{L}}|} \mu_j q(\ell_j).$$

Since

$$\sum_{j=1}^{|A_{\mathrm{L}}|} \mu_j q(\ell_j) = \frac{1}{|\mathcal{R}|} \cdot \sum_{a \in A} q(a) \geq \frac{|A|}{|\mathcal{R}|} \cdot x = \alpha \cdot |\mathcal{L}| \cdot x \geq k \cdot x$$

we conclude $\sum_{i=1}^{k} q(\ell_i)/k \geq x$ as required. □

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 [RM99], and subsequently simplified in [GPW15], 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 function has 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 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 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. *Let $h \geq 30$ and $1 \leq p \leq 2^{h/2}$ be integers, and $\delta \in (0, 1/16)$ be a real. Let $f : \{0,1\}^p \to \{0,1\}$ and $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be functions. If g has (δ, h)-hitting monochromatic rectangle-distributions then*

$$\mathcal{D}^{dt}(f) \leq \frac{5}{h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

In Section 3.3 we will show that IP_n has $(o(1), \frac{n}{5})$-hitting monochromatic rectangle-distributions, to conclude:

Corollary 3.4. *Let n be large enough integer and $p \leq 2^{n/10}$. For any function $f : \{0,1\}^p \to \{0,1\}$, $\mathcal{D}^{dt}(f) \leq \frac{25}{n} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IP}_n^p)$.*

Jakob Nordström [Nor16] recently posed to us the challenge of proving a simulation theorem for $f \circ \mathsf{IND}_N^p$ (i.e. for Indexing, not Inner-product), with a gadget size N smaller than p^3. We now sketch how our techniques actually give such a result. More careful calculations allow for the following two improvements in the bounds stated in our results above. Fix a constant $\varepsilon > 0$. The following is true.

1. IP_n has $(o(1), n(\frac{1}{2} - \varepsilon))$-hitting monochromatic rectangle-distributions.
2. Theorem 3.3 holds for $p \leq 2^{h(1-\varepsilon)}$ with the conclusion being $\mathcal{D}^{dt}(f) = O(\frac{1}{\varepsilon \cdot h} \cdot \mathcal{D}^{cc}(f \circ g^p))$.

The second improvement requires setting φ to be $4 \cdot 2^{-\varepsilon h}$ in the proof of Theorem 3.3. This allow us to significantly improve the gadget size known for the Indexing function (appearing in [RM99, GPW15]), because of the following reduction: Given an instance $(a, b) \subseteq (\{0,1\}^{np})^2$ of $f \circ \mathsf{IP}_n^p$ where $p \leq 2^{n/10}$, Alice and Bob can construct an instance of $f \circ \mathsf{IND}_N^p$ where $N = 2^n$. Bob converts his input $b \in \{0,1\}^{np}$ to $b' \in \{0,1\}^{Np}$, so that each $b'_i = [\mathsf{IP}_n(x_1, b_i)\rangle, \cdots, \mathsf{IP}_n(x_N, b_i)\rangle]$ where $\{x_1, \cdots, x_N\} = \{0,1\}^n$ is an ordering of all n-bit strings. It is easy to see that $\mathsf{IP}_n(a_i, b_i) = \mathsf{IND}_N(a_i, b'_i)$. Hence it follows as a corollary to our result for IP:

Corollary 3.5. Whenever $N \geq p^{2+\varepsilon}$, $\mathcal{D}^{dt}(f) = O(\frac{1}{\varepsilon \cdot \log N} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IND}_N^p))$.

Also, it is worth noting that the proof of Lemma 7 (projection lemma) in [GPW15] implicitly proves that IND_n has $(o(1), 3\log N/20)$-hitting rectangle-distribution. Hence we can also apply Theorem 3.3 directly to obtain a corollary similar to Corollary 3.5 (albeit with much larger gadget size N).

Notation and definitions

In the rest of Section 3, $n \geq 1$ is an integer and $\mathcal{A} = \mathcal{B} = \{0,1\}^n$. 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 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 $\alpha_I^{i,S} = \frac{|A_I^{i,S}|}{|\mathcal{A}|^{|I|}}$. For a set $B \subseteq \mathcal{B}^p$, we use β and $\beta_I^{i,S}$ for the relevant densities.

Definition 3.6 (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 3.7 (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 $d_{\text{avg}}(A, i) \geq \varphi \cdot |\mathcal{A}|^p$ 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 $|A| \geq \tau \cdot |\mathcal{A}|$.

3.1 Four lemmas exploiting the *thickness* property

The following property is from [GPW15, Lemma 6].

Lemma 3.8 (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** $d_{\min}(A^j, i) < \frac{\delta}{p}\varphi \cdot 2^n$ for some $i \in [p]$ **do**
3: Let a' be a right node of $G(A^j, i)$ with non-zero degree less than $\frac{\delta}{p}\varphi \cdot 2^n$.
4: Set $A^{j+1} = A^j \setminus \{a'\} \times_i \mathsf{Ext}(a')$, 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

$$\sum_{i\in[p]} |A_{\neq i}| = \sum_{i\in[p]} \frac{|A|}{d_{\text{avg}}(A,i)} \leq \frac{p|A|}{\varphi 2^n}.$$

As the algorithm removes at most $\frac{\delta}{p}\varphi \cdot 2^n$ 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. $\square$

Lemma 3.9. 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, $\mathsf{Ext}^{\{j\}}_A(a) \subseteq \mathsf{Ext}^{\{j'\}}_{A^{i,S}_{\neq i}}(a')$, 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 $A^{i,S}_{\neq i}$. $\square$

Lemma 3.10. 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

1. $A^{i,U}_{\neq i}$ and $B^{i,V}_{\neq i}$ is τ-thick,
2. $\alpha^{i,U}_{\neq i} \geq \frac{1}{\varphi}(1-3\delta)\alpha$,
3. $\beta^{i,V}_{\neq i} \geq (1-3\delta)\beta$,

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

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 $U \cap \mathsf{Ext}^{\{i\}}_A(a) \neq \emptyset$, and set it to 0 otherwise.

For each $a \in A_{\neq i}$, $|\mathsf{Ext}^{\{i\}}_A(a)| \geq \tau|\mathcal{A}|$, 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. Otherwise the probability that $M(a, R) = 1$ would be $< 1-\delta$.

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: $|A_{\neq i}^{i,V}| \geq (1-3\delta)|A_{\neq i}|$ and $|B_{\neq i}^{i,V}| \geq (1-3\delta)|B_{\neq i}|$. Since $|B_{\neq i}| \geq |B|/|\mathcal{B}|$, we obtain $|B_{\neq i}^{i,V}|/|\mathcal{B}|^{p-1} \geq (1-3\delta)|B_{\neq i}|/|\mathcal{B}|^{p-1} \geq (1-3\delta)\beta$. 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 $|A_{\neq i}^{i,V}|$ we obtain $|A_{\neq i}^{i,U}|/|\mathcal{A}|^{p-1} \geq (1-3\delta)\alpha/\varphi$. The thickness of $A_{\neq i}^{i,U}$ and $B_{\neq i}^{i,V}$ follows from Lemma 3.9. $\square$

Lemma 3.11. 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 3.9. 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 $(B^{(i)})_{\{i\}} \cap B_i$. Set $A^{(i+1)} = (A^{(i)})_{\neq i}^{i,\{a_i\}}$, $B^{(i+1)} = (B^{(i)})_{\neq i}^{i,\{b_i\}}$, and proceed for the next round. By Lemma 3.9, $A^{(i+1)} \times B^{(i+1)}$ 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 $b = \langle b_1, b_2, \ldots, b_p \rangle$. $\square$

3.2 Proof of the simulation theorem

Now we are ready to present the simulation theorem (Theorem 3.3). Let $h \geq 30$ and $1 \leq p \leq 2^{h/2}$ be integers, and $\delta \in (0, 1/16)$ be a real. Let $f : \{0,1\}^p \to \{0,1\}$ and $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be functions. 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 $\varphi = 4 \cdot 2^{-h/2}$ and $\tau = 2^{-h}$. The decision-tree procedure is presented in Algorithm 2. On an input $z \in \{0,1\}^p$, it uses the protocol Π to decide which bits of z to query.

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.

Algorithm 2 Decision-tree procedure

Input: $z \in \{0,1\}^p$
Output: $f(z)$

1: Set v to be the root of the protocol tree for Π, $I = [p]$, $A = \mathcal{A}^p$ and $B = \mathcal{B}^p$.
2: **while** v is not a leaf **do**
3:     **if** A_I and B_I are both φ-average-thick **then**
4:         Let v_0, v_1 be the children of v.
5:         Choose $i \in \{0,1\}$ for which there is $A' \times B' \subseteq (A \times B) \cap R_{v_i}$ such that
6:           (1) $|A'_I \times B'_I| \geq \frac{1}{4}|A_I \times B_I|$
7:           (2) $A'_I \times B'_I$ is τ-thick.
8:         Update $A = A'$, $B = B'$ and $v = v_i$.
9:     **else if** $d_{\mathrm{avg}}(A_I, j) < \varphi|\mathcal{A}|$ for some $j \in [|I|]$ **then**
10:         Query z_i, where i is the j-th (smallest) element of I.
11:         Let $U \times V$ be a z_i-monochromatic rectangle of g such that
12:           (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
13:           (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\alpha$,
14:           (3) $\beta^{i,V}_{I\setminus\{i\}} \geq (1-3\delta)\beta$,
15:         Update $A = A^{i,U}, B = B^{i,V}$ and $I = I \setminus \{i\}$.
16:     **else if** $d_{\mathrm{avg}}(B_I, j) < \varphi|\mathcal{B}|$ for some $j \in [|I|]$ **then**
17:         Query z_i, where i is the j-th (smallest) element of I.
18:         Let $U \times V$ be a z_i-monochromatic rectangle of g such that
19:           (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
20:           (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq (1-3\delta)\alpha$,
21:           (3) $\beta^{i,V}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\beta$,
22:         Update $A = A^{i,U}, B = B^{i,V}$ and $I = I \setminus \{i\}$.
23: Output $f \circ g^p(A \times B)$.

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 line 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 $A_1 = A \setminus A_0$. We can pick $i \in \{0,1\}$ such that $|(A_i)_I| \geq |A_I|/2$. Set $A'' = A_i$. Since A_I is φ-average-thick, A''_I is $\varphi/2$-average-thick. So using Lemma 3.8 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-image of A'_I obtained from the lemma.) Since $|I| \leq p \leq 2^{h/2}$, the set A'_I will be 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 3.10. 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 $5C/h$ where C is the cost of Π. In the first part of the **while** loop (line 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 $(1-3\delta)/\varphi \geq \frac{1}{2\varphi} \geq 2^{\frac{h}{2}-3}$. Since the density can be at most one, the number of queries is upper bounded by

$$\frac{2C}{\frac{h}{2}-3} = \frac{4C}{h-6} = 4\frac{C}{h} + 24\frac{C}{h(h-6)} \leq 5\frac{C}{h}, \qquad \text{when } h \geq 30.$$

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 $5C/h < p$, I is non-empty. Since $A_I \times B_I$ is τ-thick, it follows from Lemma 3.11 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. $\square$

With greater care the same argument will allow for δ close to $\frac{1}{2}$. We leave the details for the journal version of the paper.

3.3 Hitting monochromatic rectangle-distributions for IP

In this section, we will 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.,

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

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

Lemma 3.12 (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

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

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

Lemma 3.13. Consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 8 \cdot 2^{-n/4}$. Pick V to be a random linear subspace of $\{0,1\}^n$ of dimension $d \geq \frac{7}{15}n$. Then

$$\Pr_V \left[\frac{|B \cap V|}{|V|} \in (1 \pm 2^{-n/20}) \cdot \beta \right] \geq 1 - \frac{1}{2^{n/20}}.$$

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 3.12, which gives us

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

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

$$\mu \ = \ \left(1 + \frac{1}{2^n - 1}\right)(1 - 2^{-d})\beta|V| \ \in \ (1 \pm 2^{-n/5})^2 \beta |V| \ \subseteq \ (1 \pm 2^{-n/6})\beta|V|.$$

We can also show that $\mu \in (1 \pm 2^{-n/6})\beta|V|$ when $\bar{0} \in V$, because $1 - \frac{2^d-1}{2^n-1} \leq 1 \ll 2^{-n/5}\beta|V|$.

Now Lemma 3.12 also says that $\mathbf{E}[X_i X_j] \leq \mathbf{E}[X_i]\mathbf{E}[X_j]$ for all $i \neq j$. And so by the second moment method (Lemma 2.4):

$$\Pr\left[X \in \mu\left(1 \pm \frac{\varepsilon}{2}\right)\right] \geq 1 - \frac{4}{\varepsilon^2 \mu}$$

which means,

$$\Pr\left[X \in (1 \pm 2^{-n/6})(1 \pm \varepsilon/2)\beta|V|\right] \geq 1 - \frac{4}{\varepsilon^2 \beta 2^d (1 - 2^{-n/6})}$$

Taking $\varepsilon \geq 2^{-n/20}$, we get,

$$\Pr\left[X \in (1 \pm 2^{-n/20})\beta|V|\right] \geq 1 - \frac{1}{2 \cdot 2^{n/20}(1 - 2^{-n/6})} \geq 1 - \frac{1}{2^{n/20}}. \qquad \square$$

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 $d \geq 7(n-1)/15$ within $a^\perp$; then $V = a + W$.

Lemma 3.14. Consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 10 \cdot 2^{-n/4}$. Pick V as described above. Then

$$\Pr_V \left[\frac{|B \cap V|}{|V|} \in \beta(1 \pm 3 \cdot 2^{-n/20}) \right] \geq 1 - \frac{3}{2^{n/20}}.$$

Proof. Let $B' = (B - a) \cap a^\perp$ and let $\beta' = \frac{|B'|}{|a^\perp|}$. A string $a \in \{0,1\}^n$ is called *good* when

$$\beta' \stackrel{\text{def}}{=} \frac{|(B-a) \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm \cdot 2^{-n/20}).$$

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

$$\Pr_a [a \text{ is good}] \geq 1 - \frac{2}{2^{n/20}}. \qquad (*)$$

For every good a, Lemma 3.13 gives us:

$$\Pr_W\left[\frac{|B'\cap W|}{|W|}\in\beta'(1\pm 2^{-n/20})\;\middle|\; a\right]\geq 1-\frac{1}{2^{n/20}}.$$

Our result then follows by Bayes' rule.

To prove $(*)$, suppose that a is chosen to be a uniformly random non-zero string (i.e. with either even or odd Hamming weight). Then $a^\perp$ is a uniformly random subspace of dimension $n-1\geq\frac{7}{15}n$. Hence by Lemma 3.13,

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

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

$$\frac{|B\cap a^\perp|}{|a^\perp|}\in\beta\cdot(1\pm 2^{-n/20})\iff|B\cap a^\perp|\in\frac{1}{2}|B|\cdot(1\pm 2^{-n/20})\iff\frac{|B\cap a^\parallel|}{|a^\parallel|}\in\beta\cdot(1\pm 2^{-n/20}).$$

So $(**)$ holds with respect to the rightmost event. Since a uniformly random non-zero 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:

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

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

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

Lemma 3.15. For all n large enough, IP_n has $(6\cdot 2^{-n/20},n/5)$-hitting monochromatic rectangle-distributions.

Proof. 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 subspaces 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 $V\times V^\parallel$, where $V=a+W$ and $V^\parallel=a+W^\perp$.

The rectangles produced above are monochromatic as required. Also, V and $V^\perp$ of σ_0 are both random subspaces of dimension $\geq\frac{7}{15}n$ — as required by Lemma 3.13 — and V and $V^\parallel$ of σ_1 are both obtained by the the kind of procedure required in Lemma 3.14. 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 $\alpha,\beta\geq 2^{-n/5}\gg 10\cdot 2^{-n/4}$, then

$$\Pr_R\left[\frac{|A\times B\cap R|}{|R|}=(1\pm 9\cdot 2^{-n/20})\cdot\alpha\beta\right]\geq 1-\frac{6}{2^{n/20}}.$$

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

4 Regularity

We will now study a property which we believe is fundamental in the understanding of randomized composition problems.

Suppose we have an *outer function* $f : \{0,1\}^p \to \mathcal{Z}$, and an *inner function* $G : \mathcal{A}\times\mathcal{B} \to \{0,1\}^p$, and we wish to study the communication complexity of $f \circ G$. For us, G will typically be IP_n^p, and $\mathcal{A}$ and $\mathcal{B}$ will typically be $\{0,1\}^{np}$ for some n and p; but not always.

We first note that any Sub-rectangle $A \times B$ of $\mathcal{A} \times \mathcal{B}$ can be partitioned by the various inverse images of G; i.e., $A \times B = \bigcup_{z\in\{0,1\}^p} O^z_{AB}$, where

$$O^z_{AB} = O(G, z, A, B) \stackrel{\text{def}}{=} \{(a, b) \in A \times B \mid G(a, b) = z\} = G^{-1}(z) \cap (A \times B).$$

(We will write O^z_{AB} instead of $O(G, z, A, B)$ when G is clear from the context.)

We will say that a rectangle is *regular* if each part in this partition has roughly the same size; we will say that G is *regular* if every large rectangle is regular:

Definition 4.1. Let $0 \le \delta < 1$ and $G : \mathcal{A} \times \mathcal{B} \to \{0,1\}^p$. A Sub-rectangle $A \times B$ of $\mathcal{A} \times \mathcal{B}$ is said to be *δ-regular* (with respect to G), if for every $z \in \{0,1\}^p$

$$|O^z_{AB}| \in (1 \pm \delta) \cdot 2^{-p} \cdot |A \times B|.$$

The function G itself is said to be δ-regular if every Sub-rectangle $A \times B$ of $\mathcal{A} \times \mathcal{B}$ with densities $\frac{|A|}{|\mathcal{A}|} \ge \delta$ and $\frac{|B|}{|\mathcal{B}|} \ge \delta$ is δ-regular.

4.1 Lifted distributions

If we wish to prove a randomized communication complexity lower-bound for $f \circ G$ using Yao's principle, we must produce (constructively or otherwise) a *hard distribution* λ over $\mathcal{A} \times \mathcal{B}$, such that any deterministic protocol will fail to succeed with sufficient probability, when the inputs are drawn from λ.

Now suppose that, in this setting, we have a distribution μ over $\{0,1\}^p$ which we know (or believe) to be hard for f. Then there is a natural way of producing a candidate hard distribution for $f \circ G$. If we denote by O^z the entire inverse image of $z \in \{0,1\}^p$:

$$O^z = O^z_{\mathcal{A}\mathcal{B}} = G^{-1}(z) \cap (\mathcal{A} \times \mathcal{B}),$$

then what we do is distribute $\mu(z)$ probability mass uniformly inside each O^z:

Definition 4.2. Let $G : \mathcal{A} \times \mathcal{B} \to \{0,1\}^p$ and μ be some distribution over $\{0,1\}^p$. Then the *lifting* of μ (to $\mathcal{A} \times \mathcal{B}$, with respect to G) is the distribution $\lambda = \lambda_{\mathcal{A}\times\mathcal{B},G}$ over $\mathcal{A} \times \mathcal{B}$ with probability-mass function:

$$\lambda(a, b) = \lambda_{\mathcal{A}\times\mathcal{B},G}(a, b) \stackrel{\text{def}}{=} \frac{\mu(G(a, b))}{|O^{G(a,b)}|}.$$

Any distribution λ obtained in this way is called a *lifted* distribution. (Again we write λ instead of $\lambda_{\mathcal{A}\times\mathcal{B},G}$ if $\mathcal{A} \times \mathcal{B}$ and G are clear from the context.)

We may now conjecture that if f is hard under μ, in some sense which may depend on the setting, then $f \circ G$ will be hard under λ. We will prove one such result in Section 5.

4.2 Size equals λ-mass for regular rectangles and balanced G

Let us restrict our attention to G which are *balanced* in the following sense:

Definition 4.3. The inner function $G : \mathcal{A} \times \mathcal{B} \to \{0,1\}^p$ is called *balanced* if every inverse image $O^z = G^{-1}(z)$ intersects every *slice* (row and column) of $\mathcal{A} \times \mathcal{B}$ equally, i.e., if for every $a \in \mathcal{A}$, every $b \in \mathcal{B}$ and every $z \in \{0,1\}^p$

$$|\{b' \in \mathcal{B} \mid G(a,b') = z\}| = 2^{-p}|\mathcal{B}| \qquad \text{and} \qquad |\{a' \in \mathcal{A} \mid G(a',b) = z\}| = 2^{-p}|\mathcal{A}|.$$

For now let us observe the following remarkable property: if $A \times B$ is a δ-regular sub-rectangle of $\mathcal{A} \times \mathcal{B}$, with $\alpha = \frac{|A|}{|\mathcal{A}|}$, $\beta = \frac{|B|}{|\mathcal{B}|}$, and λ is a lifted distribution (lifted to $\mathcal{A} \times \mathcal{B}$), with respect to some balanced inner-function G, then

$$\lambda(A \times B) = \sum_z \mu(z) \frac{|O^z_{AB}|}{|O^z|} \in \sum_z \mu(z) \frac{2^{-p} \cdot (1 \pm \delta) \cdot |A \times B|}{2^{-p}|\mathcal{A} \times \mathcal{B}|} = (1 \pm \delta)\alpha\beta.$$

We get:

Proposition 4.4 (Size equals λ-mass for regular rectangles)**.** Let $G : \mathcal{A} \times \mathcal{B} \to \{0,1\}^p$ be a balanced inner-function. If $A \times B$ is a δ-regular sub-rectangle of $\mathcal{A} \times \mathcal{B}$ and λ is a lifted distribution, then

$$\lambda(A \times B) \in (1 \pm \delta) \cdot \alpha\beta.$$

4.3 Success probability and quality

Suppose we are given $f : \{0,1\}^p \to \mathcal{Z}$, $G : \mathcal{A} \times \mathcal{B} \to \{0,1\}^p$ and distribution μ (over $\{0,1\}^p$) as above, but we are now also given a sub-rectangle $A \times B$ of $\mathcal{A} \times \mathcal{B}$, and a deterministic protocol $\pi : A \times B \to \mathcal{Z}$. Then we may define the set of inputs where π correctly computes $f \circ G$:

$$T_{AB} \stackrel{\text{def}}{=} \{(a,b) \in A \times B \mid \pi(a,b) = f \circ G(a,b)\}.$$

We define:

Definition 4.5. The *success probability* of π on $A \times B$ (with respect to f, G and μ) is

$$\gamma \stackrel{\text{def}}{=} \Pr_{(a,b) \sim \lambda}[\pi(a,b) = f \circ G(a,b) \mid (a,b) \in A \times B] = \frac{\lambda(T_{AB})}{\lambda(A \times B)}.$$

If we let $T_{aB} = T_{AB} \cap \{a\} \times B$, we may define the success probability of π on a given string $a \in A$:

$$\gamma_a = \frac{\lambda(T_{aB})}{\lambda(\{a\} \times B)},$$

so that $\gamma = \sum_{a \in A} \lambda(\{a\} \times B \mid A \times B) \cdot \gamma_a$ — i.e. the weighted average of the γ_a is exactly γ. Working with the various γ_a is rather cumbersome, because $\lambda(\{a\} \times B)$ can vary significantly for different a. One can use a trick that appears in [RW89]: instead of measuring the λ-mass of T_{aB} with respect to the λ-mass of $\{a\} \times B$, we will measure the λ-mass of $|T_{aB}|$ with respect to the λ-mass of the entire $a \times \mathcal{B}$.

Definition 4.6. The *row-quality* (with respect to f, G, μ, A, B and π) of $a \in A$ is

$$q(a) = q_{\text{row}}(f, G, \mu, A, B, \pi, a) \stackrel{\text{def}}{=} \frac{\lambda(T_{aB})}{\lambda(\{a\} \times \mathcal{B})}.$$

When f, G, μ, A, B and π are clear from the context, and when it is clear that a denotes an element of A, we will use $q(a)$ instead of $q_{\text{row}}(f, G, \mu, A, B, \pi, a)$, and call this quantity simply the *quality* of a. Note that $q_{\text{row}}(f, G, \mu, A, B, \pi, a)$ equals $q_{\text{row}}(f, G, \mu, A', B, \pi, a)$ for any $A' \subset A$.

Recall again that we write $\alpha = \frac{|A|}{|\mathcal{A}|}$ and $\beta = \frac{|B|}{|\mathcal{B}|}$. Then we may prove the following correspondence:

Lemma 4.7. Let $0 \le \delta < \frac{1}{2}$. If $A \times B$ is δ-regular and G is balanced, then

$$\frac{1}{|A|} \sum_{a \in A} q(a) \in (1 \pm \delta) \cdot \gamma\beta.$$

(Hence if $\delta \le \frac{1}{2}$, then also $\frac{1}{|A|} \sum_{a \in A} q(a) \overset{2\delta}{\approx} \gamma\beta$.)

Proof. If G is balanced, then $\lambda(\{a\} \times \mathcal{B}) = \frac{1}{|\mathcal{A}|}$, and if furthermore $A \times B$ is δ-regular, then by Proposition 4.4 we have $\lambda(A \times B) \in (1 \pm \delta)\alpha\beta$. Then $\frac{1}{|A|} \sum_{a \in A} q(a)$ equals:

$$\frac{1}{|A|} \sum_{a \in A} \frac{\lambda(T_{aB})}{\lambda(\{a\} \times \mathcal{B})} = \sum_{a \in A} \frac{\lambda(T_{aB})}{\lambda(A \times B)} \cdot \frac{\lambda(A \times B) \cdot |\mathcal{A}|}{|A|} = \gamma \cdot \frac{\lambda(A \times B)}{\alpha} \in (1 \pm \delta)\gamma\beta. \qquad \square$$

By symmetry, the same definitions could be stated and the same lemma could be proven with respect to Bob's inputs. We then call it the *column-quality.* We will use the above correspondence several times in the rest of the paper. It tells us that if we have a protocol with good success probability, the average (row- or column-) quality must be high, and if we have several rows (or columns) with high average quality, the protocol must be successful on these rows.

4.4 The regularity property for $G = \mathsf{IP}_n^p$

We begin by recalling the well-known notion of matrix discrepancy [KN97]:

Definition 4.8. Let $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and λ be a distribution over $\mathcal{X} \times \mathcal{Y}$. The *discrepancy* of f under λ equals

$$\mathsf{Disc}_\lambda(f) = \max_{A \subseteq \{0,1\}^n, B \subseteq \{0,1\}^n} \left| \sum_{a \in A, b \in B} \lambda(a,b) \cdot (-1)^{f(a,b)} \right|.$$

It is a well-known fact that the discrepancy of IP_n is at most $2^{-n/2}$ under the uniform distribution over $\{0,1\}^{2n}$ [see KN97, for example]. We will use this to prove regularity with respect to IP_n^p:

Lemma 4.9. Let n be large enough and $p < 2^{n/10}$. Then IP_n^p is $2^{-n/10}$-regular, i.e.: If $A \times B \subseteq (\{0,1\}^{np})^2$, $\alpha = \frac{|A|}{2^{np}} \ge 2^{-n/10}$ and $\beta = \frac{|B|}{2^{np}} \ge 2^{-n/10}$, then $A \times B$ is $2^{-n/10}$-regular with respect to IP_n^p.

Proof. Let $g_j = \mathsf{IP}_n$ for $j < i$ and $g_j = 1 - \mathsf{IP}_n$ for $j \ge i$; then

$$|O_{AB}^z| = \sum_{a,b} \prod_j \frac{1 + (-1)^{g_j(a_j,b_j)}}{2},$$

where the sum is for all $a, b \in A \times B$. Expanding the product and separating out the resulting "1" term:

$$|O_{AB}^z| = 2^{-p} \cdot 2^{2np} \cdot \left(\sum_{a,b} 2^{-2np} + \sum_{\varnothing \ne S \subseteq [p]} \sigma_S \right),$$
$$\sigma_S \triangleq \sum_{a,b} 2^{-2np} \prod_{j \in S} (-1)^{g_j(a_j,b_j)}.$$

The left term is simply $\alpha\beta$, we now bound $|\sigma_S|$. Say $|S| = s$; let a' range over $A_{\bar{S}}$, and a'' over $\mathsf{Ext}(a')$; similarly for b' and b''. Then

$$|\sigma_S| \le \sum_{a',b'} 2^{-2(p-s)n} \left| \sum_{a'',b''} 2^{-2sn} \prod_{j \in S} (-1)^{\mathsf{IP}_n(a''_j,b''_j)} \right| = \sum_{a',b'} 2^{-2(p-s)n} \underbrace{\left| \sum_{a'',b''} 2^{-2sn} (-1)^{\mathsf{IP}_{sn}(a'',b'')} \right|}_{(*)}.$$

Let $D = 2^{-\frac{n}{2}}$; then the known upper-bound on the discrepancy of Inner-product tell us that $(*)$ is upper-bounded by D^s. And then so is the entire sum. But now

$$\sum_{S \neq \varnothing} |\sigma_S| \leq \sum_{s=1}^{p} \binom{p}{s} D^s = (1+D)^p - 1 \leq eDp,$$

where the last inequality holds whenever $pD \leq 1$ (this can be seen by taking the derivative of both sides with respect to D). We conclude that

$$\frac{|O^z_{AB}|}{|A \times B|} \in 2^{-p} \cdot (1 \pm \frac{ep}{\alpha\beta} D) \subseteq 2^{-p} \cdot (1 \pm 2^{-n/2+2+3n/10}) \subseteq 2^{-p} \cdot (1 \pm 2^{-n/10}). \qquad \square$$

At this point it worth mentioning the following remarkable result proven in [LSS08, Theorem 19]:

Lemma 4.10 (XOR-lemma for discrepancy)**.** Let λ^s be the s-fold product of λ, and $\oplus_s g$ be the s-fold XOR of g. Then

$$\mathsf{Disc}_{\lambda^s}(\oplus_s g) \leq 64^s \cdot \mathsf{Disc}_{\lambda}(g)^s$$

By using Lemma 4.10, it is possible to prove an analogue of the regularity property for any function of sufficiently small discrepancy. See [CrK+16] for more details. Generalizing in another direction, it is possible to prove that rectangle with sufficiently high average-thickness (as in Definition 3.7) will also be regular with respect to IP^p_n.

It should be noted here that, comparing our technique to that of [GLM+15], our technique of proving regularity uses the property that IP has small discrepancy under uniform distribution and it does not exploit the *two-source extractor property* of IP , albeit obtaining a seemingly weaker result.

5 Randomized lower-bound for $\mathsf{OS}_p \circ \mathsf{IP}^p_n$

Several lower-bounds are known for composition problems [RM99, She11, LZ10, GLM+15]; for example, the randomized communication complexity of $f \circ \mathsf{IP}^p_n$ is lower-bounded by n times the WAPP-decision-tree complexity of f [GLM+15],[1] and by n times the approximate-degree of f [SZ09, Cha09, LZ10]. However, it is a plausible conjecture that the correct lower-bound is $n \times \mathcal{R}^{dt}(f)$, and this remains an outstanding open problem.

In this section, we will prove a randomized lower-bound of $\Omega(n \log p)$ for the composition problem $\mathsf{OS}_p \circ \mathsf{IP}^p_n$. This result does not follow from either of the lower-bounds mentioned above, because the WAPP-decision-tree complexity of OS is $O(1)$, and a $\Omega(\log p)$ approximate-degree lower-bound for OS is yet unknown[2]. However, it is fairly easy to show a lower-bound of $\mathcal{R}^{dt}(\mathsf{OS}_p) = \Omega(\log p)$ as shown in Section 1.1, and we will see that our communication-complexity lower-bound for $\mathsf{OS}_p \circ \mathsf{IP}^p_n$ will follow the same overall structure of the randomized decision-tree lower-bound. We think that the techniques we develop here will lead to a *randomized* analogue of the simulation theorem of the previous section.

[1] A WAPP-decision-tree algorithm for a family of functions $f : \{0,1\}^p \to \{0,1\}$ is a probabilistic query algorithm which accepts with probability in $[0, \varepsilon\alpha]$ if $f(z) = 0$ and with probability in $[(1-\varepsilon)\alpha, \alpha]$, if $f(z) = 1$, where α is an arbitrary number (that possibly depends on p), and $\varepsilon < \frac{1}{2}$ is some constant. It is analogous to BPP algorithms, where 0 and 1 are replaced by 0 and α. By setting α sufficiently small, such algorithms can be shown to be as powerful as non-deterministic query algorithms that have a unique witness — and such an algorithm can compute OS by guessing the position of the bit-flip.

It should be mentioned that [GLM+15] prove simulation theorems for decision-tree and communication classes other than WAPP, but all of these classes are at least as powerful as non-deterministic unique-witness decision-trees.

[2] Harry Buhrman [Buh16] has provided us with a lower-bound of $\Omega(\sqrt{\log p})$ on this approximate degree. A quantum query-complexity lower-bound of $\Omega(\log p)$ appears in [HNS02] (simplified in [CL08]), but it does not directly imply the same lower-bound for approximate degree (even though a lower-bound for approximate degree would imply the same lower-bound for quantum query algorithms [BBC+01, BDW02]).

Theorem 5.1. There exists a constant c such that, if n and p are sufficiently large natural numbers and $p \leq 2^{\frac{n}{1000}}$, then

$$\frac{n}{c} \cdot \log p \leq \mathcal{R}^{cc}_{2/3}(\mathsf{OS}_p \circ \mathsf{IP}^p_n) \leq (n+1) \cdot \log p.$$

Overview of this section

The upper bound follows easily via binary search. The lower-bound proof is a careful adaptation of the approach of [RW89]. It is akin in spirit to a *round-reduction* proof, but there are actually no rounds, so one could call it a *communication-reduction* proof. In very rough terms, it proceeds as follows: we start with a protocol that solves our problem on a certain rectangle within $\{0,1\}^{np} \times \{0,1\}^{np}$, and we successively obtain a new protocol which either (I) solves the same problem with less communication on smaller rectangle in $\{0,1\}^{np} \times \{0,1\}^{np}$, or (II) solves the same problem on a denser rectangle within $\{0,1\}^{np'} \times \{0,1\}^{np'}$ for a smaller p'. Eventually we obtain a protocol that solves a non-trivial problem with zero communication, and we can prove that such a protocol does not actually exist. This rough description will be fleshed out in Section 5.1, before it is stated and proven in full precision in subsequent sections.

The argument will rely on two lemmas, which we call (I) *Sub-rectangle lemma* and (II) *Amplification lemma*. The Sub-rectangle lemma is proven in Section 5.2, with the help of the regularity property defined in Section 4. The Amplification lemma is then proven in Section 5.3; the proof makes use of a so-called *extension lemma* and some supporting claims. The extension lemma establishes a strong randomized analogue of the hitting rectangle-distribution property of Section 3, and is proven in Section 5.4. The proofs for the supporting claims are provided in Section 5.5 and Section 5.6.

5.1 The main argument

The proof proceeds by alternate applications of aforementioned two lemmas: One lemma says that if we start with a protocol π for solving $\mathsf{OS}_p \circ \mathsf{IP}^p_n$ in a very large rectangle $A \times B$, we can fix a part of the communication and get a protocol π' that solves $\mathsf{OS}_p \circ \mathsf{IP}^p_n$ in a still-somewhat-large rectangle $A' \times B'$, with a similar success probability; we call this the *Sub-rectangle lemma*.

The second lemma says that if we have a protocol π for solving $\mathsf{OS}_p \circ \mathsf{IP}^p_n$ on a somewhat-large rectangle $A \times B$, we can *zoom-in* on one of the sides of the inputs (the first part or the second part of Alice's and Bob's inputs), to obtain a new protocol that solves $\mathsf{OS}_p \circ \mathsf{IP}^{p'}_n$ — so on a smaller number of coordinates p' — and either:

1. Works on a much-denser rectangle (though perhaps loosing a little bit on the success probability), where density is with respect to $\{0,1\}^{np'} \times \{0,1\}^{np'}$; or
2. Works with better success probability (though perhaps loosing a little bit on the density of the rectangle, even when measured on the smaller $\{0,1\}^{np'} \times \{0,1\}^{np'}$.

Putting the two lemmas together will eventually give us that, if we start with a protocol for solving $\mathsf{OS}_p \circ \mathsf{IP}^p_n$ while communicating $\ll n \log p$ bits, we can obtain a zero-communication protocol solving $\mathsf{OS}_p \circ \mathsf{IP}^{p'}_n$ on a large rectangle, with success probability $\gg \frac{1}{p'}$: which we will show is impossible.

The overall structure of the proof is very similar to [RW89]. We have simplified several steps, but could not avoid making it more complicated in other respects.[3] The core new ingredient is the proof of the *extension lemma* for IP (Section 5.4).

5.1.1 The hard distribution λ

The domain of the OS_p function is the set $\mathcal{F}_{1,p} = \{1^i 0^{p-i} \mid i \in [p]\}$; let μ be a uniform distribution on $\mathcal{F}_{1,p}$ — which assigns zero probability to any $z \in \{0,1\}^p \setminus \mathcal{F}_{1,p}$ — and let

[3] Specifically, we need to be able to *zoom-in* on both Alice's and Bob's inputs, which makes the proof somewhat more delicate than [RW89], that only needed to do this for one of the players.

λ be the lifting of μ with respect to $G = \mathsf{IP}_n^p$ (as in Definition 4.2). Suppose we have a rectangle $A \times B \subseteq (\{0,1\}^{np})^2$, and a protocol $\pi : A \times B \to [p]$; define for each $i \in [p]$ the set $O^i_{AB} = \{(a,b) \in A \times B \mid \mathsf{IP}_n^p(a,b) = 1^i 0^{p-i}\}$, and let T^i_{AB} be the subset of O^i_{AB} on which $\pi(a,b) = \mathsf{OS}_p \circ \mathsf{IP}_n^p(a,b)$; let also $O_{AB} = \bigcup_{i\in[p]} O^i_{AB}$ and $T_{AB} = \bigcup_{i\in[p]} T^i_{AB}$. Then the success probability of π on $A \times B$ (with respect to $\mathsf{OS}_p, \mathsf{IP}_n^p$ and μ) is exactly

$$\Pr_{(a,b)\sim\lambda}[\pi(a,b) = \mathsf{OS}_p \circ \mathsf{IP}_n^p(a,b) \mid (a,b) \in A \times B] \stackrel{\text{def}}{=} \frac{\lambda(T_{AB})}{\lambda(A \times B)} = \frac{\frac{1}{p}\sum_i |T^i_{AB}|/|O^i|}{\frac{1}{p}\sum_i |O^i_{AB}|/|O^i|} = \frac{|T_{AB}|}{|O_{AB}|}.$$

5.1.2 Precise statements of the Sub-rectangle and Amplification lemmas

We will need to define the various parameters we want to control.

Definition 5.2 (Existence of protocol on a large rectangle)**.** Let n, p and C be positive integers, and let $\alpha, \beta, \gamma \in (0,1]$. We write

$$\mathrm{Protocol}(n, p, \alpha, \beta, C, \gamma)$$

for the following statement:

- *Large rectangle.* There exists a rectangle $A \times B \subseteq (\{0,1\}^{np})^2$ with $|A| \geq \alpha 2^{np}$, $|B| \geq \beta 2^{np}$;
- *Protocol.* And there exists a protocol $\pi : A \times B \to [p]$ for $\mathsf{OS}_p \circ \mathsf{IP}_n^p$;
- *Success probability.* And the success probability of π on $A \times B$ w.r.t. λ is at least γ:

$$\frac{|T_{AB}|}{|O_{AB}|} \geq \gamma.$$

We can now make precise statements of both lemmas. Below we show how they imply Theorem 5.1.

Lemma 5.3 (Sub-rectangle lemma)**.** Let n, p and C be sufficiently large positive integers, and let $\alpha, \beta, \gamma \in (0,1]$. If $\mathrm{Protocol}(n, p, \alpha, \beta, \gamma, C)$ and $\alpha, \beta \geq 2^{-n/20}$, then

$$\mathrm{Protocol}\left(n, p,\ 2^{-n/10000}\alpha,\ 2^{-n/10000}\beta,\ \gamma - 2 \cdot 2^{-n/10000},\ \max\left(C - \frac{n}{20000}, 0\right)\right).$$

Lemma 5.4 (Amplification lemma)**.** Let n, p and C be sufficiently large positive integers, and let $\alpha, \beta, \gamma \in (0,1]$. Suppose that $\mathrm{Protocol}(n, p, \alpha, \beta, \gamma, C)$ holds, where:

$$\alpha \geq 2 \cdot 2^{-\frac{n}{200}} \qquad \beta \geq 2 \cdot 2^{-\frac{n}{200}}$$
$$p \leq \tfrac{1}{40} 2^{\frac{n}{100}} \qquad \gamma \geq 40 \cdot p^{-1/12}$$

Then one of the following cases will hold for some $\frac{p}{200} \leq p' < p$:

Case 1 — "amplify α". $\mathrm{Protocol}(n, p', \frac{1}{8}\sqrt{\alpha}, \frac{1}{2}\beta, \frac{1}{11}\gamma, C)$.

Case 2 — "amplify γ". $\mathrm{Protocol}(n, p', \frac{1}{2}\alpha, \frac{1}{2}\beta, \frac{11}{10}\gamma, C)$.

The Amplification lemma is symmetric with respect to α and β, the only asymmetry is in the conclusion of Case 1. We will use the lemma also in the case where we reverse the role of α and β, to effectively obtain the conclusion $\mathrm{Protocol}(n, p', \frac{1}{2}\alpha, \frac{1}{8}\sqrt{\beta}, \frac{1}{11}\gamma, C)$ in Case 1.

We will prove the Sub-rectangle lemma in Section 5.2, and the Amplification lemma in Sections 5.3-5.6.

5.1.3 The formal lower-bound proof for $\mathsf{OS} \circ \mathsf{IP}$

Let us now prove Theorem 5.1 assuming the Sub-rectangle and Amplification lemmas are true. We will apply these two lemmas in turn, keeping track of the various parameters $n, p, \alpha, \beta, \gamma$ and C. Let $A_0 \times B_0 = (\{0,1\}^{np_0})^2$, where $p_0 \leq 2^{\frac{n}{1000}}$. Let μ_0 be a uniform distribution on the strings $\{1^i 0^{p_0 - i} \mid i \in [p_0]\}$, and let λ_0 be the lifting of μ_0 to $(\{0,1\}^{np_0})^2$. Suppose we are given a deterministic protocol π_0 to compute $\mathsf{OS}_{p_0} \circ \mathsf{IP}_n^{p_0}$, having success probability $2/3$ over λ_0, and using communication $C_0 \leq \frac{1}{c} \cdot n \log p_0$, where $\frac{1}{c} \in [0,1)$ is a constant to be chosen later.

We will always keep in mind some values $p, \alpha, \beta, \gamma, C$, and some protocol π and rectangle $A \times B$ witnessing Protocol$(n, p, \alpha, \beta, \gamma, C)$. We will be modifying these objects by application of the Sub-rectangle and Amplification lemmas. We begin with $\pi = \pi_0$, $A \times B = A_0 \times B_0$, $p = p_0$, $\alpha_0 = \beta_0 = 1$, $\gamma_0 = 2/3$ and $C = C_0$.

Then, as long as $C > 0$, we repeat the following three steps:

(i) We apply the Sub-rectangle lemma (Lemma 5.3) once.

(ii) We repeatedly apply Amplification lemma (Lemma 5.4) on Alice's side (Protocol$(n, p, \alpha, \beta, \gamma, C)$) until the Case 1 occurs during the application of the lemma at which point α gets amplified and we continue with Step (iii).

(iii) We repeatedly apply the Amplification lemma (Lemma 5.4) on Bob's side (Protocol$(n, p, \beta, \alpha, \gamma, C)$) until the Case 1 occurs during the application of the lemma at which point β gets amplified.

The loop will stop within $\frac{20000}{c} \cdot \log p_0$ iterations, because C decreases by $\frac{n}{20000}$ at Step (i) in each iteration. We will show that following invariants are maintained throughout:

(1) $\gamma \geq \frac{2}{3} \cdot p_0^{-1/25}$ (2) $p_0 \geq p \geq p_0^{1/2}$ (3) $\alpha, \beta \geq 2^{-n/300}$ at the onset of each step.

If these invariants hold then n, p and γ will be large enough to apply the Sub-rectangle and Amplification lemmas. Indeed, p is large enough as $p \geq p_0^{1/2}$ and p_0 is large enough, and $\gamma \geq \frac{2}{3} \cdot p_0^{-1/25} \geq \frac{2}{3} \cdot p^{-2/25} \geq 40 \cdot p^{-1/12}$. We will argue about α and β separately when discussing Invariant (3).

We will use the following constants: $c_1 = 20000/c$, $c_2 = c_1 \cdot \log 128$, and $c_3 = (c_2 / \log(11/10)) + \frac{1}{100}$. We pick constant c large enough so that $2c_1 + c_3 \leq 1/(2 \log 200)$, $c_3 \leq 1/2$ and $c_2 \leq 1/25$.

Invariant (1). Initially, $\gamma = 2/3$. At each iteration of Steps (i)-(iii), γ gets multiplied by a factor $\geq 1/128$: in Step (i) it is multiplied by $1 - o(1) \geq 121/128$ for n large enough, in Step (ii) as long as Case 2 occurs, γ is increasing and then it gets multiplied by a factor $\geq 1/11$, and in Step (iii) it is also multiplied by a factor $\geq 1/11$. Altogether, it gets multiplied by a factor $\geq 1/128$. There are $c_1 \log p_0$ iterations so $\gamma \geq \frac{2}{3} \cdot (\frac{1}{128})^{c_1 \log p_0} \geq \frac{2}{3} \cdot p_0^{-c_2} \geq \frac{2}{3} \cdot p_0^{-1/25}$, provided p_0 is large enough.

Invariant (2). As we have seen in the previous paragraph, γ can decrease by at most factor of p^{-c_2}. Each application of Case 2 of the Amplification lemma increases γ by a factor at least $11/10$. As $\gamma \leq 1$ at all times, the number of times Case 2 occurs can be upper-bounded by $\log_{11/10}(\frac{3}{2} \cdot p^{c_2}) \leq c_3 \log p_0$, for p_0 large enough. Hence, the total number of applications of the Amplification lemma is $\leq (c_3 + 2c_1) \log p_0$. Each application of the lemma shrinks p by a factor of at most $1/200$ so by properties of c_1 and c_3, p can decrease by at most $1/\sqrt{p_0}$.

Invariant (3). Assume $\alpha, \beta \geq 2 \cdot 2^{-n/300}$ before Step (i). Let us focus on α. α decreases by a factor at most $2^{-\frac{n}{10000}}$ in Step (i), thereby having a value at least $2^{-\frac{n}{300}-\frac{n}{10000}} \geq 2^{-n/290}$ before Step (ii) which is enough for an initial application of the Amplification lemma. Then α can decrease by a total factor of at most $(1/2)^{c_3 \log p_0} \geq p_0^{-c_3} \geq 2^{-\frac{n}{2000}}$ during all Case 2 applications of the lemma in Step (ii), hence before each of the applications $\alpha \geq 2^{-n/290-n/2000} \geq 2^{-n/253}$. During the last application of the Amplification lemma (Case 1), α gets amplified by a square root (times 1/8) to attain a value $\alpha \geq \frac{1}{8} \cdot \sqrt{2^{-n/253}} \geq 2^{-n/500}$. This value again permits the use of the Amplification lemma in Step (iii) as α maintains a value at least $2^{-n/500-n/2000-1} \geq 2^{-n/400} > 2^{-n/300}$, as promised. The proof for β is very similar: it is $\geq 2^{-n/300}$ before step (i) and $\geq 2^{-\frac{n}{300}-\frac{n}{10000}} \geq 2^{-n/290}$ after Step (1). It then decreases by a factor of $p_0^{-2c_3} \geq 2^{-\frac{n}{1000}}$ by all possible applications of Case 2 in Steps (ii) and (iii), thus remaining above $2^{-n/290-n/1000-1} \geq 2^{-n/224}$. After Case 1 is executed in Step (iii), $\beta \geq \frac{1}{8}\sqrt{2^{-n/224}} \geq 2^{-n/440} \geq 2^{-n/300}$. It is clear from the previous discussion that $\alpha, \beta \geq 2 \cdot 2^{-n/200}$ at the onset of each application of the Amplification lemma.

By the end of the process we conclude that Protocol$(n, p, \alpha, \beta, \gamma, 0)$ holds, for $\alpha, \beta \geq 2^{-\frac{n}{200}}$, and invariants (1) and (3) give us $\gamma \geq p_0^{-1/25} \geq p^{-1/12}$, for p_0 large enough. The protocol π does not communicate at all so, it is constant. But regularity lemma (Lemma 4.9) does not allow for a constant protocol to have such high success probability on such a large rectangle! Indeed, it implies that each O^i_{AB} has the same size, up to $1 - o(1)$ multiplicative factors. This means that the fractional size of each O^i_{AB} inside O_{AB} is approximately the same, namely $\frac{1}{p} \cdot (1 \pm o(1))$, and which gives an upper bound on the success probability of any constant protocol. Having reached this contradiction, we are forced to conclude that our initial hypothesis about the existence of a protocol communicating $\frac{1}{c} n \log p_0$ bits was false. □

5.2 Proof of the Sub-rectangle lemma

In this subsection we prove Lemma 5.3. Suppose that Protocol$(n, p, \alpha, \beta, \gamma, C)$ holds, with $\alpha, \beta \geq 2^{-n/20}$; let $A \times B$, π be the promised rectangle and protocol. Then to each prefix $w \in \{0,1\}^{\delta n}$ of the transcript of π we can associate a sub-rectangle $R_w \subseteq A \times B$, corresponding to those inputs $(a, b) \in A \times B$ for which w is the first δn bits communicated (we will set $\delta \leq 1/20$ later at our convenience). The success probability on $A \times B$ is then the average success probability over the various R_w, weighted by their λ-mass in $A \times B$:

$$\gamma = \sum_w \frac{\lambda(R_w)}{\lambda(A \times B)} \cdot \gamma_w,$$

where $\gamma_w = \Pr_{(a,b)\sim\lambda}[\pi(a, b) = \mathsf{OS}_p \circ \mathsf{IP}^p_n(a, b) \mid (a, b) \in R_w]$.

Then let us discard all R_w having size smaller than $2^{-2\delta n} \cdot |A \times B|$. By doing so, and given that there are at most $2^{\delta n}$ rectangles R_w, we have discarded at most a $2^{-\delta n}$ fraction of $A \times B$. Now notice that, as $\delta \leq 1/20$, every surviving R_w is still large enough to be $2^{-\frac{n}{10}}$-regular (by Lemma 4.9), and that the union of the surviving rectangles holds at least a $1 - 2^{-\delta n}$ fraction of the pairs in $A \times B$. By Proposition 4.4 applied on each surviving R_w, their union also holds at least a $(1 - 2^{-\frac{n}{10}})(1 - 2^{-\delta n}) \geq 1 - 2 \cdot 2^{-\delta n}$ of the λ-mass of $A \times B$.

Hence, even assuming in the worst case that all discarded rectangles have $\gamma_w = 1$, we still have

$$\sum_{\text{surviving } w} \frac{\lambda(R_w)}{\lambda(A \times B)} \cdot \gamma_w \geq \gamma - 2 \cdot 2^{-\delta n}.$$

But then there must exist a surviving $R_w = A' \times B'$ with $\gamma_w \geq \gamma - 2 \cdot 2^{-\delta n}$. Note that this R_w has size at least $2^{-\delta n} \cdot |A \times B|$ by our construction. At this point, we set δ to be 1/20000 to get Lemma 5.3. □

5.3 Proof of the Amplification lemma

In this subsection we prove the Amplification lemma (Lemma 5.4). Suppose that Protocol(n, p, α, β, γ, C) holds, where:

$$\alpha \geq 2 \cdot 2^{-\frac{n}{200}} \qquad \beta \geq 2 \cdot 2^{-\frac{n}{200}}$$
$$p \leq \tfrac{1}{40} 2^{\frac{n}{100}} \qquad \gamma \geq 40 \cdot p^{-1/12}$$

Let the rectangle $A \times B$ and protocol $\pi : A \times B \to [p]$ witness this fact.

5.3.1 Path splitting

We first split the domain $\{0,1\}^{np}$ into two sides $\{0,1\}^{np_1} \times \{0,1\}^{np_2}$, called the *prefix side* and the *suffix side*. We will do this in a way such that π still has high success probability on both sides, and that neither side is too small.

For a given split choice $p = p_1 + p_2$, let μ_1 be uniformly distributed on the strings $1^i 0^{p-i}$ for $1 \leq i \leq p_1$ and μ_2 be uniformly distributed on the strings $1^i 0^{p-i}$ for $p_1 < i \leq p$. Then (for $i \in \{1,2\}$) let λ_i be the lifting of μ_i (to $(\{0,1\}^{np})^2$, with respect to IP_n^p, see Definition 4.2), and γ_i be the success probability of π on $A \times B$ (with respect to OS_p, IP_n^p and μ_i, see Definition 4.5). If we let

$$O_{AB}^{\leq p_1} = \bigcup_{i=1}^{p_1} O_{AB}^i, \quad T_{AB}^{\leq p_1} = \bigcup_{i=1}^{p_1} T_{AB}^i, \quad O_{AB}^{>p_1} = \bigcup_{i=p_1+1}^{p} O_{AB}^i, \quad \text{and} \quad T_{AB}^{>p_1} = \bigcup_{i=p_1+1}^{p} T_{AB}^i,$$

where O_{AB}^i and T_{AB}^i were defined in Section 5.1.1, then it follows (as in Section 5.1.1) that:

$$\gamma_1 \stackrel{\text{def}}{=} \frac{\lambda_1(T_{AB})}{\lambda_1(A \times B)} = \frac{|T_{AB}^{\leq p_1}|}{|O_{AB}^{\leq p_1}|}, \qquad \gamma_2 \stackrel{\text{def}}{=} \frac{\lambda_2(T_{AB})}{\lambda_2(A \times B)} = \frac{|T_{AB}^{>p_1}|}{|O_{AB}^{>p_1}|}.$$

We may then show the following:

Claim (splitting). There is a choice of p_1 (and thus p_2) such that:

1. $p_1, p_2 \geq \frac{1}{200} p$, and
2. $\gamma_1, \gamma_2 \geq \frac{99}{100} \gamma$

Proof. For each $i \in [p]$, let $\gamma^{(i)} = \frac{|T_{AB}^i|}{|O_{AB}^i|}$; let $\delta = 2^{-\frac{n}{10}}$. For any p_1, p_2, the regularity lemma (Lemma 4.9) will give us the following approximate equalities:

$$\gamma_1 \stackrel{6\delta}{\approx} \frac{1}{p_1} \sum_{i \leq p_1} \gamma^{(i)}, \quad \gamma_2 \stackrel{6\delta}{\approx} \frac{1}{p_2} \sum_{i > p_1} \gamma^{(i)}, \quad \gamma \stackrel{6\delta}{\approx} \frac{1}{p} \sum_i \gamma^{(i)} \stackrel{6\delta}{\approx} \frac{p_1}{p} \gamma_1 + \frac{p_2}{p} \gamma_2 \tag{$*$}$$

Let us derive this only for γ, as the other two equalities follow in the same way. Let $O = \bigcup_{i \in [p]} O^i$, where $O^i = \{(a,b) \in (\{0,1\}^{np})^2 \mid \mathsf{IP}_n^p(a,b) = 1^i 0^{p-i}\}$; then:

$$\gamma = \frac{|T_{AB}|}{|O_{AB}|} \stackrel{2\delta}{\approx} \frac{|T_{AB}|}{\alpha\beta|O|} = \sum_{i=1}^{p} \frac{|T_{AB}^i|}{\alpha\beta|O|} = \frac{1}{p} \sum_{i=1}^{p} \frac{|T_{AB}^i|}{\alpha\beta|O^i|} \stackrel{2\delta}{\approx} \frac{1}{p} \sum_{i=1}^{p} \frac{|T_{AB}^i|}{|O_{AB}^i|} = \frac{1}{p} \sum_{i=1}^{p} \gamma^{(i)}.$$

Both approximate equalities follow from the regularity lemma, and all exact equalities are by definition, except for the third exact equality which follows from $|O| = p|O^i|$ (because each O^i set has exactly the same size 2^{np-p}).

To avoid encumbering the argument, let us prove the existence of p_1 and p_2 assuming that the equalities in ($*$) hold exactly. Set $L = \frac{p}{200}$ and $R = p - \frac{p}{200}$. If every choice of p_1 between L and R gives $\gamma_1, \gamma_2 \geq \frac{99}{100}\gamma$, we can just set $p_1 = \frac{p}{2}$. Otherwise, suppose without loss of generality that there is some p' between L and R for which $\gamma_1 < \frac{99}{100}\gamma$ (the case for when γ_2 is small for some p' is symmetric). Let p_1 be the smallest index in $\{p', p'+1, \ldots, R\}$ for which $\gamma_1 \geq \frac{99}{100}\gamma$. Such an index must exist, because setting $p_1 = R$ will be enough: the number of indices $i > R$ is less than $\frac{1}{200}$ fraction of all indices, so if γ_1 were less than $\frac{99}{100}\gamma$ when $p_1 = R$, the average γ could not possibly be attained.

Now notice that since $p' \geq L$, γ_1 can only increase by $\frac{1}{L}$ every time we increment p_1. Hence for this choice of p_1, it must happen that $\gamma_1 \leq \frac{99}{100}\gamma + \frac{1}{L}$, and since by assumption $\gamma \geq p^{-1/12} \gg \frac{200}{L}$, then $\gamma_1 \leq \frac{199}{200}\gamma$. That immediately implies that, to attain the average, γ_2 must be $\geq \gamma$.

It is now easy to see how the result follows from the approximate inequalities (since $\delta \ll 1/p$). $\square$

☞ From now onward, we fix the choice of p_1, p_2 to have the properties of the previous claim.

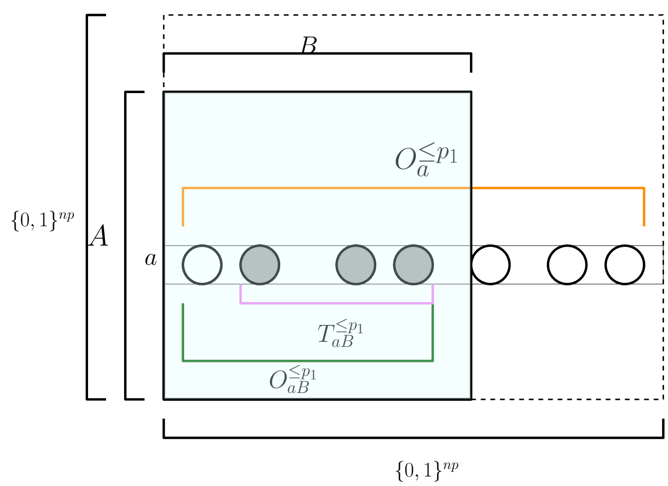
5.3.2 High-quality subsets

Given $a \in A$, the *prefix-side quality of a* is the row-quality (as in Definition 4.6) with respect to μ_1 (and $f = \mathsf{OS}_p$, $G = \mathsf{IP}_n^p$, A, B, and π). Similarly, we define the suffix-side quality of $a \in A$, $q_2(a)$, to be the row-quality with respect to μ_2. Define:

$$T_{aB}^{\leq p_1} = \{b \in B \mid \pi(a,b) = \mathsf{OS}_p \circ \mathsf{IP}_n^p(a,b) \text{ and } \mathsf{IP}_n^p(a,b) = 1^i 0^{p-i} \text{ for some } i \in [p_1]\}, \text{ and}$$

$$O_a^{\leq p_1} = \{b \in \{0,1\}^{np} \mid \mathsf{IP}_n^p(a,b) = 1^i 0^{p-i} \text{ for some } i \in [p_1]\}.$$

The following image is useful for thinking about these sets: we look at row a in the $\mathsf{OS}_p \circ \mathsf{IP}_n^p$ matrix; within this row, mark with a circle those columns b for which $\mathsf{IP}_n^p = 1^i 0^{p-i}$ for some $i \in [p_1]$; then $O_a^{\leq p_1}$ is the set of b that were marked, $O_{aB}^{\leq p_1}$ is the set of these circles within B and $T_{aB}^{\leq p_1}$ is the subset $O_a^{\leq p_1}$ where the protocol π is correct (these entries appear as gray-filled circles in the picture; $T_{aB}^{\leq p_1}$ is also a subset of B, which is why B appears in the notation).



The sets $T_{aB}^{>p_1}$ and $O_a^{>p_1}$ are similarly defined with respect to $i \in [p] \setminus [p_1]$. It then follows (as in Section 5.1.1) that

$$q_1(a) = \frac{|T_{aB}^{\leq p_1}|}{|O_a^{\leq p_1}|} \qquad\qquad q_2(a) = \frac{|T_{aB}^{>p_1}|}{|O_a^{>p_1}|},$$

Abbreviate $\delta = 2^{-\frac{n}{10}}$; then Lemma 4.7 says that the average q_i is:

$$\frac{1}{|A|}\sum_{a \in A} q_1(a) \overset{2\delta}{\approx} \gamma_1 \beta \qquad\qquad \frac{1}{|A|}\sum_{a \in A} q_2(a) \overset{2\delta}{\approx} \gamma_2 \beta.$$

So let us now focus on those $a \in A$ which attain at least $\frac{1}{10}$ of this average:

$$A_1 = \{a \in A \mid q_1(a) > \gamma_1\beta/10\} \qquad\qquad A_2 = \{a \in A \mid q_2(a) > \gamma_2\beta/10\}.$$

A_1 and A_2 are called the *high-quality subsets* of A.

5.3.3 The conditions for each of the two cases

Depending on the size of the high-quality subsets, we consider the following two exhaustive cases:

Case 1 Both $|A_1|$ and $|A_2|$ are at least $\frac{3}{4}|A|$,

Case 2 At least one of $|A_1|, |A_2|$ is less than $\frac{3}{4}|A|$.

In accordance with the statement of the Amplification lemma, if Case 1 holds we will show that Protocol$(n, p', \frac{1}{8}\sqrt{\alpha}, \frac{1}{2}\beta, \frac{1}{11}\gamma, C)$ holds, for p' equal to either p_1 or p_2, — the choice of p_i is decided on the basis of the sets A_1 and A_2, (i.e. we "amplify" α), and if Case 2 holds we will show that Protocol$(n, p', \frac{1}{2}\alpha, \frac{1}{2}\beta, \frac{11}{10}\gamma, C)$ holds, for p' equal to either p_1 or p_2 (i.e. we "amplify" γ), — the choice of p_i, again, is decided on the basis of A_i which has size smaller than $3|A|/4$.

5.3.4 Proving Case 1

Let $A' = A_1 \cap A_2$ — which is a set of size at least $|A|/2$. We will apply lemma 2.6 to find a *super-dense* side of A'. Let $\alpha' = \frac{|A'|}{2^{np}}$, $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$. For $\ell \in \mathcal{L}$, let $\mathsf{Ext}'(\ell) = \mathsf{Ext}_{A'}^{[p]\setminus[p_1]}(\ell)$ be the (possibly empty) set of $r \in \mathcal{R}$ such that $\ell \times r \in A'$ (Note that $\ell \times r$, as defined in Section 2, is the concatenation of ℓ with r); likewise, for $r \in \mathcal{R}$, let $\mathsf{Ext}'(r) = \mathsf{Ext}_{A'}^{[p_1]}(r)$ be the set of $\ell \in \mathcal{L}$ such that $\ell \times r \in A'$. We define the two sets:

$$A'_{\mathrm{L}} = \left\{ \ell \in \mathcal{L} \;\middle|\; \frac{|\mathsf{Ext}'(\ell)|}{|\mathcal{R}|} > \frac{\alpha'}{4} \right\}, \qquad\qquad A'_{\mathrm{R}} = \left\{ r \in \mathcal{R} \;\middle|\; \frac{|\mathsf{Ext}'(r)|}{|\mathcal{L}|} > \frac{\alpha'}{4} \right\}.$$

Applying Lemma 2.6, we conclude that either A'_{L} or A'_{R} is $\frac{1}{4}\sqrt{\alpha'}$-dense in $\mathcal{L}$ or $\mathcal{R}$, respectively.

The proof of Case 1 will use the following lemma (and its suffix-side analogue):

Lemma 5.5 (Zooming-in lemma, prefix side, weak version)**.** Let $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$, where $p = p_1 + p_2 \le \frac{1}{40} \cdot 2^{n/100}$ is a sufficiently large natural number. Suppose we have a rectangle $A'' \times B$, where both A'' and B are subsets of $\mathcal{L} \times \mathcal{R}$, and a C bit protocol $\pi : A'' \times B \to [p]$. Let μ_1 be a uniform distribution over the strings $1^i 0^{p-i}$ for $i \in [p_1]$, and let λ_1 be the lifting of μ_1 to $(\mathcal{L} \times \mathcal{R})^2$ with respect to IP_n^p (as in Definition 4.2). Let $q_1(a)$ denote the row-quality with respect to μ_1 (and OS_p, IP_n^p, A'', B and π). For a given $\ell \in A''_{\le p_1}$, denote by $\mathsf{Ext}''(\ell) = \mathsf{Ext}_{A''}^{[p]\setminus[p_1]}(\ell)$ the set of extensions of ℓ.

Suppose we have the following properties:

- *A'' and B have enough density.* A'' has prefix-side density $\alpha''_{\le p_1} \stackrel{\text{def}}{=} \frac{|A''_{\le p_1}|}{|\mathcal{L}|}$ which is at least $2^{-\frac{n}{10}}$, and for each $\ell \in A''_{\le p_1}$, the density of its extensions $\frac{|\mathsf{Ext}''(\ell)|}{|\mathcal{R}|}$ is at least $8 \cdot 2^{-\frac{n}{30}}$. The density of B, $\beta \stackrel{\text{def}}{=} \frac{|B|}{|\mathcal{L} \times \mathcal{R}|}$, is at least $2^{-\frac{n}{200}}$.
- *Minimum quality in A'' is high.* For some value $\tilde{\gamma} \ge 2^{-\frac{n}{1200}}$, every $a \in A''$ has $q_1(a) \ge \tilde{\gamma}\beta$.

CONCLUSION. Then Protocol$(n, p_1, (1-\delta)\alpha_{\le p_1}, (1-\delta)\beta, (1-\delta)\tilde{\gamma}, C)$ holds, where $\delta = 8 \cdot 2^{-\frac{n}{1200}}$.

Here it should be noted that a Zooming-in lemma similar to the above was implicitly proven in [RW89], though it was for the *Indexing* function. However, the asymmetry of the Indexing function allowed for their proof to be somewhat simpler than what will be afforded to us. In

Section 5.5 we will prove a stronger version of the Zooming-in lemma just presented — this stronger version will be needed in Case 2.

Now to finish Case 1: Suppose that $\frac{|A'_{\mathrm{L}}|}{|\mathcal{L}|} \geq \frac{1}{4}\sqrt{\alpha'}$. Let $A'' \subseteq \mathcal{L} \times \mathcal{R}$ contain every ℓ in A'_{L} and its extensions, i.e. $A'' = \{\ell \times r \mid \ell \in A'_{\mathrm{L}}, \ell \times r \in A'\}$, and let us see why we may apply the Zooming-in lemma above. A'' and B have *enough density* since $\frac{\alpha'}{4} \gg 2^{-\frac{n}{30}}$ and $\beta \geq 2^{-\frac{n}{200}}$. On the other hand, the quality $q_1(a)$ of every $a \in A''$ is at least $\frac{\gamma_1}{10}\beta$, because $A'' \subseteq A'$. Hence we set $\tilde{\gamma} = \frac{\gamma_1}{10}$ above, which is $\geq \frac{99}{1000}\gamma \geq \frac{99 \times 20}{1000} p^{-1/12} \geq 2^{-\frac{n}{1200}}$.

It then follows from the Zooming-in lemma that Protocol$(n, p_1, \frac{1}{8}\sqrt{\alpha}, \frac{1}{2}\beta, \frac{\gamma}{11}, C)$. A similar conclusion also follows from the analogous suffix-side Zooming-in lemma (which we state in Lemma 5.17) if A_{R} is super-dense. This concludes Case 1.

5.3.5 Proving Case 2

Let us suppose that $|A_1| < \frac{3}{4}|A|$. The intuition behind Case 2 is as follows. The average prefix-side quality of $a \in A$ is $\gamma_1\beta$, but fewer than $\frac{3}{4}$ of the inputs in A have prefix-side quality $\geq \frac{\gamma_1\beta}{10}$. That means, roughly, that $\frac{9}{10}$ of all the prefix-side quality is concentrated in less than $\frac{3}{4}$ of the strings;— then these strings must have higher than average prefix-side quality, namely $\frac{4}{3} \cdot \frac{9}{10}\gamma_1\beta = \frac{12}{10}\gamma_1\beta$. We will show that we may find a subset of the prefix-side projection of A, and a carefully selected subset of the extensions of these prefix-sides, such that we may *zoom in* on the prefix-side to get a protocol with success probability $\geq \frac{11}{10}\gamma$.

For this purpose, let $A' \subseteq A$ be the subset of A containing the $\lfloor \frac{3}{4}|A| \rfloor$ strings $a \in A$ that have highest prefix-side quality $q_1(a)$. Because $A_1 \subseteq A'$, every string $a \in A \setminus A'$ has $q_1(a) < \frac{\gamma\beta}{10}$. Let $\delta = 2^{-n/10}$. Since $A \times B$ is δ-regular (by 4.9), then from the success–quality correspondence (Lemma 4.7), we have that $\frac{1}{|A|}\sum_{a \in A} q_1(a) \geq (1-\delta)\gamma_1\beta$; it must now hold:

$$\sum_{a \in A'} q_1(a) \geq (1 - \delta - 1/10) \cdot \gamma_1\beta|A| \geq (1 - 2\delta) \cdot \frac{12}{10} \cdot \gamma_1\beta \cdot |A'|,$$

i.e., the average quality in A' is roughly $\frac{12}{10}$ times higher. We will now show that we can prune A' to obtain a set A'' with an equally dense prefix-side projection $A''_{\leq p_1}$, and such that all extensions of each $\ell \in A''_{\leq p_1}$ have fairly good quality. More precisely:

Lemma 5.6 (Min-quality lemma). Let $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$ for some sufficiently large natural numbers n, p_1 and p_2. Suppose we have a rectangle $A' \times B$, where both A' and B are subsets of $\mathcal{L} \times \mathcal{R}$, and a protocol $\pi : A' \times B \to [p]$. Let μ_1 be uniform over the strings $1^i 0^{p-i}$ for $i \in [p_1]$, and let $q_1(a)$ denote the row-quality (Definition 4.6) with respect to μ_1 (and OS, IP_n^p, A', B and π). If we have fixed a subset $A'' \subseteq A$, then for any given $\ell \in A''_{\leq p_1}$ let $\mathsf{Ext}''(\ell) = \mathsf{Ext}_{A''}^{[p] \setminus [p_1]}(\ell)$ be the set of extensions $r \in \mathcal{R}$ with $\ell \times r \in A''$, and define the *min-quality* of ℓ, $q''_{\min}(\ell)$, to be the minimum q_1 of its extensions:

$$q''_{\min}(\ell) \stackrel{\text{def}}{=} \min_{r \in \mathsf{Ext}''(\ell)} q_1(\ell \times r).$$

Now suppose we have the following properties:

- *A' and B have good density.* $\alpha' \stackrel{\text{def}}{=} \frac{|A'|}{|\mathcal{L} \times \mathcal{R}|} \geq 2^{-\frac{n}{200}}$, and $\beta \stackrel{\text{def}}{=} \frac{|B|}{|\mathcal{L} \times \mathcal{R}|} \geq 2^{-\frac{n}{200}}$.
- *Average quality is high.* For some value $Q \geq 2 \cdot 2^{-\frac{n}{150}}$ it holds that:

$$\frac{1}{|A'|}\sum_{a \in A'} q_1(a) \geq Q.$$

Conclusion. Then there is a subset $A'' \subseteq A'$ with the following properties:

- A'' **has enough density.** The size of the prefix-side projection is $|A''_{\leq p_1}| \geq \lfloor (1-2^{-\frac{n}{120}}) \cdot \alpha' |\mathcal{L}| \rfloor$, and for all $\ell \in A''_{\leq p_1}$ we have $|\mathsf{Ext}''(\ell)| \geq 8 \cdot 2^{-\frac{n}{30}} |\mathcal{L}|$;

- A'' **obeys the average min-quality condition.** The average min-quality over $\ell \in A''_{\leq p_1}$ almost matches the average quality in A':

$$\frac{1}{|A''_{\leq p_1}|} \sum_{\ell \in A''_{\leq p_1}} q''_{\min}(\ell) \geq (1 - 4 \cdot 2^{-\frac{n}{300}}) \cdot Q.$$

A variant of the above lemma appears implicitly in [RW89]. Our proof appears in Section 5.6, and is based on various proofs in that paper.

We apply this lemma to our rectangle $A' \times B$, with $Q = (1-2\delta)\frac{12}{10}\gamma_1\beta$, which is $\geq 2 \cdot 2^{-\frac{n}{150}}$ since $\beta \geq 2^{-\frac{n}{200}}$ and $\gamma_1 \geq \frac{99}{100}\gamma \geq \frac{99}{100} \cdot 40 \cdot p^{-1/12} \geq 2 \cdot 2^{-\frac{n}{600}}$. Now we have a set A'' with enough density and which obeys the min-quality condition; we can now apply the Lemma 5.16 (the *strong version* of the Zooming-in lemma which we used in Case 1), to conclude that Protocol$(n, p_1, \frac{\alpha}{2}, \frac{\beta}{2}, \frac{11}{10}\gamma, C)$ must hold. $\square$

5.4 The extension lemma

We now prove what we call an *extension lemma.* An extension lemma is a stronger version of the hitting rectangle-distribution property appearing in Section 3. The statement is somewhat technical, but let us give it now in full so that we can explain the analogy with the hitting properties.

Lemma 5.7 (0-monochromatic extension lemma)**.** Let p and n be sufficiently large natural numbers, such that $p \leq \frac{1}{40} \cdot 2^{\frac{n}{100}}$. Let $\mathcal{O}$ be some finite set.

- *Dense set of extensions.* Let $\mathsf{Ext} \subseteq \{0,1\}^{np}$ with $\alpha = \frac{|\mathsf{Ext}|}{2^{np}} \geq 8 \cdot 2^{-n/30}$.
- *Associated set.* Suppose that to each $r \in \mathsf{Ext}$ corresponds a set $T_r \subseteq \mathcal{O} \times r^{\perp}$, where

$$r^{\perp} = r_1^{\perp} \times \ldots \times r_p^{\perp} = \{r' \in \{0,1\}^{np} \mid \mathsf{IP}_n^p(r, r') = 0^p\}.$$

- *Quality.* Define the quality of r to be

$$q(r) \triangleq \frac{|T_r|}{|\mathcal{O} \times r^{\perp}|},$$

and suppose that $q(r) \geq 2^{-n/10}$ for every $r \in \mathsf{Ext}$.

- 0-*monochromatic rectangle.* Now pick a random product $V = V_1 \times \cdots \times V_p$, where each V_i is an independent and uniformly random $\lfloor \frac{n}{2} \rfloor$-dimensional random subspace of $\mathbb{F}_2^n$. Let $V^{\perp}$ denote $V_1^{\perp} \times \cdots \times V_p^{\perp}$.
- *Quality in the monochromatic rectangle.* Finally, define

$$q_V(r) = \frac{|T_r \cap (\mathcal{O} \times V^{\perp})|}{|\mathcal{O} \times V^{\perp}|}.$$

Conclusion. Then with probability $\geq 1 - 2^{-\frac{n}{150}}$ over the choice of V, there is some extension $r \in \mathsf{Ext} \cap V$ whose quality is preserved in the 0-monochromatic rectangle:

$$\exists r \in \mathsf{Ext} \text{ such that } \begin{cases} r \in V & (1) \\ q_V(r) \in (1 \pm 2^{-n/30}) \cdot q(r) & (2) \end{cases}$$

The hitting property gave us a rectangle-distribution that was almost guaranteed to hit large rectangles. Here we have a set $\bigcup_{r\in\mathsf{Ext}}\{r\}\times T_r$ which is not necessarily a rectangle — we may think of it as a union of *slices*. The extension lemma says that if there are many such slices (Ext is big), and each slice is large within $r^\perp$ (every $q(r)$ is big), then by picking a 0-monochromatic rectangle $V\times V^\perp$, we will with very high probability "hit" one of the slices, where "hitting" here means that $r\in V$ and $T_r\cap V^\perp$ has the same density within $V^\perp$ as T_r has within $r^\perp$.

This property, and the regularity of large rectangles, are the driving forces behind Theorem 5.1.

To prove the extension lemma we will need to considerably strengthen Lemmas 3.13 and 3.14, which we will do in Section 5.4.1 and Section 5.4.2. The proof of the extension lemma itself appears in Section 5.4.3, and the 1-monochromatic extension lemma is stated and proven in Section 5.4.4.

5.4.1 Generalizing Section 3.3 to multiple coordinates

We begin by extending the proofs of Section 3.3 to random product subspaces.

Lemma 5.8. Let $B\subseteq\mathcal{B}=\mathcal{B}_1\times\cdots\times\mathcal{B}_p$, where $\mathcal{B}_j=\mathbb{F}_2^n$, and the remaining $\mathcal{B}_i$'s are arbitrary finite sets. Suppose that $\beta=\frac{|B|}{|\mathcal{B}|}\geq 8\cdot 2^{-\frac{n}{4}}$. Pick V to be a random subspace of $\mathcal{B}_j$ of dimension $d\geq\frac{7n}{15}$, and let $U=\mathcal{B}_1\times\cdots\times V\times\cdots\times\mathcal{B}_p$ (V replaces $\mathcal{B}_j$). Then

$$\Pr\left[\frac{|B\cap U|}{|U|}\in\beta(1\pm 2^{-n/20})\right]\geq 1-\frac{1}{2^{n/20}}.$$

Proof. Let $b_1,\ldots,b_N$ be the elements of B_j (the projection of B into the j-th coordinate), and for each b_i let $\mathsf{Ext}(b_i)=\mathsf{Ext}_B(b_i)$ be the set of extensions of b_i into B.

Define the random variables $X_i=\theta_i[b_i\in V]$, where $\theta_i=|\mathsf{Ext}(b_i)|/|\mathcal{B}_{\neq j}|$ is the fractional size of $\mathsf{Ext}(b_i)$ in the set $\mathcal{B}_{\neq j}=\prod_{k\neq j}\mathcal{B}_k$. Note that $\sum_i\theta_i=\beta 2^n$. Then the sum $X=\sum_i X_i$ equals $\frac{|B\cap U|}{|\mathcal{B}_{\neq j}|}=\frac{|B\cap U|}{|U|}|V|$. We wish to prove that $X\in\beta|V|(1\pm 2^{-n/20})$ with high probability. To this end, let us first compute $\mathbf{E}[X]$.

$$\mu=\mathbf{E}[X]=\sum_i\mathbf{E}[X_i]=\sum_i\theta_i\Pr[b_i\in V]=\begin{cases}\beta 2^n\frac{2^d-1}{2^n-1} & \text{if }\bar{0}\notin B_j,\\ \beta 2^n\frac{2^d-1}{2^n-1}+\theta_0(1-\frac{2^d-1}{2^n-1}) & \text{otherwise.}\end{cases}$$

(θ_0 denotes θ_j for the j such that $b_j=\bar{0}$.) Note that $\theta_0\leq 1$. Hence we can bound μ as follows:

$$\beta 2^n\frac{2^d-1}{2^n-1}\leq\mu\leq\beta 2^n\frac{2^d-1}{2^n-1}+1-\frac{2^d-1}{2^n-1}.$$

As we have argued in the proof of Lemma 3.13, this implies that $\mu\in(1\pm 2^{-n/6})\cdot\beta|V|$. Using second moment method (Lemma 2.4) and noting that the X_i's are anti-correlated (Lemma 3.12), we may write:

$$\Pr\left[X\in(1\pm 2^{-n/6})(1\pm\varepsilon/2)\beta|V|\right]\geq 1-\frac{4}{\varepsilon^2\beta 2^d(1-2^{-n/6})}.$$

Taking $\varepsilon=2^{-n/20}$, we get,

$$\Pr\left[X\in(1\pm 2^{-n/20})\beta|V|\right]\geq 1-\frac{1}{2\cdot 2^{\frac{7n}{15}-\frac{n}{4}-\frac{2n}{20}}(1-2^{-n/6})}\geq 1-\frac{1}{2^{n/20}}.\qquad\square$$

We may extend this result as follows.

Lemma 5.9. Let $B \subseteq \{0,1\}^{np}$, $\beta = \frac{|B|}{2^{np}} \geq 16 \cdot 2^{-n/4}$, and $p \leq \frac{1}{4} \cdot 2^{\frac{n}{100}}$. Pick $V = V_1 \times \ldots \times V_p$ where each V_i is an independently chosen random subspace of $\mathbb{F}_2^n$, of dimension $d \geq \frac{7n}{15}$. Then

$$\Pr\left[\frac{|B \cap V|}{|V|} = \beta(1 \pm 2^{-n/25})\right] \geq 1 - \frac{1}{2^{n/25}}.$$

Proof. Apply Lemma 5.8 p times, once to each coordinate. To apply Lemma 5.8, at each time, we must ensure that the density never goes below $8 \cdot 2^{-n/4}$. This will hold, provided that $(1-2^{-n/20})^p \geq 1/2$, which is always the case for our choice of p. It follows from Bayes' rule that:

$$\Pr\left[\frac{|B \cap V|}{|V|} = \beta(1 \pm 2^{-n/20})^p\right] \geq 1 - p \cdot 2^{-n/20}, \tag{3}$$

Now, it is not hard to verify that the interval $(1 \pm 2^{-n/20})^p$ is contained in $1 \pm 2^{-n/25}$. The obvious direction is $(1-2^{-n/20})^p \geq 1 - p \cdot 2^{-n/20} \geq 1 - 2^{-n/25}$. For the other direction, note that $p \cdot 2^{-n/20} < 1$. Now, it is easy to check that for any p, δ with $p\delta < 1$, $(1 + e\delta p) \geq (1+\delta)^p$ (by taking the derivative on both sides w.r.t. δ). Hence $(1 + 2^{-n/20})^p \leq 1 + 2^{-n/25}$. □

The above results are natural extensions of the original principle. We will also need a somewhat technical variant of these results. It may be proven in the same way as Lemma 5.9, or more cleverly by noticing that each $r_i^{\perp}$ above is isomorphic to $\mathbb{F}_2^{n-1}$. (The $\frac{n}{30}$ in the statement is just a very rough lower-bound on $\frac{n-1}{25}$.)

Lemma 5.10. Let $\mathcal{L}$ be an arbitrary finite set, $r \in \{0,1\}^{np}$ and $r^{\perp} = r_1^{\perp} \times \ldots \times r_p^{\perp}$, where each $r_i^{\perp} = \{v \in \mathbb{F}_2^n \mid \mathsf{IP}_n(r,v) = 0\}$ is the perpendicular subspace to r_i. Let $p \leq \frac{1}{4} 2^{\frac{n}{100}}$, $D \subseteq \mathcal{L} \times r^{\perp}$ with $\delta = \frac{|D|}{|\mathcal{L} \times r^{\perp}|} \geq 8 \cdot 2^{-n/8}$. Now pick $V = V_1 \times \ldots \times V_p$ where each V_i is a dimension $d \geq \frac{7n}{15}$, independent random subspace of $r_i^{\perp}$. Set $U = \mathcal{L} \times V$. Then

$$\Pr\left[\frac{|D \cap U|}{|U|} = \delta \cdot (1 \pm 2^{-n/30})\right] \geq 1 - \frac{1}{2^{n/30}}.$$

5.4.2 Generalizing to the affine case

A similar result holds even if we work in the following scenario: Instead of picking V be a random subspace, instead we first pick a uniformly-random vector $a \in \mathbb{F}_2^n$ of odd Hamming weight, and then we pick W, a uniformly-random subspace of dimension $d \geq \frac{7(n-1)}{15}$ within $a^{\perp}$. We finally let $V = a + W$. The following can now be proven:

Lemma 5.11 (Analogue of Lemma 5.8)**.** Let $B \subseteq \mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_p$, where $\mathcal{B}_j = \mathbb{F}_2^n$, and the remaining $\mathcal{B}_i$'s are arbitrary finite sets. Suppose that $\beta = \frac{|B|}{|\mathcal{B}|} \geq 16 \cdot 2^{-\frac{n}{4}}$. Pick V as described above and let $U = \mathcal{B}_1 \times \cdots \times V \times \cdots \times \mathcal{B}_p$ (V replaces $\mathcal{B}_j$). Then

$$\Pr\left[\frac{|B \cap U|}{|U|} \in \beta(1 \pm 3 \cdot 2^{-n/20})\right] \geq 1 - \frac{3}{2^{n/20}}.$$

Proof. This proof uses Lemma 5.8 in the same way that the proof of Lemma 3.14 uses Lemma 3.13. Let $B' = B - a'$ where $a' \in \{0,1\}^{np}$ has $a'_j = a$ and $a'_i = \bar{0}$ for $i \neq j$. Also denote $U' = \mathcal{B}_1 \times \cdots \times a^{\perp} \times \cdots \mathcal{B}_p$. Call a string $a \in \{0,1\}^n$ *good* if

$$\beta' \stackrel{\text{def}}{=} \frac{|B' \cap U'|}{|U'|} \in \beta(1 \pm 2^{-\frac{n}{20}})$$

We show below that if a is a uniformly-random odd-Hamming-weight string in $\{0,1\}^n$, then

$$\Pr_a[a \text{ is good}] \geq 1 - \frac{2}{2^{n/20}}. \tag{$*$}$$

Assuming $(*)$, let $U'' = \mathcal{B}_1 \times \cdots \times W \times \cdots \mathcal{B}_p$; notice that Lemma 5.8 then implies

$$\Pr_{a,W}\left[\frac{|B' \cap U''|}{|U''|} \in \beta'(1 \pm 2^{-n/20}) \,\middle|\, a \text{ is good}\right] \geq 1 - \frac{1}{2^{n/20}}.$$

This is enough to prove the theorem, as we have $|U''| = |U|$ and — because $b - a' \in B' \cap U'' \iff b \in B \cap U$ — we also have $|B' \cap U''| = |B \cap U|$; the result now follows from Bayes' rule.

$(*)$ is proven in much the same way as in the proof of Lemma 3.14, with the added encumbrance of handling multiple coordinates. If we choose a to be a uniformly-random non-zero string in $\{0,1\}^n$, then $a^\perp$ is a uniformly-random subspace of dimension $n-1$. Let $U_0 = \mathcal{B}_1 \times \cdots \times \mathbb{F}_2^n \times \cdots \mathcal{B}_p$. Since $\frac{|B|}{|U_0|} = \beta$, then applying Lemma 3.14 we conclude that

$$\frac{|B \cap U'|}{|U'|} = (1 \pm 2^{-n/20}) \cdot \beta$$

must hold with probability $\geq 1 - 1/2^{n/20}$ (over the choice of a). On the other hand, since $a^\perp$ contains exactly half of the strings in $\mathbb{F}_2^n$, then for $U^\parallel = \mathcal{B}_1 \times \cdots \times a^\parallel \times \cdots \mathcal{B}_p$, we have $|U^\parallel| = |U'|$ and $\frac{|B \cap U'|}{|U'|} = (1 \pm 2^{-n/20}) \cdot \beta$ if and only if $\frac{|B \cap U^\parallel|}{|U^\parallel|} = (1 \pm 2^{-n/20}) \cdot \beta$. It then follows that, with probability $\geq 1 - 1/2^{n/20}$ over the choice of non-zero a,

$$\frac{|B \cap U^\parallel|}{|U^\parallel|} = (1 \pm 2^{-n/20}) \cdot \beta$$

Now, a uniformly-random non-zero a it will be of odd Hamming weight with probability $\geq 1/2$. Also, note that when a is an odd-Hamming-weight string, then $b \in B \cap U^\parallel \iff b - a' \in B' \cap U'$, so $|B \cap U^\parallel| = |B' \cap U'|$. We then conclude that $(*)$ must hold. $\square$

The above lemma can then be used to prove the analogue of Lemma 5.9 for $V = a + W$:

Lemma 5.12 (Analogue of Lemma 5.9)**.** Let $B \subseteq \{0,1\}^{np}$, $\beta = \frac{|B|}{2^{np}} \geq 16 \cdot 2^{-n/4}$, and $p \leq \frac{1}{12} \cdot 2^{\frac{n}{100}}$. Pick $V = V_1 \times \ldots \times V_p$ where each V_i is chosen independently to be $a_i + W_i$ as described above. Then

$$\Pr\left[\frac{|B \cap V|}{|V|} = \beta(1 \pm 2^{-n/25})\right] \geq 1 - \frac{1}{2^{n/25}}.$$

We are still missing the affine analogue of the technical variant (Lemma 5.10). Fix a vector $r \in \{0,1\}^n$ such that $r \neq 0^n$. Denote by $r^\parallel$ the set of $x \in \{0,1\}^n$ such that $\mathsf{IP}_n(x,r) = 1$. Now pick the set V by the following process: first pick a uniformly-random vector $a \in r^\parallel$ of odd Hamming weight; then pick W, a uniformly-random subspace of dimension $d \geq \frac{7(n-1)}{15}$ within $\{r,a\}^\perp$; then set $V = a + W$. The following is now true:

Lemma 5.13. Let $B \subseteq \mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_p$, where $\mathcal{B}_j = r^\parallel$ (where $r \in \{0,1\}^n$ is $\neq 0^n$), and the remaining $\mathcal{B}_i$'s are arbitrary finite sets. Suppose that $\beta = \frac{|B|}{|\mathcal{B}|} \geq 16 \cdot 2^{-\frac{n}{4}}$. Pick V as stated above and let $U = \mathcal{B}_1 \times \cdots \times V \times \cdots \times \mathcal{B}_p$ (V replaces $\mathcal{B}_j$). Then

$$\Pr\left[\frac{|B \cap U|}{|U|} \in \beta(1 \pm 5 \cdot 2^{-n/20})\right] \geq 1 - \frac{10}{2^{n/20}}.$$

Proof. This proof mimics the proof of Lemma 5.11, with some added care to deal with the fact that the ambient set $r^\parallel$ is not a subspace — it will be enough that it is a large set within $\mathbb{F}_2^n$. Let $B' = B - a'$ where $a' \in \{0,1\}^{np}$ has $a'_j = a$ and $a'_i = \bar{0}$ for $i \neq j$. Also denote $U' = \mathcal{B}_1 \times \cdots \times \{a,r\}^\perp \times \cdots \mathcal{B}_p$. Call a string $a \in \{0,1\}^n$ *good* if

$$\beta' \stackrel{\text{def}}{=} \frac{|B' \cap U'|}{|U'|} \in \beta(1 \pm 3 \cdot 2^{-\frac{n}{20}})$$

We show below that if a is a uniformly-random odd-Hamming-weight string in $r^{\|}$, then

$$\Pr_a[a \text{ is good}] \geq 1 - \frac{8}{2^{n/20}}. \tag{$*$}$$

Assuming $(*)$, let $U'' = \mathcal{B}_1 \times \cdots \times W \times \cdots \mathcal{B}_p$; notice that Lemma 3.13 then implies

$$\Pr_{a,W}\left[\frac{|B' \cap U''|}{|U''|} \in \beta'(1 \pm 2^{-n/20}) \;\middle|\; a \text{ is good}\right] \geq 1 - \frac{1}{2^{n/20}}.$$

This is enough to prove the theorem, as we have $|U''| = |U|$ and — because $b - a' \in B' \cap U'' \iff b \in B \cap U$ — we also have $|B' \cap U''| = |B \cap U|$; the result now follows from Bayes' rule.

To prove $(*)$ we now have the added encumbrance of handling multiple coordinates, one of which is $r^{\|}$ instead of $\{0,1\}^n$. If we choose a to be a uniformly-random non-zero string in $\{0,1\}^n$, then $a^{\perp}$ is a uniformly-random subspace of dimension $n-1$. Let $R^{\perp} = \mathcal{B}_1 \times \cdots \times r_i^{\perp} \times \cdots \mathcal{B}_p$, $U_0 = \mathcal{B}_1 \times \cdots \times \mathbb{F}_2^n \times \cdots \mathcal{B}_p$, and $U^{\perp} = \mathcal{B}_1 \times \cdots \times a^{\perp} \times \cdots \mathcal{B}_p$. Since $\frac{|B|}{|U_0|} = \frac{\beta}{2}$ and $\frac{|R^{\perp}|}{|U_0|} = \frac{1}{2}$, then applying Lemma 3.14 twice, we conclude that

$$\frac{|B \cap U^{\perp}|}{|U^{\perp}|} = (1 \pm 2^{-n/20}) \cdot \beta/2 \qquad \text{and} \qquad \frac{|R^{\perp} \cap U^{\perp}|}{|U^{\perp}|} = (1 \pm 2^{-n/20}) \cdot 1/2$$

must both hold with probability $\geq 1 - 2/2^{n/20}$ (over the choice of a). On the other hand, since $a^{\perp}$ contains exactly half of the strings in $\mathbb{F}_2^n$, then for $U^{\|} = \mathcal{B}_1 \times \cdots \times a^{\|} \times \cdots \mathcal{B}_p$, we have $|U^{\|}| = |U^{\perp}|$ and $\frac{|B \cap U^{\perp}|}{|U^{\perp}|} = (1 \pm 2^{-n/20}) \cdot \beta/2$ if and only if $\frac{|B \cap U^{\|}|}{|U^{\|}|} = (1 \pm 2^{-n/20}) \cdot \beta/2$. It then follows that, with probability $\geq 1 - 2/2^{n/20}$ over the choice of non-zero a,

$$\frac{|B \cap U^{\|}|}{|R^{\perp} \cap U^{\perp}|} = \frac{|B \cap U^{\|}|}{|U^{\|}|} \frac{|U^{\perp}|}{|R^{\perp} \cap U^{\perp}|} \in (1 \pm 3 \cdot 2^{-n/20})\beta.$$

Now, a uniformly-random non-zero a it will be within $r^{\|}$ and will be of odd Hamming weight with probability $\geq 1/4$ (here we use the fact that r is not an all-0 or all-1 string). Also, note that when a is an odd-Hamming-weight string in $r^{\|}$, then $b \in B \cap U^{\|} \iff b - a' \in B' \cap U'$ (here recall that $B_i \subseteq r_i^{\|}$ also), so $|B \cap U^{\|}| = |B' \cap U'|$. Finally notice that $R^{\perp} \cap U^{\perp} = U'$. We then conclude that $(*)$ must hold. □

We may apply Lemma 5.13 to each coordinate, exactly as in the proof of Lemma 5.9, to get:

Lemma 5.14 (Analogue of Lemma 5.10)**.** Let $\mathcal{L}$ be some set, $r \in \{0,1\}^{np}$ and $r^{\|} = r_1^{\|} \times \ldots \times r_p^{\|}$, where each $r_i^{\|}$ is the affine space parallel to r_i and none of the r_i is an all-0 or all-1 vector. Let $p \leq \frac{1}{40} 2^{n/100}$, $D \subseteq \mathcal{L} \times r^{\|}$ with $\beta = \frac{|D|}{|\mathcal{L} \times r^{\|}|} \geq 32 \cdot 2^{-n/4}$. Now pick a random $V = V_1 \times \ldots \times V_p$ where each V_i is picked independently as in Proposition 5.13. Then

$$\Pr\left[\frac{|D \cap U|}{|U|} = \beta(1 \pm 2^{-n/30})\right] \geq 1 - \frac{1}{2^{n/30}}.$$

5.4.3 Proof of the extension lemma

Proof of Lemma 5.7. Let us look at the following bipartite graph: on the left side we have the different V's, and on the right side we have the different $r \in \mathsf{Ext}$. We put an edge between V and r whenever $r \in V$. Let E be the number of edges, R be the number of r's, and S be the number of V's.

From Lemma 5.9, it follows that for at least $1 - 2^{-n/25}$ fraction of the V, $\deg(V) \in (1 \pm 2^{-n/25})\alpha|V|$. Hence $E \geq (1 - 2^{-n/25})S\alpha|V|$.

Now notice the following **observation:** Picking a uniformly random neighbor V of r and then taking $V^\perp$ is the same as picking a uniformly random subspace $V^\perp$ of $r^\perp$ of dimension $\lceil \frac{n}{2} \rceil$.

Call an edge (r, V) *good* if (2) holds. Then from Lemma 5.10 it follows that for every r at least a $1 - 2^{-n/30}$ fraction of its edges are good. So the total number of good edges is at least $(1 - 2^{-n/30})E$.

If we remove all the edges (r, V) from V's for which $\deg(V) \notin (1 \pm 2^{-n/25})\alpha R$, then because every V has at most $|V|$ edges, we remove no more than

$$2^{-n/25} S|V| \leq \frac{2 \cdot 2^{-n/25}}{\alpha} E \leq \frac{2 \cdot 2^{-n/25}}{8 \cdot 2^{-n/30}} E \leq \frac{1}{4} 2^{-\frac{n}{150}} E$$

edges. Hence after removing these edges, the total number of good edges is still at least $(1 - 2^{-n/30} - \frac{1}{4} 2^{-\frac{n}{150}})E \geq (1 - \frac{1}{2} 2^{-\frac{n}{150}})E$. If E' is the number of surviving edges, then we still have $(1 - \frac{1}{2} 2^{-\frac{n}{150}})E'$ good edges. Now notice that every surviving V has $(1 \pm 2^{-n/25})\alpha|V|$ edges, and so if S' is the number of surviving V's, then $E' \leq (1 + 2^{-n/25})\alpha|V|S'$, and the number of V's without good edges is at most

$$\frac{\frac{1}{2} 2^{-\frac{n}{150}} E'}{(1 - 2^{-n/25})\alpha|V|} \leq \frac{1}{2} 2^{-\frac{n}{150}} \frac{1 + 2^{-n/25}}{1 - 2^{-n/25}} S' \leq \frac{2}{3} 2^{-\frac{n}{150}} S.$$

In total, the number of V's we removed, plus the number of V's without good edges, is less than $2^{-\frac{n}{150}} S$. Any other V will have at least one neighbor r for which both conditions (1) and (2) must hold. $\square$

With greater care, we would have been able to prove that the fraction of $r \in \mathsf{Ext} \cap V$ with property (2) is $(1 \pm o(1)) \cdot \alpha$.

5.4.4 The 1-monochromatic extension lemma

We will also need the following 1-monochromatic analogue:

Lemma 5.15 (1-monochromatic extension lemma)**.** Let p and n be sufficiently large natural numbers, such that $p \leq \frac{1}{40} \cdot 2^{\frac{n}{100}}$. Let $\mathcal{O}$ be some finite set.

- *Dense set of extensions.* Let $\mathsf{Ext} \subseteq \{0,1\}^{np}$ with $\alpha = \frac{|\mathsf{Ext}|}{2^{np}} \geq 10 \cdot 2^{-n/30}$.
- *Associated set.* Suppose that to each $r \in \mathsf{Ext}$ corresponds a set $T_r \subseteq r^{\|} \times \mathcal{O}$, where

$$r^{\|} = r_1^{\|} \times \ldots \times r_p^{\|} = \{r' \in \{0,1\}^{np} \mid \mathsf{IP}_n^p(r, r') = 1^p\}.$$

- *Quality.* Define the quality of r to be

$$q(r) \triangleq \frac{|T_r|}{|r^{\|} \times \mathcal{O}|},$$

and suppose that $q(r) \geq 2^{-n/10}$ for every $r \in \mathsf{Ext}$.

- 1-*monochromatic rectangle.* Now pick a random product $V = V_1 \times \cdots \times V_p$, where each V_i is of the form $a_i + W_i$, where a_i is a random odd-hamming-weight string and W_i is a uniformly random subspace of $a_i^\perp$ of dimension $\lfloor \frac{n-1}{2} \rfloor$. Let $V^{(1)}$ denote $V_1^{(1)} \times \cdots \times V_p^{(1)}$, for $V_i^{(1)} = a_i + W_i^\perp$ — where $W_i^\perp$ is the orthogonal complement of W_i within $a_i^\perp$.
- *Quality in the monochromatic rectangle.* Finally, define

$$q_V(r) = \frac{|T_r \cap (V^{(1)} \times \mathcal{O})|}{|V^{(1)} \times \mathcal{O}|}.$$

Conclusion. Then with probability $\geq 1 - 2^{-\frac{n}{150}}$ over the choice of V, there is some extension $r \in \mathsf{Ext} \cap V$ whose quality is preserved in the 0-monochromatic rectangle:

$$\exists r \in \mathsf{Ext} \text{ such that } \begin{cases} r \in V & (1) \\ q_V(r) = (1 \pm 2^{-n/30}) \cdot q(r) & (2) \end{cases}$$

Proof. This proof is exactly the same proof as in Lemma 5.7, but we apply Lemmas 5.12 and 5.14 instead of Lemmas 5.9 and 5.10, respectively. The **observation** stated in the proof is also replaced, as follows. The left-nodes in the graph are now of the form $V = (a_1+W_1) \times \ldots \times (a_p+W_p)$. If V is a neighbor of r, i.e. if $r_i \in a_i+W_i$ for all i, then it also holds for all i that $a_i \in r^{\|}$ (i.e. $\langle r_i, a_i \rangle = 1$), and that $W_i^{\perp}$ is in $\{a_i, r_i\}^{\perp}$ (because for any $x \in W_i^{\perp}$, we have $\langle x, r_i \rangle = \langle x, a_i + w_i \rangle = 0$ for some $w_i \in W_i$).

It then holds that picking a uniformly random neighbor of r and then taking $V^{\perp}$ is the same as picking the a_i of odd hamming weight uniformly at random from $r_i^{\|}$ and then taking a uniformly random $\lceil \frac{n-1}{2} \rceil$-dimensional subspace $W_i^{\perp}$ within $\{a_i, r_i\}^{\perp}$. □

5.5 Proof of the Zooming-in lemma

We will now prove a stronger version of the Zooming-in lemma presented in Section 5.3.4.

Lemma 5.16 (Zooming-in lemma, prefix side, strong version)**.** Let $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$, where $p = p_1 + p_2 \leq \frac{1}{40} \cdot 2^{n/100}$ is a sufficiently large natural number. Suppose we have a rectangle $A \times B$, where both A and B are subsets of $\mathcal{L} \times \mathcal{R}$, and a C-bit protocol $\pi : A \times B \to [p]$. Let μ_1 be a uniform distribution over the strings 1^i0^{p-i} for $i \in [p_1]$, and let λ_1 be the lifting of μ_1 to $(\mathcal{L} \times \mathcal{R})^2$ with respect to IP_n^p (as in Definition 4.2). Let $q_1(a)$ denote the row-quality with respect to μ_1 (and OS_p, IP_n^p, A, B and π). For a given $\ell \in A_{\leq p_1}$, denote by $\mathsf{Ext}(\ell) = \mathsf{Ext}_A^{[p]\setminus[p_1]}(\ell)$ the set of extensions of ℓ, and define the *min-quality* $q_{\min}(\ell)$ to be the minimum q_1 of ℓ's extensions:

$$q_{\min}(\ell) \stackrel{\text{def}}{=} \min_{r \in \mathsf{Ext}(\ell)} q_1(\ell \times r).$$

- *Enough density.* Suppose that A has prefix-side density $\alpha_{\leq p_1} \stackrel{\text{def}}{=} \frac{|A_{\leq p_1}|}{|\mathcal{L}|}$ at least $2^{-\frac{n}{10}}$, and for each $\ell \in A_{\leq p_1}$, the density of its extensions $\frac{|\mathsf{Ext}(\ell)|}{|\mathcal{R}|}$ is at least $8 \cdot 2^{-\frac{n}{30}}$; suppose also that the density of B, $\beta \stackrel{\text{def}}{=} \frac{|B|}{|\mathcal{L} \times \mathcal{R}|}$, is at least $2^{-\frac{n}{200}}$.
- *Average min-quality condition.* Finally, suppose that the average $q_{\min}(\ell)$ is bounded by

$$\frac{1}{|A_{\leq p_1}|} \sum_{\ell \in A_{\leq p_1}} q_{\min}(\ell) \geq \tilde{\gamma}\beta,$$

for some value $\tilde{\gamma} \geq 2^{-\frac{n}{1200}}$.

- *Conclusion.* Then Protocol$(n, p_1, (1-\delta)\alpha_{\leq p_1}, (1-\delta)\beta, (1-\delta)\tilde{\gamma}, C)$ holds, where $\delta = 8 \cdot 2^{-\frac{n}{1200}}$.

This is a stronger statement than Lemma 5.5 of Section 5.3.4, because when $q_1(a) \geq \tilde{\gamma}\beta$ for every $a \in A$, obviously $q_{\min}(\ell) \geq \tilde{\gamma}\beta$ for every $\ell \in A_{\leq p_1}$.

Proof of the lemma. Fix any $\ell \in A_{\leq p_1}$. To each $r \in \mathsf{Ext}(\ell)$ corresponds the set $O_{\ell \times r}^{\leq p_1} = O_\ell \times r^{\perp}$ of possible inputs $\ell' \times r'$ of Bob for which $\mathsf{OS} \circ \mathsf{IP}(\ell \times r, \ell' \times r') \leq p_1$:

$$O_\ell = \{\ell' \in \{0,1\}^{p_1 n} \mid \exists i \in [p_1] \ \mathsf{IP}_n^{p_1}(\ell, \ell') = 1^i 0^{p_1 - i}\}$$

$$r^{\perp} = \{r' \in \{0,1\}^{p_2 n} \mid \mathsf{IP}_n^{p_2}(r, r') = 0^{p_2}\}.$$

Let us apply the extension lemma (Lemma 5.7) to each $\ell \in A_{\leq p_1}$, with $\mathsf{Ext} = \mathsf{Ext}(\ell)$, $\mathcal{O} = O_\ell$, and with $T_r = T^{\leq p_1}_{\ell \times r B} \subseteq O_\ell \times r^\perp$ being the subset of Bob's inputs $b \in O_\ell \times r^\perp$ such that $\pi(\ell \times r, b) = \mathsf{OS}_p \circ \mathsf{IP}^p_n(\ell \times r, b)$. In this case $q(r)$, as defined in the statement of Lemma 5.7, is exactly:

$$q(r) \stackrel{\text{def}}{=} \frac{|T_r|}{|O_\ell \times r^\perp|} \stackrel{\text{def}}{=} \frac{|T^{\leq p_1}_{\ell \times r B}|}{|O^{\leq p_1}_{\ell \times r}|} = \frac{\lambda_1(T^{\leq p_1}_{\ell \times r B})}{\lambda_1(O^{\leq p_1}_{\ell \times r})} \stackrel{\text{def}}{=} q_1(\ell \times r)$$

(the before-to-last equality follows as in Section 5.1.1); clearly $q(r)$ will always be greater than $q_{\min}(\ell)$. Let us define $q_V(\ell \times r)$ to be $q_V(r)$ as defined in Lemma 5.7, when applied to ℓ, i.e.:

$$q_V(\ell \times r) = \frac{|T^{\leq p_1}_{\ell \times r B} \cap (O_\ell \times V^\perp)|}{|O_\ell \times V^\perp|}.$$

Let $V = V_1 \times \cdots \times V_p$, where each V_i is an independent and uniformly random $\lfloor \frac{n}{2} \rfloor$-dimensional random subspace of $\{0,1\}^n$. Let $V^\perp$ denote $V_1^\perp \times \cdots \times V_p^\perp$. The extension lemma then says that for any such $\ell \in A_{\leq p_1}$ a random V will, with probability at least $1 - 2^{-\frac{n}{150}}$, give us a suffix-side extension $r \in \mathsf{Ext}(\ell) \cap V$ with

$$q_V(\ell \times r) > (1 - 2^{-\frac{n}{30}}) \cdot q_{\min}(\ell). \tag{I}$$

Since B is a large set, then Lemma 5.9 says that, with probability $\geq 1 - 2^{-n/25}$, it will also hold:

$$\beta_V \stackrel{\text{def}}{=} \frac{|B \cap (\mathcal{L} \times V^\perp)|}{|\mathcal{L} \times V^\perp|} \in (1 \pm 2^{-n/25}) \cdot \beta. \tag{II}$$

Therefore, we may fix a single V which satisfies (II), and which satisfies (I) for a $1 - 2 \cdot 2^{-\frac{n}{150}}$ fraction of all the $\ell \in A_{\leq p_1}$. After fixing such a V, let A'_{L} be the set of ℓ for which (I) holds, and let $B_V = B \cap (\mathcal{L} \times V^\perp)$. Associate with each $\ell \in A'_{\mathrm{L}}$ the promised string r. From the average min-quality condition we may now derive:

$$\frac{1}{|A'_{\mathrm{L}}|} \sum_{\ell \in A'_{\mathrm{L}}} q_V(\ell \times r) \geq \frac{1}{|A_{\leq p_1}|} \sum_{\ell \in A'_{\mathrm{L}}} (1 - 2^{-\frac{n}{30}}) \cdot q_{\min}(\ell \times r) \geq (1 - 2^{-\frac{n}{30}}) \tilde{\gamma} \beta - 2 \cdot 2^{-\frac{n}{150}}.$$

Together with (II) and our bounds on $\tilde{\gamma}$ and β (specifically $\tilde{\gamma} \geq 2^{-\frac{n}{1200}}$ and $\beta \geq 2^{-\frac{n}{200}}$), this implies:

$$\frac{1}{|A'_{\mathrm{L}}|} \sum_{\ell \in A'_{\mathrm{L}}} q_V(\ell \times r) \geq \left(1 - 2^{-\frac{n}{30}} - \frac{2 \cdot 2^{-\frac{n}{150}}}{\tilde{\gamma}\beta}\right) \cdot \frac{1}{1 + 2^{-\frac{n}{25}}} \cdot \tilde{\gamma} \beta_V \geq (1 - 5 \cdot 2^{-\frac{n}{1200}}) \tilde{\gamma} \beta_V.$$

What does $q_V(\ell \times r)$ mean? Let $\mathcal{A}' = \mathcal{L}$ and $\mathcal{B}_V = \mathcal{L} \times V^\perp$, and set $G_V : \mathcal{A}' \times \mathcal{B}_V \to [p]$ to $G_V(\ell, \ell' \times r') = \mathsf{IP}^p_n(\ell \times r, \ell' \times r') = \mathsf{IP}^{p_1}_n(\ell, \ell')\, 0^{p_2}$. Define a protocol $\pi_V : A'_{\mathrm{L}} \times B_V \to [p]$ to work as follows: Alice is given $\ell \in A'_{\mathrm{L}}$, and Bob is given $b \in B_V$; Alice extends ℓ with the string $r \in \mathsf{Ext}(\ell) \cap V$ that testifies (I); then Alice and Bob run π on $\ell \times r$ and b. Let $\lambda'_1 = \lambda_{\mathcal{A}' \times \mathcal{B}_V, G_V}$ be the lifting of μ_1 to $\mathcal{A}' \times \mathcal{B}_V$, with respect to G_V. Finally, let $T_{\ell B_V}$ be the set of those $\ell' \times r' \in B_V$ for which $\pi_V(\ell, \ell' \times r') = \mathsf{OS}_p \circ G_V(\ell, \ell' \times r')$. It now follows (as in Section 5.1.1) that $q_V(\ell \times r)$ equals:

$$q_V(\ell \times r) \stackrel{\text{def}}{=} \frac{|T^{\leq p_1}_{\ell \times r B} \cap (O_\ell \times V^\perp)|}{|O_\ell \times V^\perp|} = \frac{|T_{\ell B_V}|}{|O_\ell \times V^\perp|} = \frac{\lambda'_1(T_{\ell B_V})}{\lambda'_1(O_\ell \times V^\perp)},$$

which is exactly the row-quality of ℓ (Definition 4.6) w.r.t $\mathsf{OS}_p, G_V, \mu_1, A'_{\mathrm{L}}, B_V$ and π_V.

By the quality–success correspondence (Lemma 4.7), it then also follows that the success probability of π_V in $A'_{\mathrm{L}} \times B_V$, w.r.t λ'_1, is at least $(1 - 6 \cdot 2^{-\frac{n}{1200}})\tilde{\gamma}$.[4]

[4]To explain this in terms of the prefix-side projection of B_V, what is happening here is as if Bob gets a string $b' \in (B_V)_{\leq p_1}$ with probability $\frac{|\mathsf{Ext}(b')|}{|B_V|}$, i.e. *weighted according to the number of extensions*, and then runs the protocol on the string $b = b'b''$, where b'' is a uniformly chosen string in $\mathsf{Ext}(b')$ (which we can think of as Bob's private randomness).

We may now apply the quality–success correspondence (Lemma 4.7) on Bob's side. For a string $b \in B_V$, define its quality to be

$$q(b) \stackrel{\text{def}}{=} \frac{\lambda'_1(T_{A'_L b})}{\lambda'_1(A'_L \times B_V)} = \frac{|T_{A'_L b}|}{|O_b|},$$

where

$$O_b = \{\ell \in \mathcal{L} \mid \exists i \in [p_1] G_V(\ell, b) = 1^i 0^{p-i}\} \quad T_{A'_L b} = \{\ell \in A'_L \cap O_b \mid \pi_V(\ell, b) = \mathsf{OS}_p \circ G_V(\ell, b)\}.$$

Then Lemma 4.7 implies that

$$\frac{1}{|B_V|} \sum_{b \in B_V} q(b) \geq (1 - 7 \cdot 2^{-\frac{n}{1200}}) \cdot \tilde{\gamma}\alpha'_L$$

Now for a prefix-side of Bob $\ell' \in (B_V)_{\leq p_1}$, let $\mathsf{Ext}(\ell') = \mathsf{Ext}_{B_V}^{[p] \setminus [p_1]}(\ell')$ be the set of r' with $\ell' \times r' \in B_V$. Define $q_{\text{avg}}(\ell') = \frac{1}{|\mathsf{Ext}(\ell')|} \sum_{r' \in \mathsf{Ext}(\ell')} q(\ell' \times r')$, so that

$$\sum_{\ell' \in (B_V)_{\leq p_1}} \frac{|\mathsf{Ext}(\ell')|}{|B_V|} \cdot q_{\text{avg}}(\ell') \geq (1 - 7 \cdot 2^{-\frac{n}{1200}}) \cdot \tilde{\gamma}\alpha'_L$$

Then by Lemma 2.7, there exists a set $B'_L \subseteq (B_V)_{\leq p_1}$, of size $\lfloor \beta_V |\mathcal{L}| \rfloor$, such that

$$\frac{1}{|B'_L|} \sum_{\ell' \in B'_L} q_{\text{avg}}(\ell') \geq (1 - 7 \cdot 2^{-\frac{n}{1200}}) \cdot \tilde{\gamma}\alpha'_L.$$

What does the $q_{\text{avg}}(\ell')$ mean? Let μ' be the uniform distribution over the strings $1^i 0^{p_1 - i}$ for all $i \in [p_1]$. Let $\pi' : \mathcal{L} \times \mathcal{L} \to [p_1]$ be the following protocol: Alice and Bob get inputs $(\ell, \ell') \in A'_L \times B'_L$; Bob chooses some input $r' \in \mathsf{Ext}(\ell')$ such that $q(\ell' \times r') \geq q_{\text{avg}}(\ell')$, and then they play π on $\ell \times r$ and $\ell' \times r'$, where r was the input satisfying (I). Then $q(\ell' \times r')$ is exactly the *column*-quality of ℓ' with respect to OS_{p_1}, $\mathsf{IP}_n^{p_1}$, μ', A'_L, B'_L and π', and the average over B'_L is also $\geq (1 - 7 \cdot 2^{-\frac{n}{1200}}) \cdot \tilde{\gamma}\alpha'_L$.

It then follows again from the quality–success correspondence (Lemma 4.7) that π' has success probability $\geq (1 - 8 \cdot 2^{-\frac{n}{1200}})\tilde{\gamma}$ with respect to OS_{p_1}, $\mathsf{IP}_n^{p_1}$ and μ' (i.e. on the distribution λ' lifted from μ'). As shown above, the density of A'_L in $\mathcal{L}$ is $\geq (1 - 2 \cdot 2^{-\frac{n}{150}})\alpha_{\leq p_1}$, and the density of B'_L in $\mathcal{L}$ is $\geq (1 - 2^{-n/30})\beta$.[5] This concludes the proof of the lemma. □

It is to be noted that the Zooming-in lemma is symmetric with respect to p_1 and p_2, i.e., a similar argument will prove the following lemma:

Lemma 5.17 (Zooming-in lemma, suffix side, strong version)**.** Let $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$, where $p = p_1 + p_2 \leq \frac{1}{40} \cdot 2^{n/100}$ is a sufficiently large natural number. Suppose we have a rectangle $A \times B$, where both A and B are subsets of $\mathcal{L} \times \mathcal{R}$, and a C-bit protocol $\pi : A \times B \to [p]$. Let μ_2 be a uniform distribution over the strings $1^i 0^{p-i}$ for $i \in [p] \setminus [p_1]$, and let λ_2 be the lifting of μ_2 to $(\mathcal{L} \times \mathcal{R})^2$ with respect to IP_n^p (as in Definition 4.2). Let $q_2(a)$ denote the row-quality with respect to μ_2 (and OS_p, IP_n^p, A, B and π). $q_{\min}(\ell)$ is defined as in Lemma 5.16 but with respect to the extensions in the prefix-side.

- *Enough density.* Suppose that A has suffix-side density $\alpha_{>p_1}$ at least $2^{-\frac{n}{10}}$, and for each $r \in A_{>p_1}$, the density of its extensions $\frac{|\mathsf{Ext}(r)|}{|\mathcal{L}|}$ is at least $8 \cdot 2^{-\frac{n}{30}}$; suppose also that the density of B, $\beta \stackrel{\text{def}}{=} \frac{|B|}{|\mathcal{L} \times \mathcal{R}|}$, is at least $2^{-\frac{n}{200}}$.

[5] The loss from $2^{-\frac{n}{25}}$ is just a rough way of accounting for the floor.

- *Average min-quality condition.* Finally, suppose that the average $q_{\min}(r)$ is bounded by

$$\frac{1}{|A_{>p_1}|}\sum_{r\in A_{>p_1}} q_{\min}(r) \geq \tilde{\gamma}\beta,$$

for some value $\tilde{\gamma} \geq 2^{-\frac{n}{1200}}$.

- *Conclusion.* Then Protocol$(n, p_2, (1-\delta)\alpha_{>p_1}, (1-\delta)\beta, (1-\delta)\tilde{\gamma}, C)$ holds, where $\delta = 8\cdot 2^{-\frac{n}{1200}}$.

5.6 Proof of the Min-quality lemma

We first restate the lemma for convenience. As stated before, the lemma appears implicitly in [RW89], and the ideas of the proof below are all taken from that paper.

Lemma 5.18 (Min-quality lemma)**.** Let $\mathcal{L} = \{0,1\}^{np_1}$ and $\mathcal{R} = \{0,1\}^{np_2}$ for some sufficiently large natural numbers n, p_1 and p_2. Suppose we have a rectangle $A' \times B$, where both A' and B are subsets of $\mathcal{L}\times\mathcal{R}$, and a protocol $\pi : A' \times B \to [p]$. Let μ_1 be uniform over the strings $1^i 0^{p-i}$ for $i \in [p_1]$, and let $q_1(a)$ denote the row-quality (Definition 4.6) with respect to μ_1 (and OS, IP^p_n, A', B and π). If we have fixed a subset $A'' \subseteq A$, then for any given $\ell \in A''_{\leq p_1}$ let $\mathsf{Ext}''(\ell) = \mathsf{Ext}^{[p]\setminus[p_1]}_{A''}(\ell)$ be the set of extensions $r \in \mathcal{R}$ with $\ell \times r \in A''$, and define the *min-quality* of ℓ, $q''_{\min}(\ell)$, to be the minimum q_1 of its extensions:

$$q''_{\min}(\ell) \overset{\text{def}}{=} \min_{r\in\mathsf{Ext}''(\ell)} q_1(\ell\times r).$$

Now suppose we have the following properties:

- *A' and B have good density.* $\alpha' \overset{\text{def}}{=} \frac{|A'|}{|\mathcal{L}\times\mathcal{R}|} \geq 2^{-\frac{n}{200}}$, and $\beta \overset{\text{def}}{=} \frac{|B|}{|\mathcal{L}\times\mathcal{R}|} \geq 2^{-\frac{n}{200}}$.

- *Average quality is high.* For some value $Q \geq 2\cdot 2^{-\frac{n}{150}}$ it holds that:

$$\frac{1}{|A'|}\sum_{a\in A'} q_1(a) \geq Q.$$

Conclusion. Then there is a subset $A'' \subseteq A'$ with the following properties:

- A'' **has enough density.** The size of the prefix-side projection is $|A''_{\leq p_1}| \geq \lfloor (1-2^{-\frac{n}{120}})\cdot\alpha'|\mathcal{L}|\rfloor$, and for all $\ell \in A''_{\leq p_1}$ we have $|\mathsf{Ext}''(\ell)| \geq 8\cdot 2^{-\frac{n}{30}}|\mathcal{L}|$;

- A'' **obeys the average min-quality condition.** The average min-quality over $\ell \in A''_{\leq p_1}$ almost matches the average quality in A':

$$\frac{1}{|A''_{\leq p_1}|}\sum_{\ell\in A''_{\leq p_1}} q''_{\min}(\ell) \geq (1-4\cdot 2^{-\frac{n}{300}})\cdot Q.$$

5.6.1 Notation

We will start with the set A' and successively remove strings from it, thus obtaining sets $A' \supset A^{(1)} \supset A^{(2)} \supset A''$. For some set $A^* \subseteq \mathcal{L}\times\mathcal{R}$ (the notation $*$ is one of $'$, $^{(1)}$, $^{(2)}$, $''$), let $A^*_{\mathrm{L}} = A^*_{\leq p_1}$ be the projection of A^* onto its prefix side; for every $\ell \in A^*_{\mathrm{L}}$, let $\mathsf{Ext}^*(\ell)$ be the set of $r \in \mathcal{R}$ with $\ell\times r \in A^*$, and define ℓ's *average-quality* $q^*_{\mathrm{avg}}(\ell)$ to be the average prefix-side quality in $\mathsf{Ext}^*(\ell)$:

$$q^*_{\mathrm{avg}}(\ell) \overset{\text{def}}{=} \frac{1}{|\mathsf{Ext}^*(\ell)|}\sum_{r\in\mathsf{Ext}^*(\ell)} q_1(\ell\times r).$$

5.6.2 Pruning ℓ with few extensions.

We first discard all prefix-sides having a small number of suffix-side extensions — discarding the set:

$$A^{\mathsf{discard}} \stackrel{\text{def}}{=} \left\{ a = \ell \times r \in A' \;\middle|\; |\mathsf{Ext}'(\ell)| < 2^{-n/50} \cdot 2^{np_2} \right\}$$

Let $A^{(1)} = A \setminus A^{\mathsf{discard}}$. Notice that we are leaving some leverage room — we preserve only those ℓ having at least $2^{-n/50} \cdot 2^{np_2}$ extensions, but only $8 \cdot 2^{-n/30} \cdot 2^{np_2}$ are needed by the *enough density* condition. This is so that we can remove more extensions later, and still have enough.

Let us calculate the amount of quality that was lost. By our promise on Q and α', we have

$$\frac{1}{2} \cdot 2^{-n/120} Q|A'| \geq 2^{-\frac{n}{120} - \frac{n}{150} - \frac{n}{200}} \cdot 2^{np} \geq 2^{-n/50} \cdot 2^{np} \geq |A^{\mathsf{discard}}|$$

Then $\alpha^{(1)} \geq (1 - \frac{1}{2} \cdot 2^{-\frac{n}{120}}) \cdot \alpha'$, and even if all discarded a have $q_1(a) = 1$, we still have

$$\frac{1}{|A^{(1)}|} \sum_{a \in A^{(1)}} q_1(a) \geq \left(1 - \frac{1}{2} \cdot 2^{-\frac{n}{120}}\right) \cdot Q \geq \left(1 - 2^{-\frac{n}{120}}\right) \cdot Q. \tag{$\dagger$}$$

5.6.3 From weighted average to uniform average

We may rewrite ($\dagger$) as:

$$\sum_{\ell \in A_{\mathrm{L}}^{(1)}} \frac{|\mathsf{Ext}^{(1)}(\ell)|}{|A^{(1)}|} \cdot q_{\mathrm{avg}}^{(1)}(\ell) \geq (1 - 2^{-\frac{n}{120}}) \cdot Q.$$

Then by Lemma 2.7, there must exist a set $A_{\mathrm{L}}^{(2)} \subseteq A_{\mathrm{L}}^{(1)}$, with $|A_{\mathrm{L}}^{(2)}| \geq \lfloor \alpha^{(1)} \mathcal{L} \rfloor$, and such that

$$\frac{1}{|A_{\mathrm{L}}^{(2)}|} \sum_{\ell \in A_{\mathrm{L}}^{(2)}} q_{\mathrm{avg}}^{(1)}(\ell) \geq (1 - 2^{-\frac{n}{120}}) \cdot Q.$$

We then set $A^{(2)} = \{\ell \times r \mid \ell \in A_{\mathrm{L}}^{(2)} \text{ and } r \in \mathsf{Ext}^{(1)}(\ell)\}$, so that $\mathsf{Ext}^{(2)}(\ell) = \mathsf{Ext}^{(1)}(\ell)$ for every $\ell \in A_{\mathrm{L}}^{(2)}$, and $q_{\mathrm{avg}}^{(2)}(\ell) = q_{\mathrm{avg}}^{(1)}(\ell)$. It then holds that $\alpha_{\mathrm{L}}^{(2)} \geq (1 - 2^{-\frac{n}{120}}) \cdot \alpha'$, and every $\ell \in A_{\mathrm{L}}^{(2)}$ has $2^{-n/50} \cdot 2^{np_2}$ extensions. We have thus concluded that $A^{(2)}$ has *good enough* density, and that the average $q_{\mathrm{avg}}^{(2)}(\ell)$ is high enough; we will prune some more to make the average $q_{\min}(\ell)$ is high enough.

5.6.4 Forcing high min-quality

Let us *ignore* the set of $\ell \times r \in A^{(2)}$ with $q^{(2)}(\ell)$ significantly less than the average $q^{(2)}(\ell)$; i.e., we ignore the set

$$A^{\mathsf{ignore}} \stackrel{\text{def}}{=} \left\{ \ell \times r \in A^{(2)} \;\middle|\; q_{\mathrm{avg}}^{(2)}(\ell) < 2^{-\frac{n}{300}} \cdot Q \right\}.$$

Among those $\ell \times r$ which we *didn't* ignore, let us *discard* from $A^{(2)}$ those for which $q_1(\ell \times r)$ fails to be close enough to $q_{\mathrm{avg}}^{(2)}(\ell)$:

$$A^{\mathsf{discard}} = \left\{ \ell \times r \in A^{(2)} \setminus A^{\mathsf{ignore}} \;\middle|\; q_1(\ell \times r) < (1 - \varepsilon) \cdot q_{\mathrm{avg}}^{(2)}(l) \right\}$$

We will set ε later. The promised set A'' is exactly $A^{(2)} \setminus A^{\mathsf{discard}}$. I.e., we keep every ignored prefix-side and its extensions, and for each non-ignored $\ell \in A_{\mathrm{L}}^{(2)}$, we keep the set $\mathsf{Ext}''(\ell) \subseteq \mathsf{Ext}^{(2)}(\ell)$ of suffix-side extensions which attain $(1 - \varepsilon)$ of the average quality (average among the suffix-side extensions of ℓ in $A^{(2)}$). It is easy to see that, in order to attain the average $q^{(2)}_{\text{avg}}(\ell)$, the number of surviving extensions must obey:

$$|\mathsf{Ext}''(\ell)| \geq \varepsilon \cdot q^{(2)}_{\text{avg}}(\ell) \cdot |\mathsf{Ext}^{(2)}(\ell)|.$$

For a non-ignored prefix-side ℓ, $q^{(2)}_{\text{avg}}(\ell)$ is $\geq 2^{-\frac{n}{300}} \cdot Q$, and we have chosen $\mathsf{Ext}^{(2)}(\ell) = \mathsf{Ext}^{(1)}(\ell)$ to have size at least $2^{-n/50} \cdot 2^{np_2}$. So, taking $\varepsilon = 8 \cdot 2^{-\frac{n}{300}}$, we can conclude that:

$$|\mathsf{Ext}''(\ell)| \geq \underbrace{8 \cdot 2^{-\frac{n}{300}}}_{\varepsilon} \cdot \underbrace{2^{-\frac{n}{300}} \cdot 2^{-\frac{n}{150}}}_{q^{(2)}_{\text{avg}}(\ell)} \cdot \underbrace{2^{-\frac{n}{50}} \cdot 2^{np_2}}_{|\mathsf{Ext}^{(2)}(\ell)|} = 8 \cdot 2^{-n/30} \cdot 2^{np_2};$$

note also that $A''_{\text{L}} = A^{(2)}_{\text{L}}$, and so $\alpha''_{\text{L}} \geq (1 - 2^{-\frac{n}{120}}) \cdot \alpha'$ — this shows that A'' has enough density. Also, $q(\ell \times r) \geq (1 - 2^{-\frac{n}{300}}) q^{(2)}_{\text{avg}}(\ell)$ for every non-ignored prefix-side ℓ. It then follows that A obeys the average min-quality condition:

$$\begin{aligned}\frac{1}{|A''_{\text{L}}|} \sum_{\ell \in A''_{\text{L}}} q''_{\min}(\ell) &\geq \frac{1}{|A^{(2)}_{\text{L}}|} \sum_{\ell \in A^{(2)}_{\text{L}}} (1 - 2^{-\frac{n}{300}}) q''_{\text{avg}}(\ell) \; - \; 2^{-\frac{n}{300}} \cdot Q \\ &\geq \Big((1 - 2^{-\frac{n}{300}})(1 - 2^{-\frac{n}{120}}) - 2^{-n/300}\Big) \cdot Q \; \geq (1 - 4 \cdot 2^{-\frac{n}{300}}) \cdot Q. \qquad \square\end{aligned}$$

Acknowledgement

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 is partially supported by a Ramanujan Fellowship of the DST, India and the last author is partially supported by a TCS fellowship.

References

[ABB+16] Anurag Anshu, Aleksandrs Belovs, Shalev Ben-David, Mika Göös, Rahul Jain, Robin Kothari, Troy Lee, and Miklos Santha. Separations in communication complexity using cheat sheets and information complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:72, 2016.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald De Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013. 0.

[BBK+13] Joshua Brody, Harry Buhrman, Michal Kouckỳ, Bruno Loff, Florian Speelman, and Nikolay Vereshchagin. Towards a reverse newman's theorem in interactive information complexity. In *Proceedings of the 28th CCC*, pages 24–33, 2013. 0.

[BDW02] Harry Buhrman and Ronald De Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[BPSW05] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and the multiparty nof communication complexity of set disjointness. In *Proceedings of the 20th CCC*, pages 52–66, 2005.

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014.

[BRWY13a] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In *Proceedings of the 40th ICALP*, pages 232–243, 2013.

[BRWY13b] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *Proceedings of the 54th FOCS*, pages 746–755, 2013.

[Buh16] Harry Buhrman, 2016. Private communication.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. Technical Report TR08-002, Electronic Colloquium on Computational Complexity (ECCC), 2008.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[Cha07] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the 48th FOCS*, pages 449–458, 2007.

[Cha09] Arkadev Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, McGill University, 2009.

[CL08] Andrew M Childs and Troy Lee. Optimal quantum adversary lower bounds for ordered search. In *Proceedings of the 25th ICALP*, pages 869–880. Springer, 2008.

[CrK+16] Arkadev Chattopadhyay, Pavel Dvořák, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Lower bounds for elimination via weak regularity. Technical Report TR16-165, Electronic Colloquium on Computational Complexity (ECCC), 2016.

[dRNV16] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication. In *Proceedings of the 56th FOCS*, 2016.

[Dru12] Andrew Drucker. Improved direct product theorems for randomized query complexity. *Computational Complexity*, 21(2):197–244, 2012.

[GJ16] Mika Göös and TS Jayram. A composition theorem for conical juntas. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 50. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[GLM+15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the 47th STOC*, pages 257–266. ACM, 2015.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th FOCS*, 2015.

[HJMR07] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *Proceedings of the 22nd CCC*, pages 10–23, 2007.

[HNS02] Peter Høyer, Jan Neerbek, and Yaoyun Shi. Quantum complexities of ordered searching, sorting, and element distinctness. *Algorithmica*, 34(4):429–448, 2002.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th FOCS*, pages 538–545, 1995.

[Jai15] Rahul Jain. New strong direct product results in communication complexity. *Journal of the ACM*, 62(3):20, 2015.

[JKN08] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the 40th STOC*, pages 599–608, 2008.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53rd FOCS*, pages 167–176, 2012.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the 20th ICALP*, pages 300–315, 2003.

[JY12] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. Technical report, arXiv:1209.0263, 2012.

[KLL+15] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[Lev87] Leonid A. Levin. One-way functions and pseudorandom generators. *Combinatorica*, 7(4):357–363, 1987.

[LSS08] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd CCC*, pages 71–80, 2008.

[LZ10] Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 27th ICALP*, pages 475–489. Springer, 2010.

[Nor16] Jakob Nordström, 2016. Private communication.

[Pan12] Denis Pankratov. *Direct sum questions in classical communication complexity.* PhD thesis, Masters thesis, University of Chicago, 2012.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[RW89] Ran Raz and Avi Wigderson. Probabilistic communication complexity of boolean relations. In *Proceedings of the 39th STOC*, 1989.

[RY15] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. In *Proceedings of the 30th CCC*, pages 88–101, 2015.

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[She12a] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the 44th STOC*, pages 525–548, 2012.

[She12b] Alexander A Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

[She13] Alexander A. Sherstov. Communication lower bounds using directional derivatives. In *Proceedings of the 45th STOC*, pages 921–930, 2013.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[VW08] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h STOC*, pages 209–213, 1979.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *Proceedings of the 23rd FOCS*, pages 80–91, 1982.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il