

# On the Power of Statistical Zero Knowledge

Adam Bouland<sup>1</sup>, Lijie Chen<sup>2</sup>, Dhiraj Holden<sup>1</sup>, Justin Thaler<sup>3</sup>, and Prashant Nalini Vasudevan<sup>1</sup>

<sup>1</sup>CSAIL, Massachusetts Institute of Technology, Cambridge, MA USA

<sup>2</sup>IIS, Tsinghua University, Beijing, China

<sup>3</sup>Georgetown University, Washington, DC USA

## Abstract

We examine the power of statistical zero knowledge proofs (captured by the complexity class SZK) and their variants. First, we give the strongest known relativized evidence that SZK contains hard problems, by exhibiting an oracle relative to which SZK (indeed, even NISZK) is not contained in the class UPP, containing those problems solvable by randomized algorithms with unbounded error. This answers an open question of Watrous from 2002 [Aar]. Second, we “lift” this oracle separation to the setting of communication complexity, thereby answering a question of Göös et al. (ICALP 2016). Third, we give relativized evidence that *perfect* zero knowledge proofs (captured by the class PZK) are weaker than general zero knowledge proofs. Specifically, we exhibit oracles relative to which  $\text{SZK} \not\subseteq \text{PZK}$ ,  $\text{NISZK} \not\subseteq \text{NIPZK}$ , and  $\text{PZK} \neq \text{coPZK}$ . The first of these results answers a question raised in 1991 by Aiello and Håstad (Information and Computation), and the second answers a question of Lovett and Zhang (2016). We also describe additional applications of these results outside of structural complexity.

The technical core of our results is a stronger hardness amplification theorem for approximate degree, which roughly says that composing the gapped-majority function with any function of high approximate degree yields a function with high threshold degree.

# 1 Introduction

Zero knowledge proof systems, first introduced by Goldwasser, Micali and Rackoff [GMR89], have proven central to the study of complexity theory and cryptography. Abstractly, a zero knowledge proof is a form of interactive proof in which the verifier can efficiently simulate the honest prover on “yes” instances. Therefore, the verifier learns nothing other than whether its input is a “yes” or “no” instance.

In this work, we study *statistical* zero knowledge proofs systems. Here, “efficiently simulate” means that the verifier can, by itself, sample from a distribution which is statistically close to the distribution of the transcript of its interaction with the honest prover<sup>1</sup>. The resulting class of decision problems that have statistical zero knowledge proofs is denoted SZK. One can similarly define variants of this class, such as non-interactive statistical zero knowledge (where the proof system is non-interactive, denoted NISZK), or perfect zero knowledge (where the verifier can exactly simulate the honest prover, denoted PZK).

Many problems, some of which are not necessarily in NP, have been shown to admit SZK protocols. These include Graph Non-isomorphism, as well as problems believed to be hard on average, such as Quadratic Residuosity (which is equivalent to the discrete logarithm problem), and the Approximate Shortest Vector and Closest Vector problems in lattices [GMW91, GMR89, GG98, PV08]. Although SZK contains problems believed to be hard, it lies very low in the polynomial hierarchy (below  $AM \cap coAM$ ), and cannot contain NP-complete problems unless the polynomial hierarchy collapses [For87, AH91b, BHZ87]. Owing in part to its unusual property of containing problems believed to be hard but not NP-complete, SZK has been the subject of intense interest among complexity theorists and cryptographers.

Despite its importance, many basic questions about the hardness of SZK and its variants remain open. Our results in this work can be understood as grouped into three classes, detailed in each of the next three subsections. However, we prove these results via a unified set of techniques.

## 1.1 Group 1: Evidence for the Hardness of SZK

**Motivation.** Several cryptosystems have been based on the believed hardness of problems in SZK, most notably Quadratic Residuosity and the Approximate Shortest Vector and Closest Vector problems mentioned above. If one could solve SZK-hard problems efficiently, it would break these cryptosystems. Hence, a natural task is to show lower bounds demonstrating that problems in SZK cannot be solved easily. For example, one might want to show that quantum computers or other, more powerful models of computation cannot solve SZK-hard problems efficiently. This would provide evidence for the belief that problems in SZK are computationally hard.

Of course, proving such results unconditionally is very difficult, because SZK is contained in  $AM \cap coAM$  [For87, AH91b], so even proving lower bounds against classical algorithms solving SZK-hard problems would require separating P from NP.<sup>2</sup> Therefore, a more reasonable goal has been to create oracles relative to which SZK is not contained in other complexity classes; one can then unconditionally prove that “black-box” algorithms from other complexity classes cannot break SZK.

**Additional Context.** While much progress has been made in this direction (see Section 1.6 for details), the problem of giving an oracle separation between SZK and PP has been open since it was posed by Watrous in 2002 [Aar] and additionally mentioned as an open problem in [Aar12]. Here, PP is the set of decision problems decidable in polynomial time by randomized algorithms with unbounded error. Since a PP algorithm can flip polynomially many coins in its decision process, the gap between the acceptance

---

<sup>1</sup>*Computational* zero-knowledge, in which the zero-knowledge condition is that the verifier can sample from a distribution that is *computationally indistinguishable* from the transcript, has also been the subject of intense study. In this work we focus exclusively on statistical zero knowledge.

<sup>2</sup>Since  $SZK \subseteq AM \cap coAM \subseteq PH$ , if  $P \neq SZK$ , then  $P \neq PH$ , which in particular implies  $P \neq NP$ .

probabilities of yes and no instances can be exponentially small. PP is a very powerful complexity class – it contains NP and coNP (since it is trivially closed under complement) as well as  $\text{BPP}_{\text{path}}$ . Furthermore, by Toda’s theorem [Tod91],  $\text{P}^{\text{PP}}$  contains the entire polynomial hierarchy. Additionally Aaronson showed  $\text{PP} = \text{PostBQP}$ , the set of problems decidable by quantum algorithms equipped with postselection (the ability to discard all runs of an experiment which do not achieve an exponentially unlikely outcome). As a result, it is difficult to prove lower bounds against PP.

**Our Results.** We answer Watrous’ question by giving an oracle separating SZK from PP. In fact, we prove something significantly stronger: our oracle construction separates NISZK from UPP.<sup>3</sup>

**Theorem 1.1.** *There exists an oracle  $\mathcal{O}$  such that  $\text{NISZK}^{\mathcal{O}} \not\subseteq \text{UPP}^{\mathcal{O}}$ .*

## 1.2 Group 2: Limitations on the Power of Perfect Zero Knowledge

**Motivation.** Much progress has been made on understanding the relationship between natural variants of SZK [Oka96, GSV99, Fis02, Mal15, LZ16]. For example, it is known that  $\text{SZK} = \text{coSZK}$  [Oka96], and if  $\text{NISZK} = \text{coNISZK}$  then  $\text{SZK} = \text{NISZK} = \text{coNISZK}$  [GSV99]. Additionally Lovett and Zhang [LZ16] recently gave an oracle separation between NISZK and coNISZK as well as SZK and NISZK. However, many questions remain open, especially regarding the power of *perfect* zero-knowledge proof systems.

Many important SZK protocols, such as the ones for Graph Non-Isomorphism and Quadratic Nonresiduosity, are in fact PZK protocols. This illustrates the power of perfect zero knowledge. In this work, we are primarily concerned with studying the *limitations* of perfect zero knowledge. We are particularly interested in four questions: Does  $\text{SZK} = \text{PZK}$ ? What about their non-interactive variants, NISZK and NIPZK? Is PZK closed under complement, the way that SZK is? What about NIPZK? Answering any of these questions in the negative would require showing  $\text{P} \neq \text{NP}$ ,<sup>4</sup> so it is natural to try to exhibit oracles relative to which  $\text{SZK} \neq \text{PZK}$ ,  $\text{NISZK} \neq \text{NIPZK}$ ,  $\text{PZK} \neq \text{coPZK}$ , and  $\text{NIPZK} \neq \text{coNIPZK}$ .

**Additional Context.** In 1991, Aiello and Håstad [AH91a] gave evidence that PZK contains hard problems by creating an oracle relative to which PZK is not contained in BPP. On the other hand, they also gave an oracle that they *conjectured* separates SZK from PZK (but were unable to prove this). Exhibiting such an oracle requires a technique that can tell the difference between zero simulation error (PZK) and simulation to inverse exponential error (SZK), and prior to our work, no such technique was known. The question of whether  $\text{SZK} = \text{PZK}$  has been asked by Goldwasser [Gol15] as well. The analogous question for the non-interactive classes NISZK and NIPZK is also well motivated, and was explicitly asked in recent work of Lovett and Zhang [LZ16].

Determining whether variants of SZK satisfy the same closure properties as SZK is natural as well: indeed, a main result of Lovett and Zhang [LZ16] is an oracle relative to which  $\text{NISZK} \neq \text{coNISZK}$ .

**Our Results.** We give oracles separating SZK from PZK, NISZK from NIPZK, PZK from coPZK, and NIPZK from coNIPZK. The first two results answer the aforementioned questions raised by Aiello and Håstad [AH91a] (though our oracle is different from the candidate proposed by Aiello and Håstad), and Lovett and Zhang [LZ16]. Along the way, we show that PZK is contained in PP in a relativizing manner – this is in sharp contrast to SZK (see Theorem 1.1).

**Theorem 1.2.** *For any oracle  $\mathcal{O}$ ,  $\text{PZK}^{\mathcal{O}} \subseteq \text{PP}^{\mathcal{O}}$ . In addition, there exist oracles  $\mathcal{O}_1$  and  $\mathcal{O}_2$  such that  $\text{SZK}^{\mathcal{O}_1} \not\subseteq \text{PZK}^{\mathcal{O}_1}$ ,  $\text{NISZK}^{\mathcal{O}_1} \not\subseteq \text{NIPZK}^{\mathcal{O}_1}$ ,  $\text{PZK}^{\mathcal{O}_2} \not\subseteq \text{coPZK}^{\mathcal{O}_2}$ , and  $\text{NIPZK}^{\mathcal{O}_2} \not\subseteq \text{coNIPZK}^{\mathcal{O}_2}$ .*

<sup>3</sup>UPP is traditionally defined as an oracle complexity class, in which machines must output the correct answer with probability strictly greater than 1/2, and are charged for oracle queries but not for computation time. In this model, the gap between 1/2 and the probability of outputting the correct answer can be *arbitrarily* (in particular, superexponentially) small.

<sup>4</sup> $\text{P} = \text{NP}$  implies  $\text{P} = \text{PH}$ , and therefore  $\text{SZK} = \text{P}$ .

A summary of known relationships between complexity classes in the vicinity of SZK, including the new results established in this work, is provided in Figure 1.

### 1.3 Group 3: Communication Complexity

**Motivation and Context.** Paturi and Simon [PS86] introduced the model of *unbounded error communication complexity*, captured by the communication complexity class  $UPP^{cc}$ .<sup>5</sup> In this model, two parties with inputs  $(x, y)$  execute a randomized communication protocol, and are only required to output  $f(x, y)$  with probability strictly better than random guessing. Unbounded error communication protocols are extremely powerful, owing to this weak success criterion. In fact,  $UPP^{cc}$  represents the frontier of our understanding of communication complexity: it is the most powerful communication model against which we know how to prove lower bounds. We direct the interested reader to [GPW15a] for a thorough overview of communication complexity classes and their known relationships.

*What Lies Beyond the Frontier?* In an Arthur-Merlin game, a computationally-unbounded prover (Merlin) attempts to convince a computationally-bounded verifier (Arthur) of the value of a given Boolean function on a given input. The communication analogue of Arthur-Merlin games is captured by the communication complexity class  $AM^{cc}$ .

Many works have pointed to  $AM^{cc}$  as one of the simplest communication models against which we do not know how to prove superlogarithmic lower bounds. Works attempting to address this goal include [GPW15b, GPW15a, CCM<sup>+</sup>15, Lok01, LS09, PSS14, EFHK14, KP14, Kla11]. In fact, there are even simpler communication models against which we do not know how to prove lower bounds: it is known that  $NISZK^{cc} \subseteq SZK^{cc} \subseteq AM^{cc} \cap coAM^{cc} \subseteq \Sigma_2^{cc}$ , and we currently cannot prove lower bounds even against  $NISZK^{cc}$ .

Despite our inability to prove lower bounds against these classes, prior to our work it was possible that  $AM^{cc}$  is actually contained in  $UPP^{cc}$  (which, as described above, is a class against which we *can* prove lower bounds). The prior works that had come closest to ruling this out were as follows.

- $AM^{cc} \cap coAM^{cc} \not\subseteq P^{cc}$ . This was established (using a partial function) by Klauck [Kla11], who proved it by combining Vereschagin’s analogous query complexity separation with Sherstov’s pattern matrix method [She11].
- $\Sigma_2^{cc} \not\subseteq UPP^{cc}$ . This result was proved (using a total function) by Razborov and Sherstov [RS10].

Based on this state of affairs, Göös et al. [GPW15a] explicitly posed the problem of showing that  $AM^{cc} \cap coAM^{cc} \not\subseteq UPP^{cc}$ .

**Our Results.** In this work, we do even better than showing that  $AM^{cc} \not\subseteq UPP^{cc}$ . By “lifting” our oracle separation of NISZK and UPP to the communication setting, we show (using a partial function) that  $NISZK^{cc} \not\subseteq UPP^{cc}$ . Hence, if  $UPP^{cc}$  is taken to represent the frontier of our understanding of communication complexity, our result implies that  $NISZK^{cc}$  (and hence  $AM^{cc}$ ) is truly beyond the frontier. This also answers the question of Göös et al. [GPW15a].

**Theorem 1.3.** *There is a (promise) problem in  $NISZK^{cc}$  that is not in  $UPP^{cc}$ .*

---

<sup>5</sup>As is standard, given a query model  $C^{dt}$  (or a communication model  $C^{cc}$ ), we define a corresponding complexity class, also denoted  $C^{dt}$  (or  $C^{cc}$ ), consisting of all problems that have polylogarithmic cost protocols in the model.

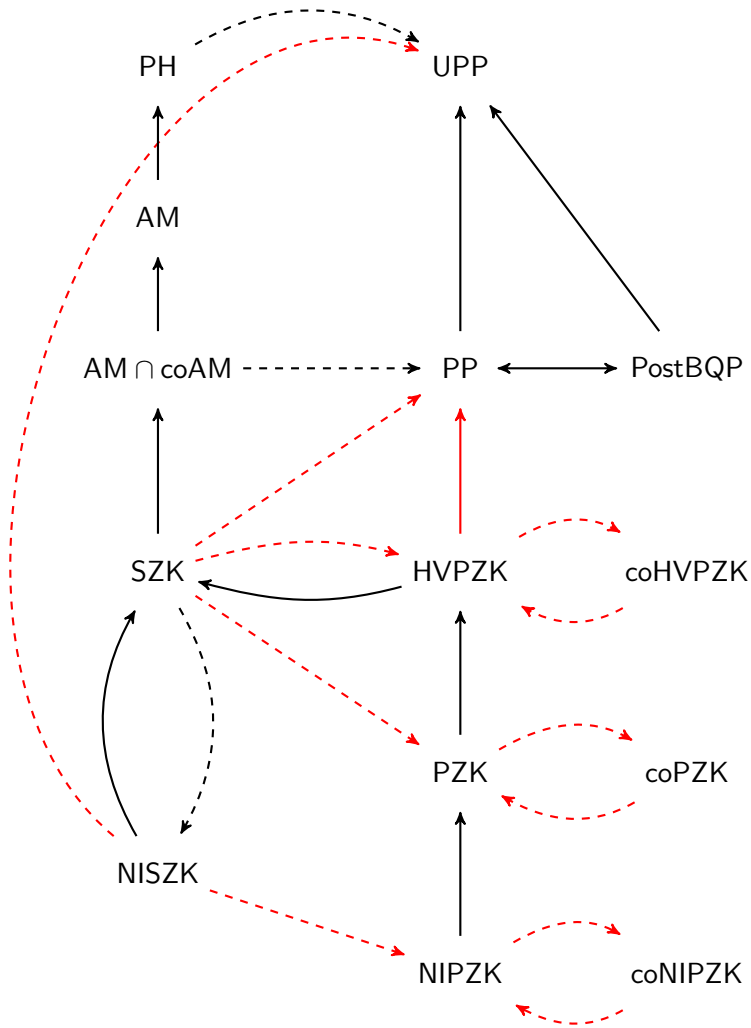


Figure 1:  $\mathcal{C}_1 \rightarrow \mathcal{C}_2$  indicates  $\mathcal{C}_1$  is contained in  $\mathcal{C}_2$  respect to *every* oracle, and  $\mathcal{C}_1 \dashrightarrow \mathcal{C}_2$  denotes that there is an oracle  $\mathcal{O}$  such that  $\mathcal{C}_1^{\mathcal{O}} \not\subseteq \mathcal{C}_2^{\mathcal{O}}$ . **Red** indicates new results. Certain non-inclusions that are depicted are subsumed by other non-inclusions (e.g., NISZK not in UPP subsumes SZK not in PP). We include some redundant arrows to facilitate comparison of our results to prior work.

## 1.4 Other Consequences of Our Results

In addition to the above oracle and communication separations, our results have a number of applications in other areas of theoretical computer science. For example, our results have implications regarding the power of complexity classes capturing the power of quantum computing with “more powerful” modified versions of quantum mechanics [ABFL16, Aar05], imply limitations on the Polarization Lemma of Sahai and Vadhan [SV03], yield novel lower bounds for certain forms of property testing algorithms, and imply upper bounds for *streaming interactive proofs* [CTY11, CCM<sup>+</sup>15]. These results are described in detail in Section 7.

## 1.5 Overview of Our Techniques

### 1.5.1 Oracle Separation of NISZK and UPP (Proof Overview for Theorem 1.1)

To describe our methods, it is helpful to introduce the notions of approximate degree and threshold degree, both of which are measures of Boolean function complexity that capture the difficulty of point-wise approximation by low-degree polynomials. The  $\varepsilon$ -approximate degree of a function  $f: \{0, 1\}^n \rightarrow \{0, 1\}$ , denoted  $\widetilde{\deg}_\varepsilon(f)$ , is the least degree of a real polynomial that point-wise approximates  $f$  to error  $\varepsilon$ . The threshold degree of  $f$ , denoted  $\deg_\pm(f)$ , is the least degree of a real polynomial that agrees in sign with  $f$  at all points. It is easy to see that threshold degree is equivalent to the limit of the approximate degree as the error parameter  $\varepsilon$  approaches  $1/2$  from below.

A recent and growing line of work has addressed a variety of open problems in complexity theory by establishing various forms of hardness amplification for approximate degree. Roughly speaking, these results show how to take a function  $f$  which is hard to approximate by degree  $d$  polynomials to error  $\varepsilon = 1/3$ , and turn  $f$  into a related function  $F$  that is hard to approximate by degree  $d$  polynomials even when  $\varepsilon$  is very close to  $1/2$ . In most of these works,  $F$  is obtained from  $f$  by block-composing  $f$  with a “hardness-amplifying function”  $g$ . We denote such a block-composition by  $g(f)$ .

The technical core of our result lies in establishing a new form of hardness amplification for approximate degree. Specifically, let  $g$  be the partial function  $\text{GapMaj}_n: \{0, 1\}^n \rightarrow \{0, 1\}$  (throughout this introduction, whenever necessary, we use subscripts after function names to clarify the number of variables on which the function is defined). Here  $\text{GapMaj}$  is the gapped majority function, defined, for some  $1 \geq \delta > 0.5$ , to be 1 if  $\geq \delta$  fraction of its inputs are 1, to be 0 if  $\geq \delta$  fraction of its inputs are 0, and to be undefined otherwise (in this introduction, we will ignore the precise choice of  $\delta$  that we use in our formal results).<sup>6</sup>

**Theorem 1.4.** (Informal) *Let  $f: \{0, 1\}^M \rightarrow \{0, 1\}$ . Suppose that  $\widetilde{\deg}_{1/3}(f) \geq d$ . Define  $F: \{0, 1\}^{n \cdot M} \rightarrow \{0, 1\}$  via  $F = \text{GapMaj}_n(f)$ . Then  $\deg_\pm(F) = \Omega(\min(d, n))$ .*

In our main application of Theorem 1.4, we apply the theorem to a well-known (partial) function  $f = \text{Col}_M$  called the Collision problem. This function is known to have approximate degree  $\tilde{\Omega}(M^{1/3})$ , so Theorem 1.4 implies that  $F := \text{GapMaj}_{M^{1/3}}(\text{Col}_M)$  has threshold degree  $\tilde{\Omega}(M^{1/3})$ . Standard results then imply that the UPP query complexity of  $F$  is  $\tilde{\Omega}(M^{1/3})$  as well. That is,  $F \notin \text{UPP}^{\text{dt}}$ .

**Corollary 1.5** (Informal). *Let  $m = M^{4/3}$ , and define  $F: \{0, 1\}^m \rightarrow \{0, 1\}$  via  $F := \text{GapMaj}_{M^{1/3}}(\text{Col}_M)$ . Then  $\text{UPP}^{\text{dt}}(F) = \tilde{\Omega}(m^{1/4})$ .*

Moreover, as we show later,  $\text{GapMaj}_{M^{1/3}}(\text{Col}_M)$  is in  $\text{NISZK}^{\text{dt}}$ . Hence, we obtain a separation between  $\text{NISZK}^{\text{dt}}$  and  $\text{UPP}^{\text{dt}}$ . The desired oracle separating  $\text{NISZK}$  from  $\text{UPP}$  follows via standard methods.

---

<sup>6</sup>We clarify that if  $f$  is a partial function then  $\text{GapMaj}_n(f)$  is technically not a composition of functions, since for some inputs  $x = (x_1, \dots, x_n)$  on which  $\text{GapMaj}_n(f)$  is defined, there may be values of  $i$  for which  $x_i$  is outside of the domain of  $f$ . See Section 2.4 for further discussion of this point.

**Comparison of Theorem 1.4 to Prior Work.** The hardness amplification result from prior work that is most closely related to Theorem 1.4 is due to Sherstov [She14]. Sherstov’s result makes use of a notion known as (positive) one-sided approximate degree [She14, BT15b]. Positive one-sided approximate degree is a measure that is intermediate between approximate degree and threshold degree – the positive one-sided approximate degree of  $f$ , denoted  $\text{deg}_\varepsilon^+(f)$ , is always at most as large as the approximate degree of  $f$  but can be much smaller, and it is always at least as large as the threshold degree of  $f$  but can be much larger (see Section 2.2 for a formal definition of positive one-sided approximate degree).<sup>7</sup>

**Theorem 1.6** (Sherstov). *Let  $f: \{0, 1\}^M \rightarrow \{0, 1\}$ . Suppose that  $\text{deg}_{1/3}^+(f) \geq d$ . Define  $F: \{0, 1\}^{n \cdot M} \rightarrow \{0, 1\}$  via  $F = \text{AND}_n(f)$ . Then  $\text{deg}_\pm(F) = \Omega(\min(d, n))$ .*<sup>8</sup>

There are two differences between Theorems 1.4 and 1.6. The first is that the hardness-amplifier in Theorem 1.4 is GapMaj, while in Theorem 1.6 it is AND. GapMaj is a “simpler” function than AND in the following sense: block-composing  $f$  with GapMaj preserves membership in complexity classes such as  $\text{NISZK}^{\text{dt}}$  and  $\text{SZK}^{\text{dt}}$ ; this is not the case for AND, as AND itself is not in  $\text{SZK}^{\text{dt}}$ . This property is essential for us to obtain threshold degree lower bounds even for functions that are in  $\widetilde{\text{NISZK}}^{\text{dt}}$ .

The second difference is that Theorem 1.4 holds under the assumption that  $\widetilde{\text{deg}}_{1/3}(f) \geq d$ , while Theorem 1.6 makes the stronger assumption that  $\text{deg}_\varepsilon^+(f) \geq d$ . While we do not exploit this second difference in our applications, ours is the first form of hardness amplification that works for approximate degree rather than one-sided approximate degree. This property has already been exploited in subsequent work [BT17].

**Proof Sketch for Theorem 1.4.** A *dual polynomial* is a dual solution to an appropriate linear program capturing the threshold degree of any function. Specifically, for a (partial) function  $f$  defined on a subset of  $\{0, 1\}^n$ , a dual polynomial witnessing the fact that  $\widetilde{\text{deg}}_\varepsilon(f) \geq d$  is a function  $\psi: \{0, 1\}^n \rightarrow \mathbb{R}$  that satisfies the following three properties.

- (a)  $\psi$  is uncorrelated with all polynomials  $p$  of total degree at most  $d$ . That is, for any  $p: \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $\text{deg}(p) \leq d$ , it holds that  $\sum_{x \in \{0, 1\}^n} \psi(x) \cdot p(x) = 0$ . We refer to this property by saying that  $\psi$  has *pure high degree*  $d$ .
- (b)  $\psi$  has  $\ell_1$  norm equal to 1, i.e.,  $\sum_{x \in \{0, 1\}^n} |\psi(x)| = 1$ .
- (c)  $\psi$  has correlation at least  $\varepsilon$  with  $f$ . That is, if  $D$  denotes the domain on which  $f$  is defined, then  $\sum_{x \in D} \psi(x) \cdot f(x) - \sum_{x \in \{0, 1\}^n \setminus D} |\psi(x)| > \varepsilon$ .

It is not hard to see that a dual witness for the fact that  $\text{deg}_\pm(f) \geq d$  is a function  $\psi$  satisfying Properties (a) and (b) above, that additionally is *perfectly* correlated with  $f$ . That is,  $\psi$  additionally satisfies

$$\sum_{x \in D} \psi(x) \cdot f(x) - \sum_{x \in \{0, 1\}^n \setminus D} |\psi(x)| = 1. \quad (1)$$

<sup>7</sup>The notion of positive one-sided approximate degree treats inputs in  $f^{-1}(1)$  and  $f^{-1}(0)$  asymmetrically. There is an analogous notion called negative one-sided approximate degree that reverses the roles of  $f^{-1}(1)$  and  $f^{-1}(0)$  [Tha14, KT14]. Our use of the positive vs. negative terminology follows prior work [Tha14, KT14] – other prior works [She14, BT15b] only used negative one-sided approximate degree, and referred to this complexity measure without qualification as one-sided approximate degree. In this paper, we exclusively use the notion of positive one-sided approximate degree.

<sup>8</sup>Sherstov stated his result for  $\text{OR}_n(f)$  under the assumption that  $f$  has large *negative* one-sided approximate degree. Our statement of Theorem 1.6 is the equivalent result under the assumption that  $f$  has large positive one-sided approximate degree.

In this case,  $\psi \cdot f$  is non-negative, and is referred to as an *orthogonalizing distribution* for  $f$ .

We prove Theorem 1.4 by constructing an explicit orthogonalizing distribution for  $\text{GapMaj}_n(f)$ . Specifically, we show how to take a dual polynomial witnessing the fact that  $\widetilde{\text{deg}}_{1/3}(f) \geq d$ , and turn it into an orthogonalizing distribution witnessing the fact that  $\text{deg}_{\pm}(F) = \Omega(\min(d, n))$ .

Our construction of an orthogonalizing distribution for  $\text{GapMaj}_n(f)$  is inspired by and reminiscent of Sherstov’s construction of an orthogonalizing distribution for  $\text{AND}_n(f)$  [She14], which in turn builds on a dual polynomial for  $\text{AND}_n(f)$  constructed by Bun and Thaler [BT15b]. In more detail, Bun and Thaler constructed a dual polynomial  $\psi_{BT}$  of pure high degree  $d$  that had correlation  $1 - 2^{-n}$  with  $\text{AND}_n(f)$ . Sherstov’s dual witness was defined as  $\psi_{BT} + \psi_{corr}$ , where  $\psi_{corr}$  is an *error-correction term* that also has pure high degree  $\Omega(d)$ . The purpose of  $\psi_{corr}$  is to “zero-out”  $\psi_{BT}$  at all points where  $\psi_{BT}$  differs in sign from  $f$ , without affecting the sign of  $\psi_{BT}$  on any other inputs.

Naively, one might hope that  $\psi_{BT} + \psi_{corr}$  is also a dual witness to the fact that  $\text{deg}_{\pm}(\text{GapMaj}_n(f))$  is large. Unfortunately, this is not the case, as it does not satisfy Equation (1) with respect to  $\text{GapMaj}_n(f)$ . It is helpful to think of this failure as stemming from two issues. First,  $\psi_{BT} + \psi_{corr}$  places non-zero weight on many inputs on which  $\text{GapMaj}_n(f)$  is undefined (i.e., on inputs for which fewer than  $\delta n$  copies of  $f$  evaluate to 1 and fewer than  $\delta n$  copies of  $f$  evaluate to 0). Second, there are inputs on which  $\text{GapMaj}_n(f)$  is defined, yet  $\text{AND}_d(f)$  does not agree with  $\text{GapMaj}_n(f)$ .

To address both of these issues, we add a *different* error-correction term  $\psi'_{corr}$  of pure high degree  $\tilde{\Omega}(\min(n, d))$  to  $\psi_{BT}$ . Our correction term does not just zero out the value of  $\psi_{BT}$  on inputs on which it disagrees in sign with  $\text{AND}_n(f)$ , but also zeros it out on inputs for which  $\text{GapMaj}_n(f)$  is undefined, and on inputs on which  $\text{AND}_n(f)$  does not agree with  $\text{GapMaj}_n(f)$ .

Moreover, we show that adding  $\psi'_{corr}$  does not affect the sign of  $\psi_{BT}$  on other inputs – achieving this requires some new ideas in both the definition of  $\psi'_{corr}$  and its analysis. Putting everything together, we obtain a dual witness  $\psi_{BT} + \psi'_{corr}$  showing that  $\text{deg}_{\pm}(\text{GapMaj}_n(f)) = \Omega(\min(n, d))$ .

## 1.5.2 Limitations on the Power of Perfect Zero Knowledge (Proof Overview For Theorem 1.2)

We begin the proof of Theorem 1.2 by showing that HVPZK (*honest verifier perfect zero knowledge*) is contained in PP in a relativizing manner (see Section 5). Since the inclusions  $\text{PP} \subseteq \text{UPP}$ ,  $\text{NIPZK} \subseteq \text{HVPZK}$ ,  $\text{PZK} \subseteq \text{HVPZK}$ , and  $\text{NISZK} \subseteq \text{SZK}$  hold with respect to any oracle, this means that our oracle separating NISZK from UPP (Theorem 1.1) also separates SZK from PZK and NISZK from NIPZK.

We then turn to showing that PZK and NIPZK are not closed under complement with respect to some oracle. Since the proofs are similar, we focus on the case of PZK in this overview.

Since both PZK and coPZK are contained in PP with respect to any oracle, our oracle separation of NISZK from PP (Theorem 1.1) does not imply an oracle relative to which  $\text{PZK} \neq \text{coPZK}$ . Instead, to obtain this result we prove a new amplification theorem for one-sided approximate degree. Using similar techniques as Theorem 1.4, we show that if  $f$  has high positive one-sided approximate degree, then block-composing  $f$  with the gapped AND function yields a function with high threshold degree. Here  $\text{GapAND}$  is partial function that outputs 1 if all inputs are 1, outputs 0 if at least a  $\delta$  fraction of inputs are 0, and is undefined otherwise.

**Theorem 1.7.** (Informal) *Let  $f : \{0, 1\}^M \rightarrow \{0, 1\}$ . Suppose that  $\text{deg}_{1/3}^+(f) \geq d$ . Then  $\text{deg}_{\pm}(\text{GapAND}_n(f)) = \Omega(\min(d, n))$ .*

We then show that (a)  $\text{PZK}^{\text{dt}}$  is closed under composition with  $\text{GapAND}$  and (b) there is a function  $f$  in  $\text{PZK}^{\text{dt}}$  whose complement  $\bar{f}$  has high one-sided positive one-sided approximate degree. If  $\text{PZK}^{\text{dt}}$  were closed under complement, then  $\bar{f}$  would be in  $\text{PZK}^{\text{dt}}$ . By amplifying the hardness of  $f$  using Theorem 1.7, we obtain a problem that is still in  $\text{PZK}^{\text{dt}}$  (this holds by property (a)) yet outside of  $\text{PP}^{\text{dt}}$  (this holds by property (b), together with Theorem 1.7). This is easily seen to contradict the fact PZK is in PP relative to



all oracles. Hence,  $\bar{f}$  is a function in  $\text{coPZK}^{\text{dt}}$  that is not in  $\text{PZK}^{\text{dt}}$ , and standard techniques translate this fact into an oracle separating  $\text{coPZK}$  from  $\text{PZK}$ . We provide details of these results in Section 5.

### 1.5.3 Lifting to Communication Complexity: Proof Overview For Theorem 1.3

To extend our separation between  $\text{NISZK}$  and  $\text{UPP}$  to the world of communication complexity, we build on recently developed methods of Bun and Thaler [BT16], who themselves used and generalized the breakthrough work of Razborov and Sherstov [RS10]. Razborov and Sherstov showed that if  $F$  has high threshold degree and this is witnessed by an orthogonalizing distribution that satisfies an additional smoothness condition, then  $F$  can be transformed into a related function  $F'$  that has high  $\text{UPP}^{\text{cc}}$  complexity (specifically,  $F'$  is obtained from  $F$  via the *pattern matrix method* introduced in [She11]). So in order to turn  $\text{GapMaj}(\text{Col})$  into a function with high  $\text{UPP}^{\text{cc}}$  complexity, it is enough to give a *smooth* orthogonalizing distribution for  $F$ .

Bun and Thaler [BT16] showed how to take the dual witness Sherstov constructed for  $\text{OR}(f)$  in the proof of Theorem 1.6 and smooth it out, assuming the inner function  $f$  satisfies some modest additional conditions. Fortunately, a variant of  $\text{Col}$  called the *Permutation Testing Problem* (PTP for short) satisfies these additional conditions, and since our construction of an orthogonalizing distribution for  $\text{GapMaj}(\text{PTP})$  is reminiscent of Sherstov’s orthogonalizing distribution for  $\text{OR}(f)$ , we are able to modify the methods of Bun and Thaler to smooth out our dual witness for  $\text{GapMaj}(\text{PTP})$ . Although there are many technical details to work through, adopting the methodology of Bun and Thaler to our setting does not require substantially new ideas, and we do not consider it to be a major technical contribution of this work. Nonetheless, it does require the careful management of various subtleties arising from our use of promise problems as opposed to total Boolean functions, and our final communication lower bound inherits many of the advantages of our Theorem 1.4 relative to prior work (such as applying to functions with high approximate degree rather than high one-sided approximate degree).

## 1.6 Other Works Giving Evidence for the Hardness of SZK

As mentioned in Section 1.2, Aiello and Håstad showed that  $\text{PZK}$  (and also  $\text{SZK}$ ) is not contained in  $\text{BPP}$  relative to some oracle [AH91a]. Agrawal et al. later used similar techniques to show that  $\text{SZK}$  is not contained in the class  $\text{SRE}$  (which can be viewed as a natural generalization of  $\text{BPP}$ ) relative to some oracle [AIKP15]. Aaronson [Aar02] gave an oracle relative to which  $\text{SZK}$  is not contained in  $\text{BQP}$  – and therefore quantum computers cannot break  $\text{SZK}$ -hard cryptosystems in a black-box manner. Building on that work, Aaronson [Aar12] later gave oracle separations against the class  $\text{QMA}$  (a quantum analogue of  $\text{NP}$ ) and the class  $\text{A}_0\text{PP}$  (a class intermediate between  $\text{QMA}$  and  $\text{PP}$ ). Therefore even quantum proofs cannot certify  $\text{SZK}$  in a black-box manner<sup>9</sup>.

Until recently, the lower bound most closely related to our oracle separation of  $\text{NISZK}$  and  $\text{UPP}$  (cf. Theorem 1.1) was Vereschagin’s result from 1995, which gave an oracle relative to which  $\text{AM} \cap \text{coAM}$  is not contained in  $\text{PP}$  [Ver95]. Our result is an improvement on Vereschagin’s because the inclusions  $\text{NISZK} \subseteq \text{SZK} \subseteq \text{AM} \cap \text{coAM}$  can be proved in a relativizing manner (cf. Figure 1). It also generalizes Aaronson’s oracle separation between  $\text{SZK}$  and  $\text{A}_0\text{PP}$  [Aar12].

Vereschagin [Ver95] also reports that Beigel claimed a simple proof of the existence of a function  $f$  that is in the query complexity class  $\text{AM}^{\text{dt}}$ , but is not in the query complexity class  $\text{UPP}^{\text{dt}}$ . Our result improves on Beigel’s in two regards. First, since  $\text{NISZK}^{\text{dt}} \subseteq \text{AM}^{\text{dt}}$ , separating  $\text{NISZK}^{\text{dt}}$  from  $\text{UPP}^{\text{dt}}$  is more difficult than separating  $\text{AM}^{\text{dt}}$  from  $\text{UPP}^{\text{dt}}$ . Second, Beigel only claimed a superlogarithmic lower bound on the  $\text{UPP}^{\text{dt}}$  query complexity of  $f$ , while we give a polynomial lower bound.

---

<sup>9</sup>Note, however, that oracle separations do not necessarily imply the analogous separations in the “real world” – see [Bar01] and [CCG<sup>+</sup>94] for instances in which the situation in the presence of oracles is far from the situation in the real world.

Theorem 1.1 also improves on very recent work of Chen [Che16a, Che16b], which gave a query separation between the classes  $\mathsf{P}^{\text{SZK}}$  and  $\mathsf{PP}$ .

## 2 Technical Preliminaries

### 2.1 Complexity Classes

**Notation.** In an interactive proof  $(P, V)$  where  $P$  is the prover and  $V$  is the verifier, we denote by  $(P, V)(x)$  the random variable corresponding to the transcript of the protocol on input  $x$ . For distributions  $D_0$  and  $D_1$ ,  $\|D_0 - D_1\|$  denotes the Total Variational Distance between them  $\left(\|D_0 - D_1\| = \frac{1}{2}|D_0 - D_1|_1\right)$ .

**Definition 2.1** (Honest Verifier Statistical Zero Knowledge). A promise problem  $L = (L_Y, L_N)$  is in HVSZK if there exists a tuple of Turing machines  $(P, V, S)$ , where the *verifier*  $V$  and *simulator*  $S$  run in probabilistic polynomial time, satisfying the following:

- $(P, V)$  is an interactive proof for  $L$  with negligible completeness and soundness errors.
- For any  $x \in L_Y$ ,

$$\|S(x) - (P, V)(x)\| \leq \text{negl}(|x|)$$

**Definition 2.2** (Non-Interactive SZK). A promise problem  $L = (L_Y, L_N)$  is in NISZK if there exist a tuple of Turing machines  $(P, V, S)$ , where the *verifier*  $V$  and *simulator*  $S$  run in probabilistic polynomial time, satisfying the following:

- $(P, V)$  is an interactive proof for  $L$  with negligible completeness and soundness errors, with the following additional conditions:
  1.  $P$  and  $V$  both have access to a long enough common random string.
  2. The interactive proof consists of a single message from  $P$  to  $V$ .
- For any  $x \in L_Y$ ,

$$\|S(x) - (P, V)(x)\| \leq \text{negl}(|x|)$$

The definitions of these classes in the presence of an oracle are the same, except that  $P$ ,  $V$ , and  $S$  all have access to the oracle.

It is easy to see that NISZK is contained in HVSZK, even in the presence of oracles. The class SZK is defined to be almost the same as HVSZK, except for the stipulation that for any verifier (even one that deviates from the prescribed protocol), there is a simulator that simulates the prover's interaction with that verifier. It was shown in [GSV98] that SZK is equal to HVSZK, and their proof continues to hold in the presence of any oracle. For this reason, NISZK is also contained in SZK in the presence of any oracle, and we shall be implicitly making use of this fact at several points where we state corollaries for SZK instead of HVSZK.

## 2.2 Approximate Degree, Threshold Degree, and Their Dual Characterizations

We first recall the definitions of approximate degree, positive one-sided approximate degree, and threshold degree for partial functions.

**Definition 2.3.** Let  $D \subseteq \{0, 1\}^M$ , and let  $f$  be a function mapping  $D$  to  $\{0, 1\}$ .

- The *approximate degree* of  $f$  with approximation constant  $0 \leq \varepsilon < 1/2$ , denoted  $\widetilde{\deg}_\varepsilon(f)$ , is the least degree of a real polynomial  $p: \{0, 1\}^M \rightarrow \mathbb{R}$  such that  $|p(x) - f(x)| \leq \varepsilon$  when  $x \in D$ , and  $|p(x)| \leq 1 + \varepsilon$  for all  $x \notin D$ . We refer to such a  $p$  as an *approximating polynomial* for  $f$ . We use  $\deg(f)$  to denote  $\widetilde{\deg}_{1/3}(f)$ .
- The *threshold degree* of  $f$ , denoted  $\deg_\pm(f)$ , is the least degree of a real polynomial  $p$  such that  $p(x) > 0$  when  $f(x) = 1$ , and  $p(x) < 0$  when  $f(x) = 0$ .
- The *positive one-sided approximate degree* of  $f$  with approximation constant  $0 \leq \varepsilon < 1/2$ , denoted  $\deg_\varepsilon^+(f)$ , is the least degree of a real polynomial  $p$  such that  $|p(x) - 1| \leq \varepsilon$  for all  $x \in f^{-1}(1)$ , and  $p(x) \leq \varepsilon$  when  $x \in f^{-1}(0)$ . We refer to such a  $p$  as a *positive one-sided approximating polynomial* for  $f$ . We use  $\deg^+(f)$  to denote  $\deg_{1/3}^+(f)$ .

**Remark.** We highlight the following subtlety in Definition 2.3: an approximating polynomial for a partial function  $f$  is required to be bounded in absolute value even outside of the domain  $D$  on which  $f$  is defined, yet this is not required of a one-sided approximating polynomial for  $f$ . The reason we choose to require an approximating polynomial to be bounded outside of  $D$  is to ensure that the Col function (defined later in Section 2.5) has large approximate degree.

There are clean dual characterizations for each of the three quantities defined in Definition 2.3. We state these characterizations without proof, and direct the interested reader to [She15, She14, BT15a] for details.

For a function  $\psi: \{0, 1\}^M \rightarrow \mathbb{R}$ , define the  $\ell_1$  norm of  $\psi$  by  $\|\psi\|_1 = \sum_{x \in \{0, 1\}^M} |\psi(x)|$ . If the support of a function  $\psi: \{0, 1\}^M \rightarrow \mathbb{R}$  is (a subset of) a set  $D \subseteq \{0, 1\}^M$ , we will write  $\psi: D \rightarrow \mathbb{R}$ . For functions  $f, \psi: D \rightarrow \mathbb{R}$ , denote their inner product by  $\langle f, \psi \rangle := \sum_{x \in D} f(x)\psi(x)$ . We say that a function

$\psi: \{0, 1\}^M \rightarrow \mathbb{R}$  has *pure high degree*  $d$  if  $\psi$  is uncorrelated with any polynomial  $p: \{0, 1\}^M \rightarrow \mathbb{R}$  of total degree at most  $d$ , i.e., if  $\langle \psi, p \rangle = 0$ .

**Theorem 2.4.** Let  $f: D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function and  $\varepsilon$  be a real number in  $[0, 1/2)$ .  $\widetilde{\deg}_\varepsilon(f) > d$  if and only if there is a real function  $\psi: \{0, 1\}^M \rightarrow \mathbb{R}$  such that:

1. (Pure high degree):  $\psi$  has pure high degree of  $d$ .
2. (Unit  $\ell_1$ -norm):  $\|\psi\|_1 = 1$ .
3. (Correlation):  $\sum_{x \in D} \psi(x)f(x) - \sum_{x \notin D} |\psi(x)| > \varepsilon$ .

**Theorem 2.5.** Let  $f: D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function.  $\deg_\pm(f) > d$  if and only if there is a real function  $\psi: D \rightarrow \mathbb{R}$  such that:

1. (Pure high degree):  $\psi$  has pure high degree of  $d$ .
2. (Sign Agreement):  $\psi(x) \geq 0$  when  $f(x) = 1$ , and  $\psi(x) \leq 0$  when  $f(x) = 0$ .

3. (Non-triviality):  $\|\psi\|_1 > 0$ .

**Theorem 2.6.** Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function and  $\varepsilon$  be a constant in  $[0, 1/2)$ .  $\deg_\varepsilon^+(f) > d$  if and only if there is a real function  $\psi : D \rightarrow \mathbb{R}$  such that:

1. (Pure high degree):  $\psi$  has pure high degree of  $d$ .
2. (Unit  $\ell_1$ -norm):  $\|\psi\|_1 = 1$ .
3. (Correlation):  $\langle \psi, f \rangle > \varepsilon$ .
4. (Negative Sign Agreement):  $\psi(x) \leq 0$  whenever  $f(x) = 0$ .

### 2.3 PP<sup>dt</sup> and UPP<sup>dt</sup>

Now we define the two natural analogues of PP complexity in the query model.

**Definition 2.7.** Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function. Let  $\mathcal{T}$  be a randomized decision tree which computes  $f$  with a probability better than  $1/2$ . Let  $\alpha$  be the maximum real number such that

$$\min_{x \in D} \Pr[\mathcal{T} \text{ outputs } f(x) \text{ on input } x] \geq \frac{1}{2} + \alpha.$$

Then we define the PP query cost of  $\mathcal{T}$  for  $f$  to be  $\text{PP}^{\text{dt}}(\mathcal{T}; f) = C(\mathcal{T}; f) + \log_2(1/\alpha)$ , where  $C(\mathcal{T}; f)$  denotes the maximum number of queries  $\mathcal{T}$  incurs on an input in the worst case. We define  $\text{UPP}^{\text{dt}}(\mathcal{T}; f) = C(\mathcal{T}; f)$ . Observe that  $\text{UPP}^{\text{dt}}(\mathcal{T}; f)$  is the same as  $\text{PP}^{\text{dt}}(\mathcal{T}; f)$ , except that the advantage  $\alpha$  of the randomized decision tree over random guessing is not incorporated into  $\text{UPP}^{\text{dt}}(\mathcal{T}; f)$ . We define  $\text{PP}^{\text{dt}}(f)$  (respectively,  $\text{UPP}^{\text{dt}}$ ) as the minimum of  $\text{PP}^{\text{dt}}(\mathcal{T}; f)$  (respectively,  $\text{UPP}^{\text{dt}}(\mathcal{T}; f)$ ) over all  $\mathcal{T}$  that computes  $f$  with a probability better than  $1/2$ .

$\text{PP}^{\text{dt}}$  is closely related to approximate degree with error very close to  $1/2$ . We have the following well-known relationship between them.

**Lemma 2.8.** Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function. Suppose  $\widetilde{\deg}_{1/2-2^{-d}}(f) > d$  for some positive integer  $d$ . Then  $\text{PP}^{\text{dt}}(f) > d/2$ .

Meanwhile,  $\text{UPP}^{\text{dt}}$  is exactly characterized by threshold degree.

**Lemma 2.9.** Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function. Then  $\text{UPP}^{\text{dt}}(f) = \deg_\pm(f)$ .

### 2.4 Gap Majority and Gap AND

In this subsection we introduce a transformation of partial functions which will be used in this paper.

**Definition 2.10.** Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function and  $n$  be a positive integer,  $0.5 < \varepsilon \leq 1$  be a real number. We define the gap majority version of  $f$ , denoted by  $\text{GapMaj}_{n,\varepsilon}(f)$ , as follows:

Given an input  $x = (x_1, x_2, \dots, x_n) \in \{0, 1\}^{M \cdot n}$ , we define  $n_{\text{Yes}}(x) := \sum_{i=1}^n \mathbf{1}_{x_i \in D \wedge f(x_i)=1}$  and  $n_{\text{No}}(x) := \sum_{i=1}^n \mathbf{1}_{x_i \in D \wedge f(x_i)=0}$ . Then

$$\text{GapMaj}_{n,\varepsilon}(f)(x) = \begin{cases} 1 & \text{when } n_{\text{Yes}}(x) \geq \varepsilon \cdot n \\ 0 & \text{when } n_{\text{No}}(x) \geq \varepsilon \cdot n \\ \text{undefined} & \text{otherwise} \end{cases}$$

Note that even on inputs  $x$  for which  $\text{GapMaj}_{n,\varepsilon}(f)(x)$  is defined, there may be some values of  $i$  for which  $x_i$  is not in  $D$ . For brevity, we will occasionally write  $\text{GapMaj}(f)$  when  $n$  and  $\varepsilon$  are clear from context.

We also define the GapAND function. This is a partial function that agrees with the total function AND wherever it is defined.

**Definition 2.11.** Let  $n$  be a positive integer,  $0 < \varepsilon < 1$  be a constant. We define the Gapped AND function,  $\text{GapAND}_{n,\varepsilon} : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^n$ , as the function that outputs 1 if all inputs are 1; outputs 0 if at least  $\varepsilon \cdot n$  inputs are 0; and is undefined otherwise.

For a partial function  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$ , we define  $\text{GapAND}_{n,\varepsilon}(f)$  to be a true block-composition of partial functions, i.e.,  $\text{GapAND}_{n,\varepsilon}(f)(x_1, \dots, x_n) = \text{GapAND}_{n,\varepsilon}(f(x_1), \dots, f(x_n))$  whenever the right hand side of the equality is defined, and  $\text{GapAND}_{n,\varepsilon}(f)$  is undefined otherwise.

**Remark 2.12.** Note that  $\text{GapMaj}_{n,\varepsilon}(f)$  is not technically a block-composition of partial functions, since  $\text{GapMaj}_{n,\varepsilon}(f)(x_1, \dots, x_n)$  is defined even on some inputs for which some  $f(x_i)$  is not defined.

## 2.5 Problems

We now recall the Collision problem. This problem interprets its input as a function  $f$  mapping  $[n]$  to  $[n]$ , and the goal is to decide whether the input is a permutation or is 2-to-1, promised that one of them is the case. We need a slightly generalized version, which asks to distinguish between permutations and  $k$ -to-1 functions.

**Definition 2.13** (Collision problem). Fix an integer  $k \geq 2$ , and assume for simplicity that  $n$  is a power of 2. The partial function  $\text{Col}_n^k$  is defined on a subset of  $\{0, 1\}^{n \log n}$ . It interprets its input as specifying a function  $f : [n] \rightarrow [n]$  in the natural way, and evaluates to 1 if  $f$  is a permutation, 0 if  $f$  is a  $k$ -to-1 function, and is undefined otherwise. When  $k$  and  $n$  are clear from context, we write  $\text{Col}$  for brevity.

This problem admits a simple SZK protocol in which the verifier makes only  $\text{polylog}(n)$  queries to the input. Specifically, the verifier executes the following sub-protocol  $\text{polylog}(n)$  times: the verifier chooses a random  $i \in [n]$ , makes a single query to learn  $f(i)$ , sends  $f(i)$  to the prover, and rejects if the prover fails to respond with  $i$ . It is easy to see that the sub-protocol has perfect completeness, constant soundness error, and is perfect zero knowledge. Because the sub-protocol is repeated  $\text{polylog}(n)$  times, the total soundness error is negligible.

In 2002, Aaronson [Aar02] proved the first non-constant lower bound for the  $\text{Col}_n^2$  problem: namely, any bounded-error quantum algorithm to solve it needs  $\Omega(n^{1/5})$  queries to  $f$ . Aaronson and Shi [AS04] subsequently improved the lower bound to  $\Omega(n^{1/3})$ , for functions  $f : [n] \rightarrow [3n/2]$ ; then Ambainis [Amb05] and Kutin [Kut05] proved the optimal  $\Omega(n^{1/3})$  lower bound for functions  $f : [n] \rightarrow [n]$ .

We need a version of the lower bound that makes explicit the dependence on  $k$  and  $\varepsilon$ .

**Theorem 2.14** (Implicit in Kutin [Kut05]).  $\widetilde{\text{deg}}_\varepsilon(\text{Col}_n^k) = \Omega(\sqrt[3]{(1/2 - \varepsilon) \cdot n/k})$  for any  $0 < \varepsilon < 1/2$  and  $k|n$ .

See also [BT15a] for a direct constructive proof (using Theorem 2.4) for the above theorem in the case that  $k = 2$ .

We will also utilize the Permutation Testing Problem, or PTP for short. This problem, which is closely related to the Collision problem, was defined in [Aar12], which also (implicitly) proved a bound on its one-sided approximate degree.

**Definition 2.15** (PTP). Given a function  $f : [n] \rightarrow [n]$  (represented as a string in  $\{0, 1\}^{n \log n}$ ),

1.  $\text{PTP}_n(f) = 1$  if  $f$  is a permutation.
2.  $\text{PTP}_n(f) = 0$  if  $f(i)$  differs from every permutation on at least  $n/8$  values of  $i$ .
3.  $\text{PTP}_n(f)$  is undefined otherwise.

**Theorem 2.16** (Implicit in [Aar12]). *For any  $0 < \varepsilon < 1/6$ ,*

$$\text{deg}_\varepsilon^+(\overline{\text{PTP}}_n) = \Omega(n^{1/3})$$

The SZK protocol described for  $\overline{\text{Col}}$  works unmodified for PTP as well.

### 3 Hardness Amplification For Approximate Degree

In this section we prove a novel hardness amplification theorem. Specifically, we show that for any function  $f$  with high approximate degree, composing  $f$  with  $\text{GapMaj}$  yields a function with high threshold degree, and hence the resulting function is hard for any UPP algorithm in the query model. Similarly, we show that if  $f$  has high positive one-sided approximate degree, then composing  $f$  with  $\text{GapAND}$  yields a function with high threshold degree.

Note that this hardness amplification theorem is tight, in the sense that if  $f$  has low approximate degree, then composing  $f$  with  $\text{GapMaj}$  yields a function that has low UPP query complexity, and the same holds for composing  $f$  with  $\text{GapAND}$  if  $f$  has low positive one-sided approximate degree. See Appendix A for details.

#### 3.1 Notation

For a partial function  $f$ , an integer  $n$  and a real  $\varepsilon \in (1/2, 1]$ , we denote  $\text{GapMaj}_{n,\varepsilon}(f)$  by  $F$  for convenience, where  $n$  and  $\varepsilon$  will always be clear in the context. We also use  $x = (x_1, x_2, \dots, x_n)$  to denote an input to  $F$ , where  $x_i$  represents the input to the  $i$ th copy of  $f$ .

The following simple lemma establishes some basic properties of dual witnesses exhibiting the fact that  $\widetilde{\text{deg}}_\varepsilon(f) > d$  or  $\text{deg}_\varepsilon^+(f) > d$ .

**Lemma 3.1.** *Let  $f: D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function,  $\varepsilon$  be a real in  $[0, 1/2)$ , and  $d$  be an integer such that  $\widetilde{\text{deg}}_\varepsilon(f) > d$ .*

*Let  $\mu: \{0, 1\}^M \rightarrow \mathbb{R}$  be a dual witness to the fact  $\widetilde{\text{deg}}_\varepsilon(f) > d$  as per Theorem 2.4. If  $f$  satisfies the stronger condition that  $\text{deg}_\varepsilon^+(f) > d$ , let  $\mu$  to be a dual witness to the fact that  $\text{deg}_\varepsilon^+(f) > d$  as per Theorem 2.6.*

*We further define  $\mu_+(x) := \max\{0, \mu(x)\}$  and  $\mu_-(x) := -\min\{0, \mu(x)\}$  to be two non-negative real functions on  $\{0, 1\}^M$ , and  $\mu_-^i$  and  $\mu_+^i$  be the restrictions of  $\mu_-$  and  $\mu_+$  on  $f^{-1}(i)$  respectively for  $i \in \{0, 1\}$ . Then the following holds:*

- $\mu_+$  and  $\mu_-$  have disjoint supports. (2)

- $\langle \mu_+, p \rangle = \langle \mu_-, p \rangle$  for any polynomial  $p$  of degree at most  $d$ . Hence,  $\|\mu_+\|_1 = \|\mu_-\|_1 = \frac{1}{2}$ . (3)

- $\|\mu_+^1\|_1 > \varepsilon$  and  $\|\mu_-^0\|_1 > \varepsilon$ . If  $\text{deg}_\varepsilon^+(f) > d$ , then  $\|\mu_+^1\|_1 = 1/2$ . (4)

The lemma follows directly from Theorem 2.4. We provide a proof in Appendix B for completeness.

### 3.2 Warm Up : A PP Lower Bound

As a warmup, we establish a simpler hardness amplification theorem for  $\text{PP}^{\text{dt}}$ .

**Theorem 3.2.** *Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function,  $n, d$  be two positive integers, and  $1/2 < \varepsilon < 1$  and  $0 < \varepsilon_2 < 1/2$  be two constants such that  $2\varepsilon_2 > \varepsilon$ . Suppose  $\widetilde{\deg}_{\varepsilon_2}(f) > d$ . Then*

$$\text{PP}^{\text{dt}}(\text{GapMaj}_{n,\varepsilon}(f)) > \Omega \left\{ \min \left( d, (2\varepsilon_2 - \varepsilon)^2 \cdot n \right) \right\}.$$

*Proof.* For  $i \in \{0, 1\}$  let  $\mu_+, \mu_-, \mu_+^i, \mu_-^i$  be functions whose existence is guaranteed by Lemma 3.1, combined with the assumption that  $\widetilde{\deg}_{\varepsilon_2}(f) > d$ .

In light of Lemma 2.8, it suffices to show that  $\widetilde{\deg}_{1/2-2^{-T}}(\text{GapMaj}_{n,\varepsilon}(f)) > T$ , for  $T = \Omega \left\{ \min \left( d, (2\varepsilon_2 - \varepsilon)^2 \cdot n \right) \right\}$ . We prove this by constructing a dual witness to this fact, as per Theorem 2.4.

We first define the following two non-negative functions on  $\{0, 1\}^{n \cdot M}$ :

$$\psi^+(x) := \prod_{i=1}^n \mu_+(x_i) \quad \text{and} \quad \psi^-(x) := \prod_{i=1}^n \mu_-(x_i).$$

Our dual witness  $\psi$  is simply their linear combination:

$$\psi := 2^{n-1} \cdot (\psi^+ - \psi^-).$$

We remark that  $\psi$  is precisely the function denoted by  $\psi_{BT}$  alluded to in Section 1.5.1. Now we verify that  $\psi$  is the dual witness we want.

**Proving the  $\psi$  has unit  $\ell_1$ -norm.** Since  $\mu_+$  and  $\mu_-$  have disjoint supports by Condition (2) of Lemma 3.1, so does  $\psi^+$  and  $\psi^-$ . Therefore  $\|\psi\|_1 = 2^{n-1} \cdot (2^{-n} + 2^{-n}) = 1$ .

**Proving the  $\psi$  has pure high degree  $d$ .** Let  $p: \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  be any monomial of degree at most  $d$ , and let  $p_i: \{0, 1\}^M \rightarrow \mathbb{R}$  be such that  $p(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i)$ . Then it holds that

$$\langle \psi^+, p \rangle = \prod_{i=1}^n \langle \mu_+, p_i \rangle = \prod_{i=1}^n \langle \mu_-, p_i \rangle = \langle \psi^-, p \rangle,$$

where the second equality holds by Condition (3) of Lemma 3.1.

As a polynomial is a sum of monomials, by linearity, it follows that  $\langle \psi, p \rangle = \langle \psi^+, p \rangle - \langle \psi^-, p \rangle = 0$  for any polynomial  $p$  with degree at most  $d$ .

**Proving that  $\psi$  has high correlation with  $F$ .** Define  $\mathcal{D}_0 := 2 \cdot \mu_-$  and  $\mathcal{D}_1 := 2 \cdot \mu_+$ . Note  $\mu_+$  and  $\mu_-$  are non-negative functions with norm  $1/2$ , so  $\mathcal{D}_0$  and  $\mathcal{D}_1$  can be thought as distributions on  $\{0, 1\}^M$ . We further define distributions  $\mathcal{U}_i$  on  $\{0, 1\}^{n \cdot M}$  for  $i \in \{0, 1\}$  as  $\mathcal{U}_i := \mathcal{D}_i^{\otimes n}$ . Observe that  $\mathcal{U}_0 = 2^n \cdot \psi^-$  and  $\mathcal{U}_1 = 2^n \cdot \psi^+$  as functions.

Then by Condition (4) of Lemma 3.1, we have  $\Pr_{x \sim \mathcal{D}_1} [f(x) = 1] = 2 \cdot \|\mu_+^1\|_1 > 2\varepsilon_2 > \varepsilon$ , and  $\Pr_{x \sim \mathcal{D}_0} [f(x) = 0] = 2 \cdot \|\mu_-^0\|_1 > 2\varepsilon_2 > \varepsilon$ .

Let  $D_F$  denote the domain of  $F$ . By the definition of  $F = \text{GapMaj}_{n,\varepsilon}(f)$  and a simple Chernoff bound, we have

$$2^n \cdot \sum_{x \in D_F} \psi^+(x) \cdot F(x) = \Pr_{x \sim \mathcal{U}_1} [F(x) = 1] \geq 1 - 2^{-c_1 \Delta^2 \cdot n}, \quad (5)$$

where  $c_1$  is a universal constant and  $\Delta := 2\varepsilon_2 - \varepsilon$ . For brevity, let  $k$  denote  $c_1 \Delta^2 \cdot n$ .

Since  $2^n \cdot \|\psi^+\|_1 = 1$ , inequality (5) further implies that

$$2^n \cdot \sum_{x \notin D_F} \psi^+(x) \leq 2^{-k}.$$

Similarly, we have

$$\Pr_{x \sim \mathcal{U}_0} [F(x) = 0] \geq 1 - 2^{-k},$$

which implies that

$$2^n \cdot \sum_{x \notin D_F} \psi^-(x) \leq 2^{-k}.$$

Putting everything together, we can calculate the correlation between  $F$  and  $\psi$  as follows:

$$\begin{aligned} & \sum_{x \in D_F} F(x)\psi(x) - \sum_{x \notin D_F} |\psi(x)| \\ & \geq 2^{n-1} \cdot \sum_{x \in D_F} \psi^+(x)F(x) - 2^{n-1} \cdot \left( \sum_{x \notin D_F} \psi^-(x) + \sum_{x \notin D_F} \psi^+(x) \right) \\ & \geq 1/2 - 2^{-k-1} - 2^{-k} \\ & > 1/2 - 2^{-k+1}. \end{aligned}$$

Setting  $T = \min(d, k-1)$ , then we can see that  $\psi$  is a dual witness for  $\widetilde{\deg}_{1-2^{-T}}(\text{GapMaj}_{n,\varepsilon}(f)) > T$ . Clearly  $T = \Omega\{\min(d, (2\varepsilon_2 - \varepsilon)^2 \cdot n)\}$ , which completes the proof.  $\square$

### 3.3 The UPP Lower Bound

The dual witness  $\psi \sim \psi^+ - \psi^-$  constructed in the previous subsection is not a dual witness for the high threshold degree of  $F = \text{GapMaj}_n(f)$  for two reasons: it puts weight on some points outside of the domain of  $F$ , and it does not satisfy the sign-agreement condition of Theorem 2.5.

In order to obtain a valid dual witness for threshold degree, we add two error correction terms  $\psi_{corr}^+$  and  $\psi_{corr}^-$  to  $\psi$ . The purpose of the error correction terms is to zero out the erroneous values, while simultaneously maintaining the high pure degree property and avoiding changing the sign of  $\psi$  on inputs at which it does not agree in sign with  $F$ . We achieve this through an error correction lemma that may be of independent interest.

**Lemma 3.3** (Error Correction Lemma). *Let  $A$  be a subset of  $\{0, 1\}^M$ , and  $\varphi$  be a function on  $\{0, 1\}^M$ . Let  $\varphi_\circ$  and  $\varphi_\times$  be the restrictions of  $\varphi$  on  $A$  and  $\{0, 1\}^M \setminus A$  respectively. That is,  $\varphi_\circ(x_i) = \varphi(x_i)$  if  $x_i \in A$  and  $\varphi_\circ(x_i) = 0$  otherwise, and similarly  $\varphi_\times(x_i) = \varphi(x_i)$  if  $x_i \notin A$  and  $\varphi_\times(x_i) = 0$  otherwise. Define  $\psi : \{0, 1\}^{n \cdot M} \rightarrow \{0, 1\}$  as  $\psi(x_1, x_2, \dots, x_n) := \prod_{i=1}^n \varphi(x_i)$ , and  $n_A(x) := \sum_{i=1}^n \mathbb{1}_{x_i \in A}$ .*

*Suppose  $\alpha = \|\varphi_\times\|_1 / \|\varphi_\circ\|_1 < 1/40$ , and let  $0.5 < \varepsilon < 1$  be a real number and  $n$  be a sufficient large integer. Then there exists a function  $\psi_{corr} : \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  such that:*

- $\psi_{corr}(x) = \psi(x)$ , when  $n_A(x) \leq \varepsilon \cdot n$ . (6)

- $|\psi_{corr}(x)| \leq \psi(x)/2$ , when  $n_A(x) > \varepsilon \cdot n$ . (7)

- $\psi_{corr}$  has pure high degree of at least  $(1 - (1 + 10\alpha) \cdot \varepsilon) \cdot n - 4$ . (8)



We defer the proof of Lemma 3.3 to Subsection 3.4. Here, we show that it implies the desired hardness amplification results.

**Theorem 3.4.** *Let  $f : D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function,  $n$  be a sufficiently large integer,  $d$  be an integer, and  $1/2 < \varepsilon < 1$  and  $0.49 < \varepsilon_2 < 1/2$  be two constants. Let  $a = \frac{2\varepsilon_2}{1 - 2\varepsilon_2}$ . Then the following holds.*

$$\text{If } \widetilde{\deg}_{\varepsilon_2}(f) > d, \text{ then } \deg_{\pm}(\text{GapMaj}_{n,\varepsilon}(f)) > \min \left( d, \left( 1 - \left( 1 + \frac{10}{a} \right) \cdot \varepsilon \right) \cdot n - 4 \right).$$

$$\text{If } \deg_{\varepsilon_2}^+(f) > d, \text{ then } \deg_{\pm}(\text{GapAND}_{n,\varepsilon}(f)) > \min \left( d, \left( 1 - \left( 1 + \frac{10}{a} \right) \cdot \varepsilon \right) \cdot n - 4 \right).$$

*Proof.* We prove both claims in the theorem by exhibiting a single dual solution that witnesses both.

As in the proof of Theorem 3.2, for  $i \in \{0, 1\}$ , let  $\mu_+, \mu_-, \mu_+^i, \mu_-^i$  denote the functions whose existence is guaranteed by Lemma 3.1, combined with the assumption that either  $\widetilde{\deg}_{\varepsilon}(f) > d$  or  $\deg_{\varepsilon}^+(f) > d$ . Also as in the proof of Theorem 3.2, define the following two non-negative functions on  $\{0, 1\}^{n \cdot M}$ :

$$\psi^+(x) := \prod_{i=1}^n \mu_+(x_i) \quad \text{and} \quad \psi^-(x) := \prod_{i=1}^n \mu_-(x_i).$$

Given an input  $x = (x_1, x_2, \dots, x_n)$ , let  $n_{\text{Yes}}(x) := \sum_{i=1}^n \mathbb{1}_{f(x_i)=1}$  and  $n_{\text{No}}(x) := \sum_{i=1}^n \mathbb{1}_{f(x_i)=0}$  as in Definition 2.10. Now apply Lemma 3.3 with the following parameters.

- Set  $A = f^{-1}(1)$ ,  $\varphi = \mu_+$ . Then for  $\alpha$  as defined in Lemma 3.3, we have  $\alpha = \frac{\|\mu_+\|_1 - \|\mu_+^1\|_1}{\|\mu_+^1\|_1} \leq \frac{1 - 2\varepsilon_2}{2\varepsilon_2} = a^{-1}$  by Conditions (3) and (4) of Lemma 3.1. Note that  $a^{-1} < 1/40$  by the assumption that  $0.49 < \varepsilon_2$ . Hence, by Lemma 3.3, there exists a function  $\psi_{\text{corr}}^+ : \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  such that:

- $\psi_{\text{corr}}^+(x) = \psi^+(x)$ , for all  $x$  such that  $n_{\text{Yes}}(x) \leq \varepsilon \cdot n$  (9)

- $|\psi_{\text{corr}}^+(x)| \leq \psi^+(x)/2$ , for all  $x$  such that  $n_{\text{Yes}}(x) > \varepsilon \cdot n$  (10)

- $\psi_{\text{corr}}^+$  has pure high degree at least  $\left( 1 - \left( 1 + \frac{10}{a} \right) \cdot \varepsilon \right) \cdot n - 4$  (11)

- Similarly, set  $A = f^{-1}(0)$ ,  $\varphi = \mu_-$ . Again by Lemma 3.3, there exists a function  $\psi_{\text{corr}}^- : \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  such that:

- $\psi_{\text{corr}}^-(x) = \psi^-(x)$ , for all  $x$  such that  $n_{\text{No}}(x) \leq \varepsilon \cdot n$  (12)

- $|\psi_{\text{corr}}^-(x)| \leq \psi^-(x)/2$ , for all  $x$  such that  $n_{\text{No}}(x) > \varepsilon \cdot n$  (13)

- $\psi_{\text{corr}}^-$  has pure high degree of at least  $\left( 1 - \left( 1 + \frac{10}{a} \right) \cdot \varepsilon \right) \cdot n - 4$  (14)

For convenience, let  $N = \left(1 - \left(1 + \frac{10}{a}\right) \cdot \varepsilon\right) \cdot n - 4$ . We are ready to construct the dual witness  $\psi$  that establishes the claimed threshold degree lower bounds. Define  $\psi: \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  by

$$\psi := (\psi^+ - \psi_{\text{corr}}^+) - (\psi^- - \psi_{\text{corr}}^-).$$

We first establish two properties of  $\psi$ .

- When  $n_{\text{Yes}}(x) \geq \varepsilon \cdot n$ ,  $\psi(x) = \psi^+(x) - \psi_{\text{corr}}^+(x) \geq \psi^+(x)/2 \geq 0$  (15)
- When  $n_{\text{No}}(x) \geq \varepsilon \cdot n$ ,  $\psi(x) = -(\psi^-(x) - \psi_{\text{corr}}^-(x)) \leq -\psi^-(x)/2 \leq 0$  (16)

**Verifying Condition (15) and (16).** To establish that Condition (15) holds, observe that since  $n_{\text{Yes}}(x) \geq \varepsilon \cdot n$ , and  $\varepsilon > 1/2$  by assumption, it follows that  $n_{\text{No}}(x) \leq (1 - \varepsilon) \cdot n \leq \varepsilon \cdot n$ . This implies that  $\psi^-(x) = \psi_{\text{corr}}^-(x)$  by Condition (12) and  $|\psi_{\text{corr}}^+(x)| \leq \psi^+(x)/2$  by Condition (10). Then  $\psi(x) = \psi^+(x) - \psi_{\text{corr}}^+(x) \geq \psi^+(x)/2 \geq 0$ , where the last inequality follows from the fact that  $\psi^+$  is non-negative.

Similarly, for Condition (16), as  $n_{\text{No}}(x) \geq \varepsilon \cdot n$ , it follows that  $n_{\text{Yes}}(x) \leq (1 - \varepsilon) \cdot n \leq \varepsilon \cdot n$ . This implies that  $\psi^+(x) = \psi_{\text{corr}}^+(x)$  by Condition (9) and  $|\psi_{\text{corr}}^-(x)| \leq \psi^-(x)/2$  by Condition (13). Note  $\psi^-$  is also non-negative. Hence  $\psi(x) = -(\psi^-(x) - \psi_{\text{corr}}^-(x)) \leq -(\psi^-(x)/2) \leq 0$ .

We now verify that  $\psi$  is a dual witness for  $\deg_{\pm}(F) > \min(d, N)$  (recall that  $F$  denotes  $\text{GapMaj}(f)$ ).

**Analyzing the pure high degree of  $\psi$ .** Write  $\psi := \psi^+ - \psi^- - \psi_{\text{corr}}^+ + \psi_{\text{corr}}^-$ . We already established that  $\psi^+ - \psi^-$  has pure high degree  $d$  in the proof of Theorem 3.2, and both  $\psi_{\text{corr}}^+$  and  $\psi_{\text{corr}}^-$  have pure high degree at least  $N$  (cf. Conditions (11) and (14)). By linearity,  $\psi$  itself has pure high degree at least  $\min(d, N)$ .

**Showing that the support of  $\psi$  is a subset of the inputs on which  $F$  is defined.** Let  $x$  be an input outside of the domain of  $F$ . Then by the definition of  $\text{GapMaj}$ , it must be the case that both  $n_{\text{Yes}}(x)$  and  $n_{\text{No}}(x)$  are strictly less than  $\varepsilon \cdot n$ . This means that  $\psi^+(x) = \psi_{\text{corr}}^+(x)$  and  $\psi^-(x) = \psi_{\text{corr}}^-(x)$  by Conditions (9) and (12), and hence  $\psi(x) = 0$ . Therefore, the support of  $\psi$  is a subset of the domain of  $F$ .

**Showing that  $\psi$  agrees in sign with  $F$ .** When  $F(x) = 1$ , by the definition of  $\text{GapMaj}$ , we have  $n_{\text{Yes}}(x) \geq \varepsilon \cdot n$ . Then  $\psi(x) \geq 0$  follows directly from Condition (15). Similarly, when  $F(x) = 0$ , we have  $n_{\text{No}}(x) \geq \varepsilon \cdot n$  and  $\psi(x) \leq 0$  by Condition (16). Therefore,  $\psi$  agrees in sign with  $F$ .

**Showing that  $\psi$  is non-trivial.** Pick an input  $x_0$  to  $f$  such that  $\mu_+^1(x_0) > 0$ , and let  $x = (x_0, x_0, \dots, x_0)$ . Then we have  $f(x_0) = 1$  and  $n_{\text{Yes}}(x) = n \geq \varepsilon \cdot n$ . Therefore,  $\psi(x) = \psi^+(x) - \psi_{\text{corr}}^+(x) \geq \psi^+(x)/2 = (\mu_+^1(x_0))^n/2 > 0$  by Condition (15). So  $\psi$  is non-trivial.

Putting everything together and invoking Theorem 2.5 proves the first claim of Theorem 3.4.

**Showing  $\psi$  is also a dual witness for  $\text{GapAND}_{n,\varepsilon}(f)$ .** Now we show that, when  $\deg_{\varepsilon_2}^+(f) > d$ , the same function  $\psi$  is also a dual witness for  $\deg_{\pm}(\text{GapAND}_{n,\varepsilon}(f)) > \min(d, N)$ .

We already proved that the pure high degree of  $\psi$  is as claimed, and that it is non-trivial. So it remains to verify  $\psi$  only puts weight in the domain of  $\text{GapAND}_{n,\varepsilon}(f)$ , and that  $\psi$  agrees in sign with  $\text{GapAND}_{n,\varepsilon}(f)$ .

By Condition (4) of Lemma 3.1, we have  $|\mu_+^1| = |\mu_+| = \frac{1}{2}$ , which means  $\mu_+$  only puts weight inputs in  $f^{-1}(1)$ . So  $\psi^+$  only takes non-zero values when  $n_{\text{Yes}}(x) = n$ . Also, note that when  $n_{\text{No}}(x) \leq \varepsilon \cdot n$ , we have  $\psi^-(x) = \psi_{\text{corr}}^-(x)$  by Condition (12). Therefore,  $\psi$  only puts weight on inputs when  $n_{\text{Yes}}(x) = n$  or  $n_{\text{No}}(x) > \varepsilon \cdot n$ . All such inputs are in the domain of  $\text{GapAND}_{n,\varepsilon}(f)$ .

Finally, we verify that  $\psi$  agrees in sign with  $\text{GapAND}_{n,\varepsilon}(f)$ . When  $\text{GapAND}_{n,\varepsilon}(f)(x) = 1$ , we have  $n_{\text{Yes}}(x) = n \geq \varepsilon \cdot n$ , hence  $\psi(x) \geq 0$  by Condition (15). When  $\text{GapAND}_{n,\varepsilon}(f)(x) = 0$ , we have  $n_{\text{No}}(x) \geq \varepsilon \cdot n$ , so  $\psi(x) \leq 0$  follows immediately from Condition (16). Applying Theorem 2.5 again, this completes the proof for the second claim of Theorem 3.4. □

### 3.4 Proof of the Error Correction Lemma

In this subsection we prove Lemma 3.3. We need two lemmas. In the first, we construct a polynomial with certain properties.

**Lemma 3.5.** *Let  $a \geq 40$ ,  $n$  be a sufficiently large integer, and  $\varepsilon$  be a real such that  $0.5 < \varepsilon < 1$ . Then there exists an (explicitly given) univariate polynomial  $P: \mathbb{R} \rightarrow \mathbb{R}$  such that:*

- $P(x) = (-a)^x$  for  $x \in \{0, \dots, \varepsilon \cdot n\}$ .
- $|P(x)| \leq a^x/2$  for  $x \in \{\varepsilon \cdot n + 1, \dots, n\}$ .
- $P$  has degree of at most  $\left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3$ .

We prove Lemma 3.5 by defining  $P$  via interpolation through carefully chosen values – we defer this proof to Appendix B.

**Lemma 3.6.** *Let  $x = (x_1, \dots, x_n) \in \{0, 1\}^{M \cdot n}$ , and let  $n_A(x) := \sum_{i=1}^n \mathbb{1}_{x_i \in A}$ . Let  $\psi$  and  $\alpha$  be as defined in the statement of Lemma 3.3. For any univariate polynomial  $P: \mathbb{R} \rightarrow \mathbb{R}$  of degree at most  $d$ , the following function on  $\{0, 1\}^{n \cdot M}$  has pure high degree  $n - d - 1$ :*

$$\psi_P(x) := \alpha^{n_A(x)} \cdot \psi(x) \cdot P(n_A(x)).$$

Before proving Lemma 3.6, we show that it and Lemma 3.5 together imply Lemma 3.3.

*Proof of Lemma 3.3.* We first deal with the special case that  $\alpha = 0$ . In this case, we note  $\psi(x) > 0$  only if  $n_A(x) = n$ . So it suffices to let  $\psi_{\text{corr}}$  be the constant function  $\mathbf{0}$ .

From now on, we assume  $\alpha > 0$ . Let  $a = \alpha^{-1}$  and  $N = (1 + 10\alpha) \cdot \varepsilon \cdot n + 3 = \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3$ .

**Construction of  $\psi_{\text{corr}}$ .** Applying Lemma 3.5, we obtain a polynomial  $P(x)$  such that:

$$\bullet \quad P(x) = (-a)^x \text{ for } x \in \{0, \dots, \varepsilon \cdot n\}. \tag{17}$$

$$\bullet \quad |P(x)| \leq a^x/2 \text{ for } x \in \{\varepsilon \cdot n + 1, \dots, n\}. \tag{18}$$

$$\bullet \quad P \text{ has a degree of at most } \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3 = N. \tag{19}$$

We define  $\psi_{\text{corr}} := \psi_P$ , where  $\psi_P$  is as defined in Lemma 3.6.

**Verification that  $\psi_{\text{corr}}$  satisfies the properties claimed in Lemma 3.3.** Given an input  $x = (x_1, \dots, x_n)$ , let  $m = n_A(x)$  for brevity. We now verify Condition (6), (7) and (8) of Lemma 3.3 in order. When  $m \leq \varepsilon \cdot n$ , we have  $P(m) = (-a)^m$  by Condition (17), hence

$$\psi_{\text{corr}}(x) = (-\alpha)^m \cdot (-a)^m \cdot \psi(x) = \psi(x),$$

so Condition (6) holds. When  $m > \varepsilon \cdot n$ , as  $|P(m)| \leq a^m/2$  by Condition (18), we have  $|\psi_{\text{corr}}(x)| \leq \alpha^m \cdot a^m/2 \cdot \psi(x) = \psi(x)/2$ , which establishes Condition (7). Finally, since  $P$  is of degree at most  $N$  as guaranteed by Condition (19), Lemma 3.6 implies that  $\psi_{\text{corr}}$  has pure high degree of at least  $n - N - 1 = \left(1 - \left(1 + \frac{10}{a}\right) \cdot \varepsilon\right) \cdot n - 4$ . So Condition (8) is verified, and this completes the proof.  $\square$

Finally, we prove Lemma 3.6.

*Proof of Lemma 3.6.* We begin by constructing some useful auxiliary functions.

**Definition and analysis of auxiliary functions**  $\psi_k : \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$ . Recall the definitions of  $\varphi_o$  and  $\varphi_\times$  from the statement of Lemma 3.3:  $\varphi_o(x_i) = \varphi(x_i)$  if  $x_i \in A$  and  $\varphi_o(x_i) = 0$  otherwise, and  $\varphi_\times(x_i) = \varphi(x_i)$  if  $x_i \notin A$  and  $\varphi_\times(x_i) = 0$  otherwise. For each integer  $k \in \{0, \dots, n\}$ , we define

$$\psi_k(x) := \sum_{S \subseteq [n], |S|=k} \left( \prod_{i \in S} \varphi(x_i) \cdot \prod_{i \notin S} (\varphi_\times - \alpha \cdot \varphi_o)(x_i) \right). \quad (20)$$

We claim that:

$$\psi_k \text{ has pure high degree at least } n - k - 1. \quad (21)$$

To establish this, it suffices to show that for every  $|S| = k$ , the following function

$$\psi_S(x) := \prod_{i \in S} \varphi(x_i) \cdot \prod_{i \notin S} (\varphi_\times - \alpha \cdot \varphi_o)(x_i) \quad (22)$$

has pure high degree at least  $n - k - 1$ , as  $\psi_k$  is simply a sum of  $\psi_S$ 's with  $|S| = k$ .

Let  $[n] := \{1, \dots, n\}$ ,  $p : \{0, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  be any monomial of degree at most  $n - k - 1$ , and let  $p_i : \{0, 1\}^M \rightarrow \mathbb{R}$  be such that  $p(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i)$ . Then  $\sum_{i=1}^n \deg(p_i) = \deg(p) \leq n - k - 1$ , which means there are at least  $k + 1$   $p_i$ 's which have degree zero, i.e., are constant functions. Since  $k + 1 + |[n] \setminus S| = n + 1 > n$ , there must exist an index  $i^*$  such that  $p_{i^*}$  is a constant function  $p_{i^*}(x) \equiv c$ , and  $i^* \notin S$ . Now, since  $\varphi_\times$  and  $\varphi_o$  have disjoint supports and  $\alpha = \|\varphi_\times\|_1 / \|\varphi_o\|_1$  by definition, we have  $\langle \varphi_\times - \alpha \cdot \varphi_o, p_{i^*} \rangle = (\|\varphi_\times\|_1 - \alpha \cdot \|\varphi_o\|_1) \cdot c = 0$ . Therefore, by Equation (22),

$$\langle \psi_S, p \rangle = \langle \varphi_\times - \alpha \cdot \varphi_o, p_{i^*} \rangle \cdot \prod_{i \in S} \langle \varphi, p_i \rangle \cdot \prod_{i \notin S, i \neq i^*} \langle \varphi_\times - \alpha \cdot \varphi_o, p_i \rangle = 0.$$

As a polynomial is a sum of monomials, by linearity, we conclude that  $\psi_S$  has pure high degree  $n - k - 1$  for all  $|S| = k$ , and so does  $\psi_k$ .

Now, fix an input  $x = (x_1, x_2, \dots, x_n)$ . Write  $m = n_A(x)$  (note that  $m$  is actually a function of  $x$ ). We are going to re-express  $\psi_k(x)$  in a convenient form, as follows. We assume that the first  $m$  inputs  $x_1, x_2, \dots, x_m$  satisfy  $x_i \in A$  – this is without loss of generality by symmetry.

Fix a set  $S \subseteq [n]$  with  $|S| = k$ . We claim that

$$\psi_S(x) = (-\alpha)^{m - |S \cap [m]|} \cdot \prod_{i=1}^n \varphi(x_i).$$

To see this, first consider  $i \in S \cap [m]$ . Then there is a factor  $\varphi(x_i)$  appearing in  $\psi_S(x)$ . Now fix an  $i \in ([m] \setminus S)$ . Then there is a factor of  $(\varphi_\times - \alpha \cdot \varphi_o)(x_i) = (-\alpha) \cdot \varphi(x_i)$  appearing in  $\psi_S(x)$ . Finally, fix any  $i \notin [m]$ . If  $i \in S$ , then there is a factor  $\varphi(x_i)$  appearing in  $\psi_S(x)$ , and if  $i \notin S$ , the factor appearing in  $\psi_S(x)$  is  $(\varphi_\times(x_i) - \alpha \cdot \varphi_o(x_i)) = \varphi_\times(x_i) = \varphi(x_i)$ .

Hence, we may write

$$\begin{aligned}
\psi_k(x) &= \sum_{S \subseteq [n], |S|=k} (-\alpha)^{|S \cap [m]|} \cdot \prod_{i=1}^n \varphi(x_i) \\
&= \prod_{i=1}^n \varphi(x_i) \cdot \left( \sum_{j=0}^k \binom{m}{j} \cdot \binom{n-m}{k-j} \cdot (-\alpha)^{m-j} \right) \\
&= (-\alpha)^m \cdot \psi(x) \cdot \left( \sum_{j=0}^k \binom{m}{j} \cdot \binom{n-m}{k-j} \cdot (-\alpha)^{-j} \right). \tag{23}
\end{aligned}$$

**Definition and analysis of auxiliary univariate polynomials**  $P_k: \mathbb{R} \rightarrow \mathbb{R}$ . Let

$$P_k(m) := \sum_{i=0}^k \binom{m}{i} \cdot \binom{n-m}{k-i} \cdot (-\alpha)^{-i}.$$

Then  $P_k(m)$  is polynomial in  $m$  of degree at most  $k$ , and by Equation (23),

$$\psi_k(x) = (-\alpha)^m \cdot \psi(x) \cdot P_k(m). \tag{24}$$

Expanding the binomial coefficients, we have

$$P_k(m) = \sum_{i=0}^k \frac{\prod_{j=0}^{i-1} (m-j)}{i!} \cdot \frac{\prod_{j=0}^{k-i-1} (n-m-j)}{(k-i)!} \cdot (-\alpha)^{-i}. \tag{25}$$

Observe that the coefficient of  $m^k$  in Equation (25) is

$$\begin{aligned}
&\sum_{i=0}^k \frac{1}{i! \cdot (k-i)!} \cdot (-\alpha)^{-i} \cdot (-1)^{k-i} \\
&= (-1)^k \cdot \frac{1}{k!} \cdot \sum_{i=0}^k \alpha^{-i} \cdot \binom{k}{i} \\
&= (-1)^k \cdot \frac{1}{k!} \cdot (1 + \alpha^{-1})^k \neq 0,
\end{aligned}$$

where the second equality follows from the equation  $(a+b)^k = \sum_{i=0}^k a^i b^{k-i} \cdot \binom{k}{i}$ , and the last inequality follows because  $\alpha \geq 0$  by definition.

So  $P_k(m)$  is a polynomial of degree exactly  $k$ . Therefore, the set  $\{P_k(m)\}_{k=0}^d$  generates all the polynomials in  $m$  with degree at most  $d$ .

**Verification that  $\psi_P$  has the pure high degree claimed in Lemma 3.6.** Let  $P: \mathbb{R} \rightarrow \mathbb{R}$  be a polynomial of degree at most  $d$ . By the previous paragraph, we may write  $P(m)$  as

$$P(m) = \sum_{k=0}^d \beta_k \cdot P_k(m), \tag{26}$$

for some real numbers  $\beta_0, \dots, \beta_d$ .

Then we have

$$\begin{aligned}
\psi_P(x) &= \alpha^{m(x)} \cdot \psi(x) \cdot P(m(x)) \\
&= \alpha^{m(x)} \cdot \psi(x) \cdot \sum_{k=0}^d (\beta_k \cdot P_k(m(x))) \\
&= \sum_{k=0}^d \beta_k \cdot \psi_k(x).
\end{aligned}$$

Here, the first equality holds by definition of  $\psi_P$ , the second by Equation (26), and the third by Equation (24).

Each  $\psi_k$  appearing in the above sum has pure high degree at least  $n - d - 1$  by Property (21). Hence, by linearity,  $\psi_P$  has pure high degree of  $n - d - 1$ . This completes the proof.  $\square$

## 4 NISZK<sup>O</sup> $\not\subseteq$ UPP<sup>O</sup>

In this section we construct an oracle  $\mathcal{O}$  such that NISZK<sup>O</sup>  $\not\subseteq$  UPP<sup>O</sup>. We will use the function  $\text{GCol}_n := \text{GapMaj}_{n^{1/4}, 1 - \frac{1}{3 \log n}}(\text{Col}_{n^{3/4}}^{3 \log n})$  to attain the desired oracle separation.

We first show that its complement  $\overline{\text{GCol}_n}$  is easy for NISZK by providing a reduction from it to the statistical distance from uniform (SDU) problem. SDU is complete for NISZK and so has an NISZK protocol [GSV99]. We first introduce the problem SDU.

**Definition 4.1** (Statistical Distance from Uniform (SDU) [GSV99]). The promise problem Statistical Distance from Uniform, denoted  $\text{SDU} = (\text{SDU}_{\text{YES}}, \text{SDU}_{\text{NO}})$ , consisted of

$$\begin{aligned}
\text{SDU}_{\text{YES}} &= \{X : \|X - U\| < 1/n\} \\
\text{SDU}_{\text{NO}} &= \{X : \|X - U\| > 1 - 1/n\}
\end{aligned}$$

where  $X$  is a distribution encoded as a circuit outputting  $n$  bits, and  $U$  is the uniform distribution on  $n$  bits, and  $\|X - U\|$  denotes the statistical distance between  $X$  and  $U$ .

**Theorem 4.2.** *There is a polylog( $n$ )-time NISZK protocol for  $\overline{\text{GCol}_n}$ .*

*Proof.* For simplicity, we assume  $n$  is a power of 2. We prove this theorem by showing a reduction from  $\overline{\text{GCol}_n}$  to an instance of SDU with distributions on  $\log n$  bits.

Now, let  $m = n^{1/4}$ ,  $k = n^{3/4}$  and  $x = (f_1, f_2, \dots, f_m)$  be an input to  $\text{GCol}_n$ , where each  $f_i$  is interpreted as a function from  $[k] \rightarrow [k]$ . We construct the distribution  $\mathcal{D}(x)$  as follows: to generate a sample from  $\mathcal{D}(x)$ , we pick a pair  $(i, j) \in [m] \times [k]$  at uniformly random, and output the sample  $(i, f_i(j))$ . Clearly  $\mathcal{D}(x)$  is polylog( $n$ )-time preparable.

Now we show this is a valid reduction. Let  $\mathcal{U}$  be the uniform distribution on  $[m] \times [k]$  and  $\mathcal{U}_k$  be the uniform distribution on  $[k]$ . For a function  $f : [k] \rightarrow [k]$ , let  $\mathcal{D}_f$  be the distribution obtained by outputting  $f(i)$  for an index  $i \sim \mathcal{U}_k$ . Then we can see  $\mathcal{D}(x) = \frac{1}{m} \sum_{i=1}^m \{i\} \times \mathcal{D}_{f_i}$ .

When  $\overline{\text{GCol}_n}(x) = 1$ , we have

$$\|\mathcal{D}(x) - \mathcal{U}\| = \frac{1}{m} \sum_{i=1}^m \|\mathcal{U}_k - \mathcal{D}_{f_i}\| \leq \frac{1}{3 \log n} < \frac{1}{\log n}.$$

Here, the first inequality holds because at least a  $1 - \frac{1}{3 \log n}$  fraction of  $f_i$ 's are permutations, which implies that  $\|\mathcal{U}_k - \mathcal{D}_{f_i}\| = 0$ .

When  $\overline{\text{GCol}}_n(x) = 0$ , we have

$$\|\mathcal{D}(x) - \mathcal{U}\| = \frac{1}{m} \sum_{i=1}^m \|\mathcal{U}_k - \mathcal{D}_{f_i}\| \geq \left(1 - \frac{1}{3 \log n}\right) \cdot \left(1 - \frac{1}{3 \log n}\right) > 1 - \frac{1}{\log n}.$$

Here, the first inequality holds because at least a  $1 - \frac{1}{3 \log n}$  fraction of  $f_i$ 's are 3 log  $n$ -to-1, which implies that  $\|\mathcal{U}_k - \mathcal{D}_{f_i}\| = 1 - \frac{1}{3 \log n}$ .

Putting everything together, we have shown  $\mathcal{D}(x)$  that is a valid reduction to SDU. This completes the proof.  $\square$

Then by a straightforward application of Theorem 3.2, we can show  $\text{GCol}_n$  is hard for any UPP algorithm.

**Theorem 4.3.**  $\text{UPP}^{\text{dt}}(\text{GCol}_n) = \Omega(n^{1/4}/\log n)$ .

*Proof.* Observe that  $\widetilde{\text{deg}}_{1/2 - \frac{1}{50 \log n}}(\text{Col}_{n^{3/4}}^{3 \log n}) = \Omega(n^{1/4}/\log^{2/3} n)$  by Theorem 2.14. Applying Theorem 3.4 with  $a = \frac{1 - 2 \cdot \frac{1}{50 \log n}}{2 \cdot \frac{1}{50 \log n}} = 25 \log n - 1$  (recall  $a = \frac{2\varepsilon_2}{1 - 2\varepsilon_2}$  in Theorem 3.4), we have that

$$\begin{aligned} & \text{UPP}^{\text{dt}}\left(\text{GapMaj}_{n^{1/4}, 1 - \frac{1}{3 \log n}}(\text{Col}_{n^{3/4}}^{3 \log n})\right) \\ & \geq \text{deg}_{\pm}\left(\text{GapMaj}_{n^{1/4}, 1 - \frac{1}{3 \log n}}(\text{Col}_{n^{3/4}}^{3 \log n})\right) \quad (\text{by Lemma 2.9}) \\ & \geq \min\left\{\left(1 - \left(1 + \frac{10}{a}\right) \cdot \left(1 - \frac{1}{3 \log n}\right)\right) \cdot n^{1/4} - 4, \widetilde{\text{deg}}_{1/2 - \frac{1}{50 \log n}}(\text{Col}_{n^{3/4}}^{3 \log n})\right\} \\ & \geq \Omega(n^{1/4}/\log n). \end{aligned}$$

$\square$

Now Theorem 1.1 from the introduction follows from standard diagonalization methods and the observation that UPP is closed under complement.

## 5 Limitations on Perfect Zero Knowledge Proofs (Proof of Theorem 1.2)

In this section, we study the limitations of perfect zero knowledge in the presence of oracles.

**Definition 5.1** (Honest Verifier Perfect Zero Knowledge). A promise problem  $L = (L_Y, L_N)$  is in HVPZK if there exist a tuple of Turing machines  $(P, V, S)$ , where the *verifier*  $V$  and *simulator*  $S$  run in probabilistic polynomial time, satisfying the following:

- $(P, V)$  is an interactive proof for  $L$  with negligible completeness and soundness errors.
- $S(x)$  is allowed to fail (by outputting  $\perp$ ), but with probability at most  $1/2$ .

- For any  $x \in L$ , let  $\hat{S}(x)$  denote the distribution of  $S(x)$  conditioned on it not failing. Then,

$$\|\hat{S}(x) - (P, V)(x)\| = 0$$

The class PZK is defined similarly but, as in the case of SZK, with the additional stipulation that for any verifier that deviates from the protocol, there is a simulator that simulates the prover's interaction with that verifier. It is easy to see that  $\text{PZK} \subseteq \text{HVPZK}$  in the presence of any oracle. (The definitions of these classes in the presence of an oracle are the same, except that  $P$ ,  $V$ , and  $S$  all have access to the oracle.)

Note that the probability of failure of the simulator can be made negligible (in  $|x|$ ) by repeating it a polynomial number of times and taking its output to be that from the first time that it succeeds. We use this implicitly in the rest of our development. The variant where the simulator is not allowed to fail is called Super-Perfect Zero Knowledge by Goldreich and Teichner [GT14]. There (and also elsewhere), this definition is considered to be “oversimplified” as such proof systems are not known for problems outside BPP. However, in the setting of honest verifiers with small but non-zero completeness error, the class thus defined turns out to be equal to HVPZK [GT14]. While sometimes these classes are defined with the requirement of perfect completeness in the zero knowledge proofs, note that defining them as above only makes our results stronger – requiring perfect completeness can only make HVPZK smaller, and our oracle separation between PZK and coPZK continues to hold when both these classes are defined with perfect completeness.

## 5.1 A Preliminary Lemma

We will need the following lemma in the proof of the theorems that follow.

**Lemma 5.2.** *There is an oracle Turing Machine  $M_2$  that is such that when given sample access to two distributions  $p$  and  $q$ ,  $M_2$  uses two samples and,*

$$\Pr[M_2^{p,q} \text{ accepts}] = \frac{1}{2} + \frac{\|p - q\|_2^2}{8}$$

*Proof.*  $M_2^{p,q}$  behaves as follows:

1. With probability  $\frac{1}{4}$ , sample  $y_1, y_2$  from  $p$ .
  - If  $y_1 = y_2$ , accept with probability 1.
  - Else, accept with probability  $\frac{1}{2}$ .
2. With probability  $\frac{1}{4}$ , do the same with samples from  $q$ .
3. With probability  $\frac{1}{2}$ , sample  $y_1$  from  $p$  and  $y_2$  from  $q$ .
  - If  $y_1 = y_2$ , reject with probability 1.
  - Else, accept with probability  $\frac{1}{2}$ .

$$\begin{aligned} \Pr[M_2^{p,q} \text{ accepts}] &= \frac{1}{4} \left[ (1 - \|p\|_2^2) \frac{1}{2} + \|p\|_2^2 \right] + \frac{1}{4} \left[ (1 - \|q\|_2^2) \frac{1}{2} + \|q\|_2^2 \right] + \frac{1}{2} \left[ (1 - \langle p, q \rangle) \frac{1}{2} + \langle p, q \rangle \cdot 0 \right] \\ &= \frac{1}{2} + \frac{\|p - q\|_2^2}{8} \end{aligned}$$

□



## 5.2 Showing $\text{HVPZK} \subseteq \text{PP}$ Relative to Any Oracle

The first step in our proof of Theorem 1.2 is to show that HVPZK is contained in PP in a relativizing manner.

**Theorem 5.3.**  $\text{HVPZK} \subseteq \text{PP}$ . *Further, this is true in the presence of any oracle.*

*Proof.* Let  $L$  be a language with an HVPZK proof system  $(P, V, S)$ . We will show how to decide membership in  $L$  in PP. Fix any input length  $n$ , and let the number of messages in the proof system for any input of this length be  $m$ , and the length of each message be  $\ell$  (these are without loss of generality). Also suppose that the first message is always sent by the verifier. Let the number of random bits used by  $V$  on an input of length  $n$  be  $v$ , and the number of random bits used by  $S$  be  $s$ .

For any  $x \in \{0, 1\}^n$ , we write the output of the simulator  $S$  on input  $x$  using randomness  $r$  as  $S(x; r) = (R_V(x; r), T_1(x; r), \dots, T_m(x; r))$ , where  $R_V$  is the simulated randomness of the verifier, and  $T_i$  is the simulated  $i$ th message in the protocol. Let  $S_i$  denote  $S$  truncated at  $T_i$ . Denote by  $V_S$  the verifier simulated by  $S$ , and by  $P_S$  the simulated prover, both conditioned on the simulator not failing.

**Claim 1.** *An input  $x$  is in  $L$  if and only if the following three conditions are satisfied:*

1.  $V_S$  on input  $x$  behaves like the actual verifier  $V$ . This involves the following:
  - $R_V(x)$ , conditioned on not being  $\perp$ , is distributed uniformly over  $\{0, 1\}^v$ .
  - For any non-failing transcript  $(r_V, t_1, \dots, t_m)$  output by  $S(x)$ , the verifier's responses in  $(t_1, \dots, t_m)$  are consistent with what  $V$  would have sent when using  $r_V$  as randomness.
2.  $P_S$  on input  $x$  is a valid prover.
  - This means that the distribution of the prover's simulated messages  $(T_{2i}(x))$  should depend only on the messages in the transcript so far  $(T_1(x), \dots, T_{2i-1}(x))$ , and should be independent of the verifier's simulated randomness  $(R_V(x))$ .
3.  $S(x)$  is an accepting transcript with probability at least  $3/4$ .

For any  $x \in L$ , the transcript of the actual protocol satisfies the above properties, and so does the simulation, since it is perfect conditioned on not failing.

The other direction follows on noting that if all three conditions are satisfied for some  $x$ , then  $P_S$  is a prover strategy that convinces the actual verifier  $V$  that  $x \in L$ . By the soundness of the  $(P, V)$  proof system, this can only happen if  $x$  is indeed in  $L$ .

So to decide the membership of  $x$  in  $L$  in PP, it is sufficient to be able to decide each of the above three properties of  $S(x)$  in PP (since PP is closed under conjunction [BRS91]). Of these, property (3) is easily seen to be decidable in BPP, and hence in PP.

Lemma 5.2 says, in particular, that testing whether two polynomial-time-samplable distributions  $p$  and  $q$  are identical can be done in PP. Let  $U_S(x)$  be the distribution sampled by first running  $S(x)$ , outputting  $\perp$  if it fails, and a uniform sample from  $\{0, 1\}^v$  if it doesn't. The first check on  $V_S$  is the same as checking whether  $R_V(x)$  is identical to  $U_S(x)$ . The other check required on  $V_S$  is a coNP statement, and hence can be done in PP.

Let  $M_2$  be the TM from Lemma 5.2. To check that  $P_S$  is a valid prover, consider the TM – call it  $M_P$  – that works as follows on input  $x$ .

1. Select  $i \in \left[-1, \frac{m}{2}\right]$  at random.
2. If  $i = 0$ , run  $M_2$  on the distributions  $R_V(x)$  and  $U_S(x)$ .

3. If  $i = -1$ , check the consistency of transcripts produced by  $S(x)$  with the simulated randomness.
  - This is done by selecting  $r_S \in \{0, 1\}^s$ , and running  $S(x; r_S)$  to get  $(r_V, t_1, \dots, t_m)$ .
  - If this transcript is failing or consistent, accept with probability  $1/2$ , else with probability  $1$ .
4. Else, select at random  $t_1, \dots, t_{2i} \in \{0, 1\}^\ell$ ,  $r_V^1, r_V^2 \in \{0, 1\}^v$ , and  $r_S^1, r_S^2 \in \{0, 1\}^s$ .
5. If  $S(x; r_S^1)$  does not have  $(r_V^1, t_1, \dots, t_{2i-1})$  as a prefix or  $S(x; r_S^2)$  does not have  $(r_V^2, t_1, \dots, t_{2i-1})$  as a prefix, accept with probability  $1/2$ .
6. Let  $p$  be the distribution over  $\{0, 1\}$  such that  $p(1) = \Pr[S_{2i}(x) = (r_V^1, t_1, \dots, t_{2i})]$ , and  $q$  be the same but with  $r_V^2$  instead of  $r_V^1$ .
7. Run  $M_2$  on the distributions  $p$  and  $q$ .

**Claim 2.**  $M_P(x)$  accepts with probability at most  $\frac{1}{2}$  if and only if  $V_S$  is a valid verifier and  $P_S$  is a valid prover on input  $x$ .

Suppose  $V_S$  is a valid verifier and  $P_S$  is a valid prover on input  $x$ . If  $M_P$  selects  $i = 0$  or  $i = -1$ , then it accepts with probability  $\frac{1}{2}$  because  $V_S$  is a valid verifier.

If  $i \notin \{-1, 0\}$ , and  $M_P$  picks  $r_V^1, r_V^2, t_1, \dots, t_{2i}$ . If this fails the check in step 5, then  $M_P$  again accepts with probability  $1/2$ . If this does not happen and  $r_V^1, r_V^2, t_1, \dots, t_{2i-1}$  are in the support of  $S_{2i-1}(x)$ ,

$$\begin{aligned} \Pr[S_{2i}(x) = (r_V^1, t_1, \dots, t_{2i})] \\ &= \Pr[S_{2i-1}(x) = (r_V^1, t_1, \dots, t_{2i-1})] \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^1, t_1, \dots, t_{2i-1})] \\ &= \Pr[S_{2i-1}(x) = (r_V^1, t_1, \dots, t_{2i-1})] \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^2, t_1, \dots, t_{2i-1})] \end{aligned}$$

where the second equality is because  $P_S$  is a valid prover, so its responses do not depend on the simulated randomness of the verifier. We can write the first term in the product above as:

$$\begin{aligned} \Pr[S_{2i-1}(x) = (r_V^1, t_1, \dots, t_{2i-1})] \\ &= \Pr[S_{2i-2}(x) = (r_V^1, t_1, \dots, t_{2i-2})] \Pr[T_{2i-1}(x) = t_{2i-1} \mid S_{2i-2}(x) = (r_V^1, t_1, \dots, t_{2i-2})] \\ &= \Pr[S_{2i-2}(x) = (r_V^1, t_1, \dots, t_{2i-2})] \end{aligned}$$

where the second equality is because  $V_S$  is a valid verifier and is deterministic once  $R_V$  is fixed, and step 5 was there precisely to check that this probability is non-zero.

Now starting from the fact that  $\Pr[S_0(x) = r_V^1] = \Pr[S_0(x) = r_V^2]$ , and using the above relationships, we can inductively prove that  $\Pr[S_{2i}(x) = (r_V^1, t_1, \dots, t_{2i})] = \Pr[S_{2i}(x) = (r_V^2, t_1, \dots, t_{2i})]$ . This implies that the call to  $M_2$  in step 7 of  $M_P$  accepts with probability  $1/2$ , as the distributions  $p$  and  $q$  there are identical. So in all cases,  $M_P$  accepts with probability  $1/2$ .

To prove the converse, we start by noting that each branch of  $M_P$  always accepts with probability  $1/2$  or more. So even if one of the branches accepts with probability strictly more than  $1/2$ , the acceptance probability of  $M_P$  as a whole will be strictly more than  $1/2$ .

Now suppose  $V_S$  is not a valid verifier. Then  $M_P$  would accept with probability strictly more than  $1/2$  because either  $i = 0$  or  $i = -1$  would accept with probability more than  $1/2$ .

The remaining case is where  $V_S$  is a valid verifier but  $P_S$  is not a valid prover. This means that at some point the distribution of  $P_S$ 's responses depended on the simulated verifier's randomness. Specifically, there

must exist an  $i \in [m/2]$  and  $r_V^1, r_V^2, t_1, \dots, t_{2i}$  such that  $(\{r_V^1, r_V^2\}, t_1, \dots, t_{2i-1})$  are in the support of  $S_{2i-1}(x)$  and:

$$\Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^1, t_1, \dots, t_{2i-1})] \neq \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^2, t_1, \dots, t_{2i-1})]$$

For this  $r_V^1$  and  $r_V^2$ , let  $i_0$  be the least  $i$  such that there exist  $t_1, \dots, t_{2i_0}$  where such an inequality holds.  $i_0$  being the smallest such  $i$  implies, by the same induction arguments above and the validity of  $V_S$  as a verifier, that:

$$\Pr[S_{2i_0-1}(x) = (r_V^1, t_1, \dots, t_{2i_0-1})] = \Pr[S_{2i_0-1}(x) = (r_V^2, t_1, \dots, t_{2i_0-1})]$$

Putting the above two relations together, we get:

$$\Pr[S_{2i_0}(x) = (r_V^1, t_1, \dots, t_{2i_0})] \neq \Pr[S_{2i_0}(x) = (r_V^2, t_1, \dots, t_{2i_0})]$$

So when  $M_P$  chooses  $i = i_0$  and these values of  $r_V^1, r_V^2$  and  $t_1, \dots, t_{2i_0}$ , it will accept with probability strictly greater than  $1/2$ , and so it will do so overall as well. This proves Claim 2.

Due to the fact that PP is closed under complement and Claim 2, we have now established that the conditions in Claim 1 can be checked in PP. And so by Claim 1,  $L$  can be decided in PP. It is also easy to see that this proof still works relative to any oracle, as it only makes black-box use of  $S$ .  $\square$

The following theorem follows immediately from Corollary 1.1 and Lemma 5.3.

**Theorem 5.4.** *There is an oracle  $\mathcal{O}$  such that  $\text{NISZK}^{\mathcal{O}} \not\subseteq \text{HVPZK}^{\mathcal{O}}$ . Consequently,  $\text{SZK}^{\mathcal{O}} \not\subseteq \text{PZK}^{\mathcal{O}}$  and  $\text{NISZK}^{\mathcal{O}} \not\subseteq \text{NIPZK}^{\mathcal{O}}$ .*

### 5.3 A Relativized Separation of PZK and coPZK

**Theorem 5.5.** *There is an oracle  $\mathcal{O}$  such that  $\text{PZK}^{\mathcal{O}} \neq \text{coPZK}^{\mathcal{O}}$ .*

*Proof.* In order to prove Theorem 5.5, we first show that HVPZK is closed under “composition” with GapAND.

**Lemma 5.6.** *Let  $f : D \rightarrow \{0, 1\}$  with  $D \in \{0, 1\}^M$  be a partial function and  $n$  be a positive integer,  $1/2 < \varepsilon < 1$  be a constant. If  $f$  has a  $\text{polylog}(M)$ -time HVPZK protocol, then  $\text{GapAND}_{n,\varepsilon}(f)$  has a  $\text{polylog}(nM)$ -time HVPZK protocol.*

*Proof.* For convenience, denote  $\text{GapAND}_{n,\varepsilon}(f)$  by  $g$ . Given an HVPZK protocol  $(P, V, S)$  for  $f$ , we will construct an HVPZK protocol  $(P', V', S')$  for  $g$ . Given an input  $x = (x_1, \dots, x_n)$  for  $g$ ,  $V'$  selects, say,  $\log^2(n)$  values of  $i \in [n]$ , and  $P'$  and  $V'$  run the interactive protocol  $(P, V)$  on each of the corresponding  $x_i$ 's independently.  $V'$  accepts if and only if  $(P, V)$  accepts on all these  $x_i$ 's. Completeness and soundness follows easily from standard arguments and the definition of  $g$ .

On a similar input, the simulator  $S'$  simply selects the same number of  $i$ 's, and runs  $S$  on the corresponding  $x_i$ 's. Since in a YES instance all of the  $x_i$ 's are such that  $f(x_i) = 1$ ,  $S$  simulates the transcripts of  $(P, V)$  on all of these exactly, except with a negligible probability when it fails on one or more of these. Hence  $S'$  simulates  $(P', V')$  exactly as well, again failing only with a negligible probability.  $\square$

We will need the following implication of the constructions in [DGOW95].

**Lemma 5.7** (Implied by [DGOW95]). *Any partial Boolean function that has a  $\text{polylog}(n)$ -time NIPZK protocol also has a  $\text{polylog}(n)$ -time PZK protocol.*

We will use the function  $\text{PTP}_n$  (cf. Definition 2.15) to establish our separation. The following are immediate consequences of Theorems 2.16 and 3.4 and Lemma 2.8.

**Corollary 5.8.**  $\text{PP}^{\text{dt}}(\text{GapAND}_{n,7/8}(\overline{\text{PTP}_n})) = \Omega(n^{1/3})$

**Lemma 5.9.**  $\text{PTP}_n$  has a  $\text{polylog}(n)$ -time PZK protocol.

*Proof.* We will show this by presenting a  $\text{polylog}(n)$ -time NIPZK protocol for  $\text{PTP}_n$  and invoking Lemma 5.7. The protocol is very similar to the one described in Section 2.5. Given a function  $f : [n] \rightarrow [n]$  as input, an  $r \in [n]$  is chosen at random using the common random string.  $P$  is then supposed to send an  $x$  to  $V$  such that  $f(x) = r$ .  $V$  accepts if this is true. Completeness, soundness and perfect zero-knowledge are all easily argued using the definition of  $\text{PTP}_n$ .  $\square$

Now we have everything we need to prove Theorem 5.5. Suppose  $\text{PZK}^{\mathcal{O}} = \text{coPZK}^{\mathcal{O}}$  with respect to all oracles  $\mathcal{O}$ . This implies that any language that is in  $\text{polylog}(n)$ -time PZK is also in  $\text{polylog}(n)$ -time  $\text{coPZK}$ , and vice versa – if this were not true for some language, then we would be able to use that language to construct an oracle that separates the two classes by diagonalization. In particular, this hypothesis and Lemma 5.9 imply that  $\overline{\text{PTP}_n}$  has a  $\text{polylog}(n)$ -time PZK (and hence HVPZK) protocol. Then, by Lemma 5.6,  $\text{GapAND}_{n,7/8}(\overline{\text{PTP}_n})$  has a  $\text{polylog}(n)$ -time HVPZK protocol.

This fact, along with the lower bound in Corollary 5.8, can be used to construct an oracle separating HVPZK from PP by standard diagonalization. But by Lemma 5.3, such an oracle cannot exist. So there has to be some oracle separating PZK and  $\text{coPZK}$ , completing the proof of Theorem 5.5.  $\square$

An argument identical to the proof of Theorem 5.5 (without the need to invoke Lemma 5.7) shows that the same oracle separates NIPZK and  $\text{coNIPZK}$ , as well as HVPZK and  $\text{coHVPZK}$ .

**Theorem 5.10.** *The oracle  $\mathcal{O}$  witnessing Theorem 5.5 also satisfies  $\text{NIPZK}^{\mathcal{O}} \neq \text{coNIPZK}^{\mathcal{O}}$ , as well as  $\text{HVPZK}^{\mathcal{O}} \neq \text{coHVPZK}^{\mathcal{O}}$ .*

Combining Theorems 5.3, 5.4, 5.5, and 5.10 yields Theorem 1.2 from Section 1.2.

## 6 Communication Separation Between NISZK and UPP

Based on the framework of Razborov and Sherstov [RS10], and Bun and Thaler [BT16], we are able to generalize Section 4’s separation between the query complexity classes  $\text{NISZK}^{\text{dt}}$  and  $\text{UPP}^{\text{dt}}$  to communication complexity. That is, we prove the following theorem (the communication complexity classes  $\text{NISZK}^{\text{cc}}$  and  $\text{UPP}^{\text{cc}}$  are formally defined in Appendix C).

**Theorem 6.1.**  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$ .

In light of Razborov and Sherstov’s framework, proving the above theorem boils down to identifying some  $f$  in  $\text{NISZK}^{\text{dt}}$  such that  $\text{deg}_{\pm}(\text{GapMaj}(f))$  is large, and moreover there is a dual witness to this fact that satisfies an additional *smoothness* condition. Unfortunately, the dual witness for  $\text{GapMaj}(f)$  constructed in Theorem 3.4 is not smooth.

Bun and Thaler described methods for “smoothing out” certain dual witnesses. However, their methods were specifically described in the context of functions of the form  $\text{OR}(f)$ , while we must consider functions of the form  $\text{GapMaj}(f)$ . Nonetheless, we are able to apply Bun and Thaler’s methodology to smooth out the dual witness for  $\text{GapMaj}(f)$  that we constructed in Theorem 3.4, for a non-trivial class of functions  $f$ . We thereby obtain the claimed separation of Theorem 6.1. While the proof of this result largely follows the same lines of Bun and Thaler’s, there are many details and a few subtleties to work through. We present the proof in Appendix C for completeness.

## 7 Additional Consequences

### 7.1 Consequences In Structural Complexity

Theorem 1.1’s oracle separation between NISZK (or SZK) and PP also answers a number of open questions in structural complexity; for instance it trivially implies oracle separations between SZK and  $\text{BPP}_{\text{path}}$ , as well as separations between CQP & DQP and PP. The latter classes, defined by Aaronson [Aar05], and Aaronson, Bouland, Fitzsimons and Lee [ABFL16] are complexity classes capturing the power of quantum computing with “more powerful” modified versions of quantum mechanics. The authors of [Aar05, ABFL16] showed these classes are in EXP and  $\text{BPP}^{\#P}$ , respectively, and ask if one could improve the upper bounds on these classes to e.g. PP (which is an upper bound on BQP) as an open problem. Since these classes contain SZK, our result implies that one cannot place their classes in PP using relativizing techniques. This partially explains the difficulty in substantially improving their upper bounds. This was part of our original motivation for studying this problem.

### 7.2 Consequences for Polarization

A polarization algorithm is an algorithm that is given black-box sampling access to two distributions, and outputs two new distributions that are either extremely close in total variation distance (if they were initially somewhat close) or extremely far in total variation distance (if they were originally somewhat far). In this section we describe how our oracle separation between SZK and PP implies lower bounds on polarization algorithms. In particular we show black-box polarization algorithms are limited in how close they can push the statistical difference to 0 or 1 relative to the number of bits in the output distribution.

The concept of polarization first arose in work of Sahai and Vadhan [SV03]. In their work, Sahai and Vadhan showed that the statistical difference problem is complete for the class SZK. The statistical distance problem is formulated as follows: Let  $P_b(x)$  be poly-sized classical circuits. Let  $D_b$  be the distribution on  $\{0, 1\}^n$  induced by inputting a uniformly random input  $x$  to  $P_b(x)$ . The statistical difference problem is, given circuits  $P_0$  and  $P_1$ , determine if either  $\|D_0 - D_1\| \leq 1/3$  or if  $\|D_0 - D_1\| \geq 2/3$ , promised one is the case. Here  $\|D_0 - D_1\|$  indicates the total variation distance between the distributions  $D_0$  and  $D_1$ .

In their paper, Sahai and Vadhan also showed a remarkable property of the statistical difference problem – namely that the constants  $1/3$  and  $2/3$  in the Statistical Difference problem can be amplified to be exponentially close to 0 and 1 [SV03]. This property is not immediately obvious, because it cannot be obtained by simply repeatedly sampling from  $D_0$  and  $D_1$ . Nevertheless, they showed the following: given black-box distributions  $D_0$  and  $D_1$ , and a number  $k$  expressed in unary, then in polynomial time one can sample from distributions  $D'_0$  and  $D'_1$  (using polynomially many samples from  $D_0$  and  $D_1$ ) such that, if  $\|D_0 - D_1\| \leq 1/3$ , then  $\|D'_0 - D'_1\| \leq \varepsilon$  and if  $\|D_0 - D_1\| \geq 2/3$ , then  $\|D'_0 - D'_1\| \geq 1 - \varepsilon$ , where  $\varepsilon = 2^{-k}$ . Hence without loss of generality, one can assume that the distributions in the statistical difference problem are exponentially close to 0 or 1; their transformation “polarizes” the distributions to be either very close or very far from one another. This is known as the Polarization Lemma, and is a key part of the proof that Statistical Difference is SZK-complete<sup>10</sup>.

Given this fundamental result, it is natural to ask whether or not one can improve the parameters of the Polarization Lemma. For instance, Sahai and Vadhan noted in their paper that their algorithm could only polarize distributions under the promise  $\|D_0 - D_1\| > \alpha$  or  $\|D_0 - D_1\| < \beta$  in the case that  $\alpha^2 > \beta$ . So their algorithm can polarize  $\alpha = 2/3$  and  $\beta = 1/3$ , but not  $\alpha = 5/9$  and  $\beta = 4/9$ . A natural question is whether or not this limitation could be removed. Holenstein and Renner answered this question in the negative for certain types of black-box polarization [HR05]. In particular, they showed that any form of

---

<sup>10</sup>In statistical zero-knowledge proof systems, the verifier must be able to simulate the honest prover to negligibly small ( $1/\text{superpoly}$ ) total variation distance. The ability to polarize distributions allows the statistical difference problem to have this property.

black-box polarization which works by drawing strings  $b, c \in \{0, 1\}^\ell$ , and then sets  $D'_0 = D_{b_1} \otimes \dots \otimes D_{b_\ell}$  and  $D'_1 = D_{c_1} \otimes \dots \otimes D_{c_\ell}$  cannot polarize in the case where  $\alpha^2 < \beta$ . As Sahai and Vadhan's polarization algorithm took this form, this was strong evidence that this limitation was fundamental. Note, however, that it remains open to show that polarization cannot occur when  $\alpha^2 < \beta$  using *arbitrary* black-box algorithms. For instance, one could feed the random outputs of  $D_0$  back into the circuit for  $D_1$  in order to help polarize the distributions. While it is not clear how these sorts of operations could help one polarize, it is difficult to rule out the possibility that such operations might lead to a stronger polarization algorithm.

In this paper we consider different parameters of the Polarization Lemma - namely how small can the security parameter  $\varepsilon$  be relative to the size of the range of the output distributions. For example, if one is given distributions  $D_0$  and  $D_1$  over  $n$ -bit strings with total variation distance  $> 2/3$  or  $< 1/3$ , then can one create distributions  $D'_0$  and  $D'_1$  over  $n'$ -bit string such that the total variation distance is  $\leq \varepsilon$  or  $\geq 1 - \varepsilon$  where  $\varepsilon = 2^{-n'}$ , or  $2^{-n'^2}$ ? At first it might appear the answer to the above question is trivially yes - because one can simply set  $k = -n^2$  (or  $k = n^c$  for any constant  $c$ ) and run the Polarization Lemma. However this does not work because the Polarization Lemma increases the size of the domains of the distributions as it polarizes; in other words  $n'$  is some polynomial function of  $n$  and  $k$ . By tweaking the parameters of the Polarization Lemma slightly [SV03], one can polarize distributions on  $n$  bits to distributions on  $n' = \text{poly}(n)$  bits which are polarized to roughly  $\varepsilon \approx 2^{-\sqrt{n'}}$ . However, it seems difficult to do better than  $\varepsilon = 2^{-\sqrt{n'}}$  using the proof techniques of Sahai and Vadhan [SV03]. This is because their proof alternates between two lemmas, one which total variation distance towards 1 in the case the distributions are far apart, and another which pushes the total variation distance towards zero in the case the distributions are close. In order to make the distributions  $2^{-k}$ -close or  $1 - 2^{-k}$ -far, one must apply both lemmas, each of which increases the number of bits output by the distribution by a factor of  $k$ . Hence using Sahai and Vadhan's Lemma with  $k = n^c$ , the best one can achieve are distributions on  $n' = n^{2c+1}$  bits which are either  $2^{-n^c}$ -close or  $(1 - 2^{-n^c})$ -far. For large constant  $c$  this gives  $\varepsilon \approx 2^{-\sqrt{n'}}$ . It seems difficult to improve their lemma further using the techniques of their paper.

A natural question is therefore: what is the smallest value of  $\varepsilon$  that one can achieve relative to the size of the output distributions  $n'$ ? In this work, we show that if  $\varepsilon$  can be made very small relative to  $n'$ , then that would place  $\text{SZK}^\mathcal{O} \subseteq \text{PP}^\mathcal{O}$  (and even  $\text{SZK}^\mathcal{O} \subseteq \text{BPP}_{\text{path}}^\mathcal{O}$ ) for all oracles  $\mathcal{O}$ . Therefore, as a corollary of our main result,  $\varepsilon$  cannot be made very small by any poly-time black-box polarization algorithm. More specifically, we achieve a lower bound of  $\varepsilon > 2^{-n'/2-1}$  for any poly-time polarization algorithm.

More specifically, we prove two theorems showing that a stronger version of polarization places SZK in PP relative to all oracles. Therefore, a stronger polarization algorithm cannot exist as a corollary of Theorem 1.1.

**Theorem 7.1.** *Suppose that there is an algorithm running in  $\text{poly}(n)$  time, which given black box distributions  $D_0, D_1$  on strings of length  $n$  which obey either  $|D_0 - D_1| < 1/3$  or  $|D_0 - D_1| > 2/3$ , produces two output distributions  $D'_0$  and  $D'_1$  on strings of length  $n' = \text{poly}(n)$  such that either  $|D'_0 - D'_1| < \varepsilon$  (in the first case) or  $|D'_0 - D'_1| > 1 - \varepsilon$  (in the second case) where  $\varepsilon \leq 2^{-n'/2-1}$ . Then  $\text{SZK}^\mathcal{O} \subseteq \text{PP}^\mathcal{O}$  for all oracles  $\mathcal{O}$ .*

**Theorem 7.2.** *Suppose that there is an algorithm running in  $\text{poly}(n)$  time, which given black box distributions  $D_0, D_1$  on strings of length  $n$  which obey either  $|D_0 - D_1| < 1/3$  or  $|D_0 - D_1| > 2/3$ , produces two output distributions  $D'_0$  and  $D'_1$  on strings of length  $n' = \text{poly}(n)$  such that either  $|D'_0 - D'_1| < \varepsilon$  (in the first case) or  $|D'_0 - D'_1| > 1 - \varepsilon$  (in the second case) where  $\varepsilon \leq 2^{-2n'/3-1}$ . Then  $\text{SZK}^\mathcal{O} \subseteq (\text{BPP}_{\text{path}})^\mathcal{O}$  for all oracles  $\mathcal{O}$ .*

Therefore as a corollary of Theorem 1.1, there do not exist poly-time polarization algorithms achieving  $\varepsilon = 2^{-n'/2-1}$ . In fact one could have achieved such a lower bound even if one had merely given an

oracle separation between SZK and  $\text{BPP}_{\text{path}}$ . It remains open to close the gap between our lower bound of  $\varepsilon = 2^{-n'/2-1}$  and the upper bound of  $\varepsilon = 2^{-n'/2+\delta}$  for any  $\delta > 0$  given by Sahai and Vadhan [SV03].

The proof of Theorem 7.1 is relatively straightforward. Suppose one can polarize to  $\varepsilon' \ll 2^{-n'/2}$ . Then the output distributions now have a promise on the  $\ell_2$  distance between the output distributions - in particular the  $\ell_2$  distance between them is more or less than some (exponentially small) threshold. It is easy to decide this problem PP - this is because the  $\ell_2$  distance square is a degree-two polynomial in the output probabilities. To see this, say you're trying to determine if  $S = \sum_{x \in \{0,1\}^n} (D'_0(x) - D'_1(x))^2$  is more or less

than some threshold  $t$ , consider the following algorithm: pick at random  $x$ , pick a random number 1,2,3 or 4. If the number is 1 (respectively 4) sample two samples from  $D'_0$  (respectively  $D'_1$ ) and accept if they both give output  $x$ , otherwise output accept/reject using a 50-50 coin flip. If the number is 2 or 3 sample one sample from  $D'_0$  and  $D'_1$  and reject iff they collide, otherwise output a 50-50 coin flip. The probability this machine accepts is  $1/2 + S/2$  - which is more or less than a known threshold  $(1 + t)/2$ . Therefore by correcting the bias of the machine with an initial coin flip, this is a PP algorithm to decide the problem. In short, deciding thresholds for the  $\ell_2$  norm is easy for PP because it is a low-degree polynomial, while deciding thresholds for the  $\ell_1$  norm is hard for PP because the  $\ell_1$  norm is not a low degree polynomial.

On the other hand, the proof of Theorem 7.2 is involved - it works by examining the algorithms of Aaronson [Aar05] and Aaronson, Bouland, Fitzsimons and Lee [ABFL16] showing that certain modified versions of quantum mechanics can be used to solve SZK-hard problems in polynomial time. These algorithms are not based on postselection (otherwise they would place  $\text{SZK} \subseteq \text{PostBQP} = \text{PP}$  for all oracles, a contradiction with our main result). However, it turns out that if one has a very strong polarization lemma, then one can turn them into postselected quantum algorithms (and even postselected classical algorithms) for statistical difference. Interestingly, this was part of our original motivation for this work. We include this proof in Appendix F for the interested reader.

### 7.3 Consequences for Property Testing

**Lower Bounds for Property Testers That Barely Do Better Than Random Guessing.** For any NISZK-hard property testing problem  $P$ , our query complexity lower bounds immediately imply that any property testing algorithm for  $P$  that outputs the correct answer with probability strictly greater than  $1/2$  requires  $n^{\Omega(1)}$  queries. For concreteness, we highlight the result we obtain for the NISZK-complete problem of entropy approximation. Specifically, given a distribution  $D$  over  $n$  elements, a natural problem is to ask how many samples from  $D$  are required to estimate the entropy of  $D$  to additive error. In 2011, Valiant and Valiant [VV11] showed that to achieve any constant additive error less than  $\log 2/2$ , it is both necessary and sufficient to take  $\Theta(n/\log n)$  samples from  $D$ . However, their bounds assume that one wishes to estimate the entropy with high probability, say with probability  $1 - o(1/\text{poly}(n))$ . Quantitatively, our  $\text{UPP}^{\text{dt}}$  query lower bounds imply the following.

**Corollary 7.3.** *Any algorithm which decides if the entropy of  $D$  (over domain size  $n$ ) is  $\leq k - 1$  or  $\geq k + 1$  and succeeds with probability  $> \frac{1}{2}$  requires  $\Omega(n^{1/4}/\log n)$  samples from  $D$ .*

In other words, estimating the entropy of a distribution to additive error 2 requires  $\tilde{\Omega}(n^{1/4})$  samples, even if the algorithm is only required to have an arbitrarily small bias in deciding the answer correctly.

### 7.4 Consequences for Delegating Computation

In this section, we point out an easy implication of our results: two-message streaming interactive proofs (SIPs) [CTY11] of logarithmic cost can compute functions outside of  $\text{UPP}^{\text{cc}}$ .

In a SIP, a verifier with limited working memory makes a single streaming pass over an input  $x$ , and then interacts with an untrusted prover, who evaluates a function  $f$  of the input, and attempts to convince the verifier of the value of  $f(x)$ . The protocol must satisfy standard notions of completeness and soundness. The cost of the protocol is the size of the verifier’s working memory and the total length of the messages exchanged between the prover and verifier. We direct the interested reader to [CTY11] for the formal definition.

It follows from our analysis in Appendix C that the  $(O(n), n, F)$ -pattern matrix of the function  $F := \text{GapMaj}_{n^{1/4}, 499}(\text{PTP}_{n^{3/4}})$  specifies a communication problem that is outside of  $\text{UPP}^{\text{cc}}$  (see Appendix C.1.5 for a definition of pattern matrices). For our purposes, the relevant properties of such pattern matrices are as follows. In the communication problem  $F^{\text{cc}}(x, y)$  corresponding to the pattern matrix of a function  $F: \{0, 1\}^n \rightarrow \{0, 1\}$ , Alice’s input  $x$  and Bob’s input  $y$  together specify a vector  $u(x, y) \in \{0, 1\}^n$ , and  $F^{\text{cc}}(x, y)$  is defined to equal  $F(u(x, y))$ . Moreover, each coordinate of  $u(x, y)$  depends on  $O(1)$  entries of  $x$  and  $y$  respectively.

Observe that  $F$  is computed by a simple two-message interactive proof in which the verifier makes  $O(\log n)$  *non-adaptive* queries to bits of the input  $x$ , uses  $O(\log n)$  bits of working memory, and the total communication cost is also  $O(\log n)$ : the verifier picks an instance of  $\text{PTP}_{n^{3/4}}$  at random, and runs the SZK protocol for PTP described in Section 2.5 on that instance (we do not need the zero-knowledge property of the SZK protocol here). Completeness and soundness follow from the definition of  $\text{GapMaj}$  and completeness and soundness of the SZK protocol for PTP.

Consider a data stream consisting of Alice and Bob’s inputs to  $F^{\text{cc}}$ , and a SIP verifier who wishes to compute  $F^{\text{cc}}(x, y)$ . That is, the first part of the stream specifies  $x$  and the second part specifies  $y$ . There is a simple two-message SIP for evaluating  $F^{\text{cc}}(x, y)$ : the SIP verifier simulates the verifier in the above interactive proof for  $F(u(x, y))$ . Since the latter verifier only needs to know  $O(\log n)$  bits of  $u(x, y)$ , and each bit of  $u(x, y)$  depends on  $O(1)$  bits of  $x$  and  $y$ , the SIP verifier can compute the bits of  $u(x, y)$  that are necessary to run the simulation, using just  $O(\log n)$  bits of memory and a single streaming pass over the input. We obtain the following theorem.

**Theorem 7.4.** *There is a communication problem  $F^{\text{cc}}(x, y): \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  such that  $F^{\text{cc}}$  is not in  $\text{UPP}^{\text{cc}}$ , yet given a data stream specifying  $x$  followed by  $y$ , there is a 2-message SIP of cost  $O(\log n)$  computing  $F^{\text{cc}}(x, y)$ .*

Theorem 7.4 provides an explanation for why prior work attempting to understand the power of 2-message SIPs has succeeded only in proving lower bounds on special classes of such protocols [CCM<sup>+</sup>15]. Indeed, taking  $\text{UPP}^{\text{cc}}$  to represent the limit of our methods for proving lower bounds in communication complexity, the fact that 2-message SIPs and their analogous two-party communication model (called  $\mathbf{OIP}_+^{[2]}$  in [CCM<sup>+</sup>15]) can compute functions outside of  $\text{UPP}^{\text{cc}}$  means that proving superlogarithmic lower bounds for 2-message SIPs will require new lower bound techniques.

### 7.4.1 Consequences for the Algebrization Barrier

Arithmetization is a powerful technique in complexity theory that is used, for example, to prove  $\text{IP} = \text{PSPACE}$  [Sha92, LFKN92] and many other celebrated theorems. While arithmetization circumvents the *relativization* barrier, Aaronson and Wigderson [AW09] proposed a new *algebrization* barrier. Roughly speaking, their results show that arithmetization alone will not suffice to resolve many open questions in complexity theory. Informally, one of their key results was the following.

**Theorem 7.5** (Informal, implicit in Theorem 5.11 of [AW09]). *For two complexity classes  $\mathcal{C}$  and  $\mathcal{D}$ , if  $\mathcal{C}^{\text{cc}} \not\subseteq \mathcal{D}^{\text{cc}}$ , then arithmetization techniques alone cannot prove  $\mathcal{C} \subseteq \mathcal{D}$ .*



From the above theorem, our communication class separation  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$  immediately implies the following informal corollary.

**Corollary 7.6** (Informal). *Arithmetization techniques alone cannot prove that  $\text{NISZK} \subseteq \text{PP}$ .*

## 8 Open Problems

Our work leaves a number of open related problems. First, we have shown that the function  $\text{GapMaj}(f)$  is hard for  $\text{UPP}^{\text{dt}}$ , for any function  $f$  of high approximate degree, and that  $\text{GapAND}(f)$  is hard for  $\text{UPP}^{\text{dt}}$ , for any function of high positive one-sided approximate degree. Can one extend this work to characterize when  $f \circ g$  is hard for  $\text{UPP}^{\text{dt}}$ , based on some properties of  $f$  and  $g$ ? We conjecture that the  $\text{UPP}^{\text{dt}}$  complexity of  $\text{GapMaj}(f)$  (respectively,  $\text{GapAND}(f)$ ) is characterized by the *rational approximate degree* of  $f$  (respectively, positive one-sided approximate degree of  $f$ ). Such a result would complement the characterization of the threshold degree of  $\text{AND}(f)$  in terms of positive one-sided rational approximate degree given in [She14].

Additionally, we have shown a lower bound on certain parameters of the polarization lemma. Is there a polarization algorithm which matches our lower bound?

It would also be interesting to determine whether our lower bounds on property testing algorithms that output the correct answer with probability strictly greater than  $1/2$  are quantitatively tight. For example, is there an algorithm that, given query access to a distribution  $D$  (over domain size  $n$ ) that is promised to have entropy  $\leq k - 1$  or  $\geq k + 1$ , decides which is the case with probability greater than  $1/2$ , using  $\tilde{O}(n^{1/4})$  samples from  $D$ ?

Finally, the main open question highlighted by our work is to break through the UPP frontier in communication complexity. We formalize this question via the following challenge: prove any superlogarithmic lower bound for an explicit problem in a natural communication model that cannot be efficiently simulated by  $\text{UPP}^{\text{cc}}$ . Our work shows that any communication model capable of efficiently computing the pattern matrix of  $\text{GapMaj}(\text{PTP})$  is a candidate for achieving this goal. Thomas Watson has suggested the following as perhaps the simplest such candidate: consider the  $\text{NISZK}^{\text{cc}}$  model, but restricted to be one-way, in the sense that neither Merlin nor Bob can talk to Alice. This model effectively combines the key features of the  $\text{NISZK}^{\text{cc}}$  and  $\mathbf{OIP}_+^{[2]}$  (cf. [CCM<sup>+</sup>15]) communication models. There is a logarithmic cost “one-way  $\text{NISZK}$ ” protocol for the pattern matrix of  $\text{GapMaj}(\text{PTP})$ , so this model cannot be efficiently simulated by  $\text{UPP}^{\text{cc}}$ . Curiously, despite the ability of this model to compute functions outside of  $\text{UPP}^{\text{cc}}$ , to the best of our knowledge it is possible that even the  $\text{INDEX}$  function requires polynomial cost in this model. Note that while Chakrabarti et al. [CCM<sup>+</sup>15] gave an efficient  $\mathbf{OIP}_+^{[2]}$  communication protocol for  $\text{INDEX}$ , their protocol is not zero-knowledge.

## Acknowledgments

We thank Scott Aaronson, Jayadev Acharya, Shalev Ben-David, Clément Canonne, Oded Goldreich, Mika Göös, Gautam Kamath, Robin Kothari, Tomoyuki Morimae, Harumichi Nishimura, Ron Rothblum, Mike Saks, Salil Vadhan and Thomas Watson for helpful discussions. Adam Bouland was supported in part by the NSF Graduate Research Fellowship under grant no. 1122374 and by the NSF Alan T. Waterman award under grant no. 1249349. Lijie Chen was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61361136003. Dhiraj Holden was supported in part by an Akamai Presidential fellowship, by NSF MACS - CNS-1413920, and by a SIMONS Investigator award Agreement Dated 6-5-12. Prashant Vasudevan was

supported in part by the Qatar Computing Research Institute under the QCRI-CSAIL partnership, and by the National Science Foundation Frontier grant CNS 1413920.

## References

- [Aar] Scott Aaronson. Personal communication.
- [Aar02] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 635–642. ACM, 2002.
- [Aar05] Scott Aaronson. Quantum computing and hidden variables. *Physical Review A*, 71(3):032325, 2005.
- [Aar12] Scott Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Information & Computation*, 12(1-2):21–28, 2012.
- [ABFL16] Scott Aaronson, Adam Bouland, Joseph Fitzsimons, and Mitchell Lee. The space “just above” BQP. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, pages 271–280, 2016.
- [AH91a] William Aiello and Johan Håstad. Relativized perfect zero knowledge is not BPP. *Information and Computation*, 93:223–240, 1991.
- [AH91b] William Aiello and Johan Håstad. Statistical zero-knowledge languages can be recognized in two rounds. *J. Comput. Syst. Sci.*, 42(3):327–345, 1991.
- [AIKP15] Shweta Agrawal, Yuval Ishai, Dakshita Khurana, and Anat Paskin-Cherniavsky. Statistical randomized encodings: A complexity theoretic view. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 1–13, 2015.
- [Amb05] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.
- [AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 51(4):595–605, 2004.
- [AW09] Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *ACM Transactions on Computation Theory (TOCT)*, 1(1):2, 2009.
- [Bar01] Boaz Barak. How to go beyond the black-box simulation barrier. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 106–115, 2001.
- [BFS86] Laszlo Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Foundations of Computer Science, 1986., 27th Annual Symposium on*, pages 337–347. IEEE, 1986.
- [BHZ87] Ravi B. Boppana, Johan Håstad, and Stathis Zachos. Does co-np have short interactive proofs? *Inf. Process. Lett.*, 25(2):127–132, 1987.

- [BRS91] Richard Beigel, Nick Reingold, and Daniel A. Spielman. PP is closed under intersection (extended abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 1–9, 1991.
- [BT15a] Mark Bun and Justin Thaler. Dual polynomials for collision and element distinctness. *arXiv preprint arXiv:1503.07261*, 2015.
- [BT15b] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *International Colloquium on Automata, Languages, and Programming*, pages 268–280. Springer, 2015.
- [BT16] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of  $AC^0$ . *Electronic Colloquium on Computational Complexity (ECCC)*, 23:75, 2016.
- [BT17] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of  $ac^0$ . *Electronic Colloquium on Computational Complexity (ECCC)*, 24:51, 2017.
- [CCG<sup>+</sup>94] Richard Chang, Benny Chor, Oded Goldreich, Juris Hartmanis, Johan Håstad, Desh Ranjan, and Pankaj Rohatgi. The random oracle hypothesis is false. *J. Comput. Syst. Sci.*, 49(1):24–39, 1994.
- [CCM<sup>+</sup>15] Amit Chakrabarti, Graham Cormode, Andrew McGregor, Justin Thaler, and Suresh Venkatasubramanian. Verifiable stream computation and arthur-merlin communication. In David Zuckerman, editor, *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, volume 33 of *LIPICs*, pages 217–243. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.
- [Che16a] Lijie Chen. Adaptivity vs postselection. *arXiv:1606.04016*, 2016.
- [Che16b] Lijie Chen. A note on oracle separations for BQP. *arXiv:1605.00619*, 2016.
- [CTY11] Graham Cormode, Justin Thaler, and Ke Yi. Verifying computations with streaming interactive proofs. *PVLDB*, 5(1):25–36, 2011.
- [DGJ<sup>+</sup>10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010.
- [DGOW95] Ivan Damgård, Oded Goldreich, Tatsuaki Okamoto, and Avi Wigderson. Honest verifier vs dishonest verifier in public coin zero-knowledge proofs. In *Advances in Cryptology - CRYPTO '95, 15th Annual International Cryptology Conference, Santa Barbara, California, USA, August 27-31, 1995, Proceedings*, pages 325–338, 1995.
- [DR96] Devdatt P Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *BRICS Report Series*, 3(25), 1996.
- [EFHK14] Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors. *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, volume 8572 of *Lecture Notes in Computer Science*. Springer, 2014.

- [Fis02] Marc Fischlin. On the impossibility of constructing non-interactive statistically-secret protocols from any trapdoor one-way function. In *Proceedings of the The Cryptographer's Track at the RSA Conference on Topics in Cryptology, CT-RSA '02*, pages 79–95, London, UK, UK, 2002. Springer-Verlag.
- [For87] Lance Fortnow. The complexity of perfect zero-knowledge (extended abstract). In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, pages 204–209, 1987.
- [GG98] Oded Goldreich and Shafi Goldwasser. On the limits of non-approximability of lattice problems. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 1–9. ACM, 1998.
- [GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on computing*, 18(1):186–208, 1989.
- [GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM (JACM)*, 38(3):690–728, 1991.
- [Gol15] Shafi Goldwasser. Zero knowledge probabilistic proof systems. <https://www.youtube.com/watch?v=J4TkHuTmHsg#t=1h15m20s>, 2015.
- [GPW15a] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:49, 2015.
- [GPW15b] Mika Göös, Toniann Pitassi, and Thomas Watson. Zero-information protocols and unambiguity in arthur-merlin communication. In Tim Roughgarden, editor, *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 113–122. ACM, 2015.
- [GSV98] Oded Goldreich, Amit Sahai, and Salil P. Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 399–408, 1998.
- [GSV99] Oded Goldreich, Amit Sahai, and Salil Vadhan. Can statistical zero knowledge be made non-interactive? or on the relationship of SZK and NISZK. In *Advances in Cryptology CRYPTO99*, pages 467–484. Springer, 1999.
- [GT14] Oded Goldreich and Liav Teichner. Super-perfect zero-knowledge proofs. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:97, 2014.
- [HR05] Thomas Holenstein and Renato Renner. One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, pages 478–493, 2005.
- [Kla11] Hartmut Klauck. On arthur merlin games in communication complexity. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, June 8-10, 2011*, pages 189–199. IEEE Computer Society, 2011.

- [KP14] Hartmut Klauck and Supartha Podder. Two results about quantum messages. In Erzsébet Csuhaaj-Varjú, Martin Dietzfelbinger, and Zoltán Ésik, editors, *Mathematical Foundations of Computer Science 2014 - 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part II*, volume 8635 of *Lecture Notes in Computer Science*, pages 445–456. Springer, 2014.
- [KT14] Varun Kanade and Justin Thaler. Distribution-independent reliable learning. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 3–24. JMLR.org, 2014.
- [Kut05] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.
- [LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM (JACM)*, 39(4):859–868, 1992.
- [Lok01] Satyanarayana V. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *J. Comput. Syst. Sci.*, 63(3):449–473, 2001.
- [LS09] Nathan Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability & Computing*, 18(1-2):227–245, 2009.
- [LZ16] Shachar Lovett and Jiapeng Zhang. On the impossibility of entropy reversal, and its application to zero-knowledge proofs. *ECCC TR16-118*, July 31 2016.
- [Mal15] Lior Malka. How to achieve perfect simulation and a complete problem for non-interactive perfect zero-knowledge. *Journal of Cryptology*, 28(3):533–550, 2015.
- [Oka96] Tatsuaki Okamoto. On relationships between statistical zero-knowledge proofs. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 649–658. ACM, 1996.
- [Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 468–474. ACM, 1992.
- [PS86] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.
- [PSS14] Periklis A. Papakonstantinou, Dominik Scheder, and Hao Song. Overlays and limited memory communication. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 298–308. IEEE Computer Society, 2014.
- [PV08] Chris Peikert and Vinod Vaikuntanathan. Noninteractive statistical zero-knowledge proofs for lattice problems. In *Annual International Cryptology Conference*, pages 536–553. Springer, 2008.
- [RS10] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of  $AC^0$ . *SIAM J. Comput.*, 39(5):1833–1855, 2010.
- [Sha92] Adi Shamir.  $IP=PSPACE$ . *Journal of the ACM (JACM)*, 39(4):869–877, 1992.

- [She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.
- [She14] Alexander A Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 223–232. ACM, 2014.
- [She15] Alexander A Sherstov. The power of asymmetry in constant-depth circuits. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 431–450. IEEE, 2015.
- [SV03] Amit Sahai and Salil Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM (JACM)*, 50(2):196–249, 2003.
- [Tha14] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:150, 2014.
- [Tod91] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.
- [Ver95] Nikolai K. Vereshchagin. Lower bounds for perceptrons solving some separation problems and oracle separation of AM from PP. In *Third Israel Symposium on Theory of Computing and Systems, ISTCS 1995, Tel Aviv, Israel, January 4-6, 1995, Proceedings*, pages 46–51, 1995.
- [VV11] Gregory Valiant and Paul Valiant. Estimating the unseen: an  $n/\log(n)$ -sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 685–694, 2011.

## A Upper Bounds on $\text{UPP}^{\text{dt}}(\text{GapMaj}(f))$ and $\text{UPP}^{\text{dt}}(\text{GapAND}(f))$

In Section 3, we proved that if  $f$  has a high *approximate degree* (*positive one-sided approximate degree*), then  $\text{GapMaj}(f)$  ( $\text{GapAND}(f)$ ) is hard for UPP algorithms. In this section we show that condition is also necessary: when  $\text{deg}^+(f)$  is small,  $\text{GapAND}(f)$  has a lower UPP query complexity; and when  $\widetilde{\text{deg}}(f)$  is small,  $\text{GapMaj}(f)$  has a low UPP query complexity. Formally, we have:

**Theorem A.1.** *For a partial function  $f$  with input length  $n$ , and a positive integer  $m$ .*

$$\begin{aligned} \text{UPP}^{\text{dt}}(\text{GapMaj}_{m,2/3}(f)) &= O(\widetilde{\text{deg}}(f)). \\ \text{UPP}^{\text{dt}}(\text{GapAND}_{m,2/3}(f)) &= O(\text{deg}^+(f)). \end{aligned}$$

*Recall that  $\text{deg}^+(f) := \text{deg}_{1/3}^+(f)$ , and  $\widetilde{\text{deg}}(f) := \widetilde{\text{deg}}_{1/3}(f)$ .*

The choice of the constants  $2/3$  and  $1/3$  is only for convenience. We need the following standard fact for approximate degree and positive one-sided approximate degree, (for example, see [DGJ<sup>+</sup>10, Claim 4.3] and [She15, Fact 2.4]).

**Claim 3.** *For any constant  $0 < \varepsilon < 0.5$ , and any partial function  $f$ ,  $\widetilde{\text{deg}}_\varepsilon(f) = \Theta(\widetilde{\text{deg}}(f))$  and  $\text{deg}_\varepsilon^+(f) = \Theta(\text{deg}^+(f))$ .*

Now we prove Theorem A.1.

*Proof of Theorem A.1.* We first upper bound  $\text{UPP}^{\text{dt}}(\text{GapMaj}_{m,2/3}(f))$ . By Lemma 2.9, it suffices to upper bound  $\text{deg}_{\pm}(\text{GapMaj}_{m,2/3}(f))$ . By Claim 3, we have  $\widetilde{\text{deg}}_{1/20}(f) = O(\widetilde{\text{deg}}(f))$ ; let  $p$  be the corresponding approximating polynomial for  $f$ . Now, let the input for  $\text{GapMaj}_{m,2/3}(f)$  be  $x = (x_1, x_2, \dots, x_m)$ , where each  $x_i$  is the input to the  $i$ th copy of  $f$ . Let  $q$  be a polynomial on  $\{0, 1\}^{nm}$  defined as  $q(x) := \frac{1}{m} \cdot \sum_{i=1}^m p(x_i)^2 - 0.5$ .

Now, evidently  $\text{deg}(q) = 2 \text{deg}(p)$ . From the definition of  $\widetilde{\text{deg}}_{1/20}(f)$ , we can see when  $\text{GapMaj}_{m,2/3}(f)(x) = 1$ ,

$$q(x) \geq \frac{2}{3} \cdot 0.95^2 - 0.5 > 0;$$

and when  $\text{GapMaj}(f)(x) = 0$ ,

$$q(x) \leq \frac{2}{3} \cdot 0.05^2 + \frac{1}{3} \cdot 1.05^2 - 0.5 < 0.$$

Hence, by the definition of  $\text{deg}_{\pm}$  (cf. Definition 2.3), we conclude that  $\text{deg}_{\pm}(\text{GapMaj}_{m,2/3}(f)) = O(\widetilde{\text{deg}}(f))$ , and this completes the proof for the first claim.

Similarly, in order to upper bound  $\text{UPP}^{\text{dt}}(\text{GapAND}_{m,2/3}(f))$ , it suffices to consider  $\text{deg}_{\pm}(\text{GapAND}_{m,2/3}(f))$ . By Claim 3, we have  $\text{deg}_{1/20}^+(f) = O(\text{deg}^+(f))$ ; let  $p$  be a corresponding positive one-sided approximating polynomial for  $f$ . Let  $q$  be a polynomial on  $\{0, 1\}^{nm}$  defined as  $q(x) := \frac{1}{m} \cdot \sum_{i=1}^m p(x_i) - 0.5$ .

Clearly  $\text{deg}(q) = \text{deg}(p)$ . From the definition of  $\text{deg}_{1/20}^+(f)$ , we can see when  $\text{GapAND}_{m,2/3}(f)(x) = 1$ ,

$$q(x) \geq 0.95 - 0.5 > 0;$$

and when  $\text{GapAND}_{m,2/3}(f)(x) = 0$ ,

$$q(x) \leq \frac{2}{3} \cdot 0.05 + \frac{1}{3} \cdot 1.05 - 0.5 < 0.$$

Therefore,  $\text{deg}_{\pm}(\text{GapAND}_{m,2/3}(f)) = O(\text{deg}^+(f))$ , and this completes the whole proof.  $\square$

## B Missing Proofs From Section 3

In this section we provide the missing proofs from Section 3. We begin with Lemma 3.5, restating the lemma here for convenience.

**Lemma 3.5** (restated) *Let  $a \geq 40$ ,  $n$  be a sufficiently large integer and  $\varepsilon$  be a real such that  $0.5 < \varepsilon < 1$ . Then there exists an (explicitly given) univariate polynomial  $P: \mathbb{R} \rightarrow \mathbb{R}$  such that:*

- $P(x) = (-a)^x$  for  $x \in \{0, \dots, \varepsilon \cdot n\}$ .
- $|P(x)| \leq a^x/2$  for  $x \in \{\varepsilon \cdot n + 1, \dots, n\}$ .
- $P$  has degree of at most  $\left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3$ .

*Proof of Lemma 3.5.* We begin by constructing the polynomial  $P: \mathbb{R} \rightarrow \mathbb{R}$  whose existence is claimed by the lemma.

**Construction of  $P$ .** Let  $N = \left\lceil \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 2 \right\rceil$ . We define  $P$  through interpolation to be the unique polynomial of degree at most  $N$  satisfying the following properties.

- $P(x) = (-a)^x$  for  $x \in \{0, \dots, \varepsilon \cdot n\}$ .
- $P(x) = 0$  for  $x \in \{\varepsilon \cdot n + 1, \dots, N\}$ .

**Analysis of  $P$ .** Under the above definition, it is obvious that the first and the last conditions in Lemma 3.5 are satisfied by  $P$ . In the rest of the proof, we establish that  $P$  also satisfies the second condition claimed by Lemma 3.5, i.e.,

$$|P(x)| \leq a^x / 2 \text{ for } x \in \{\varepsilon \cdot n + 1, \dots, n\}. \quad (27)$$

When  $\varepsilon$  is a constant strictly between  $1/2$  and  $1$  and  $a$  is a sufficiently large constant, Equation (27) is an easy consequence of standard bounds on the growth rate of low-degree polynomials defined through interpolation (cf. [RS10, Lemma 3.1]). However, our applications require us to consider  $\varepsilon \approx 1 - 1/3 \log n = 1 - o(1)$  and  $a = \Theta(\log n)$ . To handle this parameter regime, a more delicate analysis seems to be required.

For each  $i \in \{0, \dots, \varepsilon \cdot n\}$ , define the polynomial  $e_i$  as

$$e_i(x) := \prod_{j \in \{0, \dots, N\} \setminus \{i\}} \frac{x - j}{i - j}. \quad (28)$$

Observe that

$$\text{when } x \in \{0, \dots, N\}, e_i(x) \text{ is equivalent to } \mathbb{1}_{x=i}. \quad (29)$$

Moreover, each  $e_i(x)$  has degree at most  $N$ . Hence, we may write

$$P = \sum_{i \in \{0, \dots, \varepsilon \cdot n\}} e_i \cdot (-a)^i. \quad (30)$$

Indeed, the right hand side of Equation (30) is a polynomial of degree at most  $N$ , and by Observation (29), the right hand side agrees with  $P$  at all  $N$  inputs in  $\{0, 1, \dots, N\}$ . It follows that the right hand side of Equation (30) and  $P$  are equal as formal polynomials.

Thus, for any  $x$ ,  $P(x)$  can be expressed as follows.

$$P(x) := \sum_{i \in \{0, \dots, \varepsilon \cdot n\}} e_i(x) \cdot (-a)^i. \quad (31)$$

For  $x \in \{\varepsilon \cdot n + 1, \dots, N\}$ , as  $P(x) = 0$ , Equation (27) trivially satisfied. So we assume  $x \in \{N + 1, \dots, n\}$  from now on. Observe that for each  $i \in \{0, \dots, \varepsilon \cdot n\}$ ,

$$\prod_{j \in \{0, \dots, N\} \setminus \{i\}} (x - j) = \prod_{j \in \{x-N, \dots, x\} \setminus \{x-i\}} j = \frac{x - N}{x - i} \prod_{j=x-N+1}^x j \leq \prod_{j=x-N+1}^x j = x! / (x - N)! \quad (32)$$

and

$$\prod_{j \in \{0, \dots, N\} \setminus \{i\}} |i - j| = \prod_{j=0}^{i-1} (i - j) \cdot \prod_{j=i+1}^N (j - i) = i! \cdot (N - i)!. \quad (33)$$



Using Equation (33) and Inequality (32), we can bound  $|e_i(x)|$  by

$$\begin{aligned}
|e_i(x)| &\leq \frac{x!/(x-N)!}{i! \cdot (N-i)!} \\
&= \frac{x!}{(x-N)! \cdot N!} \cdot \frac{N!}{i! \cdot (N-i)!} \\
&= \binom{x}{N} \cdot \binom{N}{i}.
\end{aligned} \tag{34}$$

$$\tag{35}$$

Combining Expression (34) with Equation (31), we can bound  $|P(x)|$  by

$$\begin{aligned}
|P(x)| &\leq \sum_{i=0}^{\varepsilon \cdot n} |e_i(x)| \cdot a^i \\
&\leq \sum_{i=0}^{\varepsilon \cdot n} \binom{x}{N} \cdot \binom{N}{i} \cdot a^i \\
&= \binom{x}{N} \cdot \sum_{i=0}^{\varepsilon \cdot n} \binom{N}{i} \cdot a^i.
\end{aligned} \tag{36}$$

Now, we are going to bound  $\sum_{i \in \{0, \dots, \varepsilon \cdot n\}} \binom{N}{i} \cdot a^i$ , as it is independent of the variable  $x$ . Note that for  $i \in \{0, \dots, \varepsilon \cdot n - 1\}$ , we have

$$\begin{aligned}
&\left[ \binom{N}{i+1} \cdot a^{i+1} \right] / \left[ \binom{N}{i} \cdot a^i \right] \\
&= \frac{N-i}{i+1} \cdot a \\
&\geq \frac{N-\varepsilon \cdot n}{\varepsilon \cdot n} \cdot a && (i \leq \varepsilon \cdot n - 1) \\
&\geq \frac{\left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n - \varepsilon \cdot n}{\varepsilon \cdot n} \cdot a && (N \geq \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n) \\
&\geq \frac{10}{a} \cdot a \geq 2.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{i \in \{0, \dots, \varepsilon \cdot n\}} \binom{N}{i} \cdot a^i \\
&\leq \sum_{i \in \{0, \dots, \varepsilon \cdot n\}} \binom{N}{\varepsilon \cdot n} \cdot a^{\varepsilon \cdot n} \cdot 2^{-\varepsilon \cdot n + i} \\
&\leq 2 \cdot \binom{N}{\varepsilon \cdot n} \cdot a^{\varepsilon \cdot n}.
\end{aligned} \tag{37}$$

$$\tag{38}$$

Combining Expression (37) with Expression (36), it follows, in order to establish that Equation (27) holds, it suffices to show that

$$2 \cdot \binom{x}{N} \cdot \binom{N}{\varepsilon \cdot n} \cdot a^{\varepsilon \cdot n} / a^x \leq 1/2 \text{ for } x \in \{N+1, \dots, n\}. \quad (39)$$

Note the left side of inequality (39) is maximized if and only if the function

$$f(x) := \binom{x}{N} / a^x \quad (40)$$

is maximized. So now we are going to derive the value  $x^* \in \{N+1, \dots, n\}$  maximizing  $f(x^*)$ .

When  $x \in \{N+1, \dots, n\}$ , we have

$$\begin{aligned} f(x+1)/f(x) &= \left[ \binom{x+1}{N} / a^{x+1} \right] / \left[ \binom{x}{N} / a^x \right] \\ &= \frac{x+1}{a(x-N+1)} = \frac{1}{a} \cdot \left( 1 + \frac{N}{x-N+1} \right). \end{aligned} \quad (41)$$

By Equation (41), we can see that  $f(x+1)/f(x)$  is a decreasing function in  $x$ . Therefore,  $f(x)$  is maximized when  $x$  is the smallest integer such that  $f(x+1)/f(x) \leq 1$ , which is equivalent to

$$\begin{aligned} 1 + \frac{N}{x-N+1} &\leq a. \\ \implies \frac{N}{x-N+1} &\leq a-1 \\ \implies (a-1) \cdot x &\geq a \cdot N - (a-1). \end{aligned}$$

Therefore, the maximizer of  $f(x)$  is

$$x^* = \left\lceil \frac{a \cdot N}{a-1} - 1 \right\rceil. \quad (42)$$

Now, it suffices to verify that inequality (39) holds when  $x = x^*$ , i.e.,

$$2 \cdot \binom{x^*}{N} \cdot \binom{N}{\varepsilon \cdot n} \cdot a^{\varepsilon \cdot n} \leq a^{x^*} / 2. \quad (43)$$

**Establishing Inequality (43).** We claim that

$$\binom{x^*}{N} \leq (2e \cdot a)^{x^* - N}. \quad (44)$$

It is easy to see that  $x^* \geq N$  by Equation (42). When  $x^* = N$ , we have  $\binom{x^*}{N} = 1 \leq (2e \cdot a)^{x^* - N}$ . And when  $x^* > N$ , we have  $x^* - N = \left\lceil \frac{N}{a-1} - 1 \right\rceil \geq 1$ , which in turn means  $\frac{N}{a-1} > 1$ .

If  $\frac{N}{a-1} > 2$ , we have  $\left\lceil \frac{N}{a-1} - 1 \right\rceil \geq \frac{N}{a-1} - 1 \geq \frac{N}{2(a-1)}$ . Otherwise,  $\frac{N}{a-1} \in (1, 2]$ , and we also have  $\left\lceil \frac{N}{a-1} - 1 \right\rceil = 1 \geq \frac{N}{2(a-1)}$ . Putting them together, we can see that when  $x^* > N$ ,

$$\frac{x^*}{x^* - N} = \left\lceil \frac{a \cdot N}{a-1} - 1 \right\rceil / \left\lceil \frac{N}{a-1} - 1 \right\rceil \leq \frac{a \cdot N}{a-1} / \frac{N}{2(a-1)} \leq 2a. \quad (45)$$

Combining Inequality (45) with the inequality  $\binom{n}{m} \leq \left(\frac{en}{m}\right)^m$ , we have

$$\binom{x^*}{N} = \binom{x^*}{x^* - N} \leq \left(\frac{e \cdot x^*}{x^* - N}\right)^{x^* - N} \leq (2e \cdot a)^{x^* - N},$$

when  $x^* > N$ . This proves our Claim (44).

Now we bound  $\binom{N}{\varepsilon \cdot n}$ . As  $N \geq \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n$  and  $a \geq 40$ , we have

$$\frac{e \cdot N}{N - \varepsilon \cdot n} = e \cdot \left(1 + \frac{\varepsilon \cdot n}{N - \varepsilon \cdot n}\right) \leq e \cdot \left(1 + \frac{a}{10}\right) \leq \frac{ae}{5},$$

and

$$\binom{N}{\varepsilon \cdot n} = \binom{N}{N - \varepsilon \cdot n} \leq \left(\frac{e \cdot N}{N - \varepsilon \cdot n}\right)^{N - \varepsilon \cdot n} \leq \left(\frac{ae}{5}\right)^{N - \varepsilon \cdot n}. \quad (46)$$

Putting Inequalities (44) and (46) together, we have

$$2 \cdot \binom{x^*}{N} \cdot \binom{N}{\varepsilon \cdot n} \cdot a^{\varepsilon \cdot n} \quad (47)$$

$$\leq 2 \cdot (2e \cdot a)^{x^* - N} \cdot \left(\frac{ae}{5}\right)^{N - \varepsilon \cdot n} \cdot a^{\varepsilon \cdot n} \quad (48)$$

$$\leq 2 \cdot (2e)^{x^* - N} \cdot \left(\frac{e}{5}\right)^{N - \varepsilon \cdot n} \cdot a^{x^*}. \quad (49)$$

As  $\left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 2 \leq N = \left\lceil \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 2 \right\rceil \leq \left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3$  and  $a \geq 40$ , we have

$$x^* - N = \left\lceil \frac{N}{a-1} - 1 \right\rceil \leq \frac{N}{a-1} \leq \frac{1}{a-1} \cdot \left[\left(1 + \frac{10}{a}\right) \cdot \varepsilon \cdot n + 3\right] \leq \frac{2}{a} \cdot \varepsilon \cdot n + \frac{1}{10},$$

and

$$N - \varepsilon \cdot n \geq \frac{10}{a} \cdot \varepsilon \cdot n + 2.$$

Therefore, we can further bound (49) by

$$2 \cdot 2^{\frac{2}{a} \cdot \varepsilon \cdot n + \frac{1}{10}} \cdot \left(\frac{e}{5}\right)^{\frac{10}{a} \cdot \varepsilon \cdot n + 2} \cdot a^{x^*} \leq a^{x^*} / 2,$$

which establishes Inequality (43) and completes the proof.  $\square$

Now we provide the simple proof for Lemma 3.1. We first restate it for the reader's convenience.

**Lemma 3.1** (restated) *Let  $f: D \rightarrow \{0, 1\}$  with  $D \subseteq \{0, 1\}^M$  be a partial function,  $\varepsilon$  be a real in  $[0, 1/2)$ , and  $d$  be an integer such that  $\deg_\varepsilon(f) > d$ .*

*Let  $\mu: \{0, 1\}^M \rightarrow \mathbb{R}$  be a dual witness to the fact  $\widetilde{\deg}_\varepsilon(f) > d$  as per Theorem 2.4. If  $f$  satisfies the stronger condition that  $\deg_\varepsilon^+(f) > d$ , let  $\mu$  to be a dual witness to the fact that  $\deg_\varepsilon^+(f) > d$  as per Theorem 2.6.*

*We further define  $\mu_+(x) := \max\{0, \mu(x)\}$  and  $\mu_-(x) := -\min\{0, \mu(x)\}$  to be two non-negative real functions on  $\{0, 1\}^M$ , and  $\mu_-^i$  and  $\mu_+^i$  be the restrictions of  $\mu_-$  and  $\mu_+$  on  $f^{-1}(i)$  respectively for  $i \in \{0, 1\}$ . Then the following holds:*

- $\mu_+$  and  $\mu_-$  have disjoint supports.
- $\langle \mu_+, p \rangle = \langle \mu_-, p \rangle$ , for any polynomial  $p$  of degree at most  $d$ . In particular,  $\|\mu_+\|_1 = \|\mu_-\|_1 = \frac{1}{2}$ .
- $\|\mu_+^1\|_1 > \varepsilon$  and  $\|\mu_-^0\|_1 > \varepsilon$ .
- If  $\deg_\varepsilon^+(f) > d$ , then  $\|\mu_+^1\|_1 = 1/2$ .

*Proof of Lemma 3.1.* The first two claims follows directly from Theorem 2.4 and the definitions of  $\mu_+$  and  $\mu_-$ .

For the third claim, by Theorem 2.4, we have

$$\sum_{x \in D} f(x) \cdot \mu(x) - \sum_{x \notin D} |\mu(x)| > \varepsilon.$$

$$\text{Hence, } \|\mu_+^1\|_1 - \|\mu_-^1\|_1 - \sum_{x \notin D} |\mu_-(x)| > \varepsilon.$$

$$\text{This implies that } \|\mu_+^1\|_1 - \|\mu_-^1\|_1 - (0.5 - \|\mu_-^1\|_1 - \|\mu_-^0\|_1) > \varepsilon.$$

$$\text{Hence, } \|\mu_+^1\|_1 - (0.5 - \|\mu_-^0\|_1) > \varepsilon.$$

Therefore,  $\|\mu_+^1\|_1 > \varepsilon$ , and  $(0.5 - \|\mu_-^0\|_1) < 0.5 - \varepsilon$ , which means  $\|\mu_-^0\|_1 > \varepsilon$ .

Finally, the last claim follows directly from Theorem 2.6. □

## C Proof of Theorem 6.1: $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$

In this appendix, we define the communication complexity classes  $\text{NISZK}^{\text{cc}}$  and  $\text{UPP}^{\text{cc}}$  and prove that  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$ .

### C.1 Preliminaries

#### C.1.1 Representation of Boolean Functions

Up until this point of the paper, we have considered functions  $f: \{0, 1\}^n \rightarrow \{0, 1\}$ . However, in order to define and reason about  $\text{UPP}^{\text{cc}}$  communication complexity, it will be highly convenient to consider functions  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  instead, where 1 is interpreted as logical FALSE and  $-1$  is interpreted as logical TRUE. Hence, throughout this appendix, a total Boolean function  $f$  will map  $\{-1, 1\}^n \rightarrow \{-1, 1\}$ . We will sometimes refer to the outputs of  $f$  as TRUE and FALSE rather than  $-1$  and 1 for clarity.

The following additional convention will be highly convenient: we define partial Boolean functions to map undefined inputs to 0. That is, a partial Boolean function on  $\{-1, 1\}^n$  will be thought as a map from  $\{-1, 1\}^n \rightarrow \{-1, 0, 1\}$ . We use  $D_f$  to denote the domain of  $f$ , i.e.,  $D_f := \{x \in \{-1, 1\}^n : f(x) \in \{-1, 1\}\}$ .

### C.1.2 Entropy Estimation

The definition of NISZK (cf. Definition 2.2) is quite technical. Fortunately, there is a simple complete problem for NISZK named Entropy Estimation, which basically asks one to estimate the entropy of a distribution. Formally, it is defined as follows.

**Definition C.1** (Entropy Estimation (EA) [GSV99]). The promise problem Entropy Estimation, denoted  $EA = (EA_{\text{YES}}, EA_{\text{NO}})$ , is defined such that  $EA^{-1}(1) = EA_{\text{YES}}$  and  $EA^{-1}(0) = EA_{\text{NO}}$ , where

$$\begin{aligned} EA_{\text{YES}} &:= \{(X, k) : H(X) > k + 1\} \\ EA_{\text{NO}} &:= \{(X, k) : H(X) < k - 1.\} \end{aligned}$$

Here,  $k$  is an integer specified as part of the input in binary and  $X$  is a distribution encoded as a circuit outputting  $n$  bits.

### C.1.3 Definition of Communication Models

**Sign-Rank and  $UPP^{\text{cc}}$ .** The original definition of *unbounded error communication complexity* ( $UPP^{\text{cc}}$ ) defined by Babai et al. [BFS86] is for *total functions*, but it is straightforward to generalize the definition to partial functions. Consider a partial Boolean function  $f : X \times Y \rightarrow \{-1, 0, 1\}$ . In a UPP protocol of  $f$ , Alice receives an input  $x \in X \subseteq \{-1, 1\}^{n_x}$ , and Bob receives an input  $y \in Y \subseteq \{-1, 1\}^{n_y}$ . Each has an unlimited source of private randomness, and their goal is to compute the joint function  $f(x, y)$  of their inputs with minimal communication for all pair  $(x, y)$  such that  $f(x, y) \neq 0$ . We say the protocol computes  $f$  if for any input  $(x, y) \in f^{-1}(\{-1, 1\})$ , the output of the protocol is correct with probability strictly greater than  $1/2$ . The cost of a protocol for computing  $f$  is the maximum number of bits exchanged on any input  $(x, y)$ . The unbounded error communication complexity  $UPP(f)$  of a function  $f$  is the minimum cost of a protocol computing  $f$ . A partial function  $f$  is in the complexity class  $UPP^{\text{cc}}$  if  $UPP(f) = O(\log^c n)$  for some constant  $c$ , where  $n = \max\{n_x, n_y\}$ .

The *sign-rank* of a matrix  $A$  with entries in  $\mathbb{R}$  is the least rank of a real matrix  $B$  with  $A_{ij} \cdot B_{ij} > 0$  for all  $i, j$  such that  $A_{i,j} \neq 0$ .

Paturi and Simon [PS86] showed that  $UPP(f) = \log(\text{sign-rank}([f(x, y)]_{x \in X, y \in Y})) + O(1)$ .<sup>11</sup> Therefore, the sign-rank characterizes the  $UPP^{\text{cc}}$  complexity of a communication problem.

**Definition of  $NISZK^{\text{cc}}$ .** We now define the  $NISZK^{\text{cc}}$  complexity of a Boolean function  $f$ . This is a natural extension of the original definition by Goldreich, Sahai and Vadhan [GSV99], and follows the canonical method of turning a complexity class into its communication analogue introduced by Babai, Frankl and Simon [BFS86].

**Definition C.2** ( $NISZK^{\text{cc}}$ ). In a NISZK-protocol for a partial Boolean function  $f : X \times Y \rightarrow \{-1, 0, 1\}$  with  $X \subseteq \{-1, 1\}^{n_x}$  and  $Y \subseteq \{-1, 1\}^{n_y}$ , there are three computationally-unbounded parties Alice, Bob, and Merlin. Alice holds an input  $x \in \mathcal{X}$  while Bob holds an input  $y \in \mathcal{Y}$ . The goal of Merlin is to convince Alice and Bob that  $f(x, y) = 1$ , in a non-interactive and zero knowledge fashion.

<sup>11</sup>Paturi and Simon's proof was in the context of total functions, but their result is easily seen to apply to partial functions as well.

Specifically, there is a public random string shared between the three parties  $\sigma \in \{0, 1\}^r$ . Additionally, Alice and Bob can also use shared randomness between them, which is not visible to Merlin. The protocol starts by Merlin sending a message  $m = m(x, y, \sigma)$  to Alice (without loss of generality, we can assume the message is a function of  $x, y$  and  $\sigma$ ). Then Alice and Bob communicate, after which Alice outputs “accept” or “reject”. A NISZK communication protocol for  $f$  also must satisfy additional conditions. The first two are the standard notions of completeness and soundness for probabilistic proof systems.

- **Completeness:** For all  $(x, y) \in f^{-1}(1)$ , there is a strategy  $m^*$  for Merlin that causes Alice to output accept with probability  $\geq 2/3$  (where the probability is taken over both the public random string  $\sigma$  and the shared randomness between Alice and Bob that is not visible to Merlin).
- **Soundness:** For all  $(x, y) \in f^{-1}(0)$ , and for every strategy for Merlin, Alice outputs accept with probability  $\leq 1/3$ .

Let the worst case communication cost be  $w_V$ , where this cost includes both the length of Merlin’s message  $m^*(x, y, \sigma)$  and the total number of bits exchanged by Alice and Bob. Finally, a NISZK communication protocol must also satisfy the following zero knowledge condition

- **Zero Knowledge:** There is a public-coin randomized communication protocol  $S$  with output in  $\{0, 1\}^k$  and worst case communication complexity  $w_S$ , such that for all  $(x, y) \in F^{-1}(1)$ , the statistical distance between the following two distributions is smaller than  $1/n$ , where  $n = \max(n_x, n_y)$ .
  - (A) Choose  $\sigma$  uniformly from  $\{0, 1\}^r$ , and output  $m^*(x, y, \sigma)$ .
  - (B) The output distribution of  $S$  on  $(x, y)$ .

Finally, the cost of the NISZK communication protocol is defined as  $r + \max(w_V, w_S)$ . The quantity  $\text{NISZK}^{\text{cc}}(f)$  is defined as the minimum of the cost of all NISZK protocols for  $f$ .

**Remark C.3.** *In the original definition in [GSV99] (see also Definition 2.2 in Section 2.1 of this work), it is required that the statistical difference between the distributions (A) and (B) is negligible. In the context of communication complexity, where protocols of cost  $\text{polylog}(n)$  are considered efficient, negligible corresponds to  $1/\log^{\omega(1)}(n)$ . However, since for polylogarithmic cost protocols, the difference between the distributions (A) and (B) can be amplified from  $1/\text{polylog}(n)$  to  $1/n^{\omega(1)}$  with a polylogarithmic blowup in cost (cf. Lemma 3.1 in [GSV99]), we simply require the difference to be at most  $1/n$  here.*

#### C.1.4 Approximate Degree, Threshold Degree, and Their Dual Characterizations

Since we are now considering Boolean functions mapping  $\{-1, 1\}$  to  $\{-1, 1\}$  rather than  $\{0, 1\}$ , it is convenient to redefine approximate degree and threshold degree in this new setting and state the appropriate dual formulations.

**Definition C.4.** Let  $f : \{-1, 1\}^M \rightarrow \{-1, 0, 1\}$  be a partial function. Recall that the domain of  $f$  is  $D_f := \{x \in \{-1, 1\}^M : f(x) \in \{-1, 1\}\}$ .

- The *approximate degree* of  $f$  with approximation constant  $0 \leq \varepsilon < 1$ , denoted  $\widetilde{\text{deg}}_\varepsilon(f)$ , is the least degree of a real polynomial  $p : \{-1, 1\}^M \rightarrow \mathbb{R}$  such that  $|p(x) - f(x)| \leq \varepsilon$  when  $x \in D_f$ , and  $|p(x)| \leq 1 + \varepsilon$  for all  $x \notin D_f$ . We refer to such a  $p$  as an *approximating polynomial* for  $f$ . We use  $\widetilde{\text{deg}}(f)$  to denote  $\widetilde{\text{deg}}_{1/3}(f)$ .
- The *threshold degree* of  $f$ , denoted  $\text{deg}_\pm(f)$ , is the least degree of a real polynomial  $p$  such that  $p(x) \cdot f(x) > 0$  for all  $x \in D_f$ .

**Remark C.5.** All the results from earlier in this paper regarding partial functions mapping (subsets of)  $\{0, 1\}^n$  to  $\{0, 1\}$  can be translated to results regarding functions mapping  $\{-1, 1\}^n$  to  $\{-1, 0, 1\}$  as considered in this appendix. Specifically, given a partial function  $f : \{-1, 1\}^M \rightarrow \{-1, 0, 1\}$  with domain  $D_f$ , let  $D_f^{\{0,1\}} := \left\{ \left( \frac{1-x_1}{2}, \dots, \frac{1-x_M}{2} \right) : (x_1, \dots, x_M) \in D_f \right\}$ . Consider the partial function  $f^{\{0,1\}} : D_f^{\{0,1\}} \rightarrow \{0, 1\}$  defined as

$$f^{\{0,1\}}(x_1, \dots, x_M) = \frac{1 - f((-1)^{x_1}, \dots, (-1)^{x_M})}{2}.$$

Then it is easy to see that  $\widetilde{\deg}_\varepsilon(f)$  is equal to  $\widetilde{\deg}_{\varepsilon/2}(f^{\{0,1\}})$ , for any  $\varepsilon \in [0, 1)$ .

We recall the definitions of norm, inner product and high pure degree, now with respect to Boolean representation in  $\{-1, 1\}$ . For a function  $\psi : \{-1, 1\}^M \rightarrow \mathbb{R}$ , define the  $\ell_1$  norm of  $\psi$  by  $\|\psi\|_1 = \sum_{x \in \{-1, 1\}^M} |\psi(x)|$ . If the support of a function  $\psi : \{-1, 1\}^M \rightarrow \mathbb{R}$  is (a subset of) a set  $D \subseteq \{-1, 1\}^M$ , we

will write  $\psi : D \rightarrow \mathbb{R}$ . For functions  $f, \psi : D \rightarrow \mathbb{R}$ , denote their inner product by  $\langle f, \psi \rangle := \sum_{x \in D} f(x)\psi(x)$ .

We say that a function  $\psi : \{-1, 1\}^M \rightarrow \mathbb{R}$  has *pure high degree*  $d$  if  $\psi$  is uncorrelated with any polynomial  $p : \{-1, 1\}^M \rightarrow \mathbb{R}$  of total degree at most  $d$ , i.e., if  $\langle \psi, p \rangle = 0$ .

**Theorem C.6.** Let  $f : \{-1, 1\}^M \rightarrow \{-1, 0, 1\}$  be a partial function and  $\varepsilon$  be a constant in  $[0, 1)$ .  $\widetilde{\deg}_\varepsilon(f) > d$  if and only if there is a real function  $\psi : \{-1, 1\}^M \rightarrow \mathbb{R}$  such that:

1. (Pure high degree):  $\psi$  has pure high degree of  $d$ .
2. (Unit  $\ell_1$ -norm):  $\|\psi\|_1 = 1$ .
3. (Correlation):  $\sum_{x \in D_f} \psi(x)f(x) - \sum_{x \notin D_f} |\psi(x)| > \varepsilon$ .

**Theorem C.7.** Let  $f : \{-1, 1\}^M \rightarrow \{-1, 0, 1\}$  be a partial function.  $\deg_\pm(f) > d$  if and only if there is a real function  $\psi : \{-1, 1\}^M \rightarrow \mathbb{R}$  such that:

1. (Zero Outside of Domain):  $\psi(x) = 0$  when  $x \notin D_f$ .
2. (Pure high degree):  $\psi$  has pure high degree of  $d$ .
3. (Sign Agreement):  $\psi(x) \cdot f(x) \geq 0$  for all  $x \in D_f$ .
4. (Non-triviality):  $\|\psi\|_1 > 0$ .

**Orthogonalizing Distributions.** If  $\psi$  is a dual witness for the fact that  $\deg_\pm(f) > d$  as per Theorem C.7, then  $\psi \cdot f$  is a  $d$ -orthogonalizing distribution for  $f$ , as defined next.

**Definition C.8.** A distribution  $\mu : \{-1, 1\}^n \rightarrow [0, 1]$  is  $d$ -orthogonalizing for a function  $h : \{-1, 1\}^n \rightarrow \{-1, 0, 1\}$  if

$$\mathbb{E}_{x \sim \mu}[h(x)p(x)] = 0$$

for every polynomial  $p : \{-1, 1\}^n \rightarrow \mathbb{R}$  with  $\deg p \leq d$ . In other words,  $\mu$  is  $d$ -orthogonalizing for  $h$  if the function  $\mu(x)h(x)$  has pure high degree  $d$ .

### C.1.5 Pattern Matrices

As indicated in Section 6, Razborov and Sherstov [RS10] showed that in order to turn a function  $f: \{-1, 1\}^n \rightarrow \{-1, 0, 1\}$  that has high threshold degree into a matrix  $M$  with high sign-rank (and hence high UPP<sup>cc</sup> complexity), it suffices to show exhibit a dual witness  $\psi$  to the fact that  $\deg_{\pm}(f)$  is large, such that  $\psi$  satisfies an additional *smoothness* condition. The transformation from  $f$  to the matrix  $M$  relies on the *pattern matrix* method introduced by Sherstov [She11]. Pattern matrices are defined as follows.

Let  $n$  and  $N$  be positive integers for which  $n$  divides  $N$ . Let  $\mathcal{P}(N, n)$  denote the collection of subsets  $S \subset [N]$  for which  $S$  contains exactly one member of each block  $\{1, 2, \dots, N/n\}, \{N/n+1, \dots, 2N/n\}, \dots, \{(n-1)N/n+1, \dots, N\}$ . For  $x \in \{-1, 1\}^N$  and  $S \in \mathcal{P}(N, n)$ , let  $x|_S$  denote the restriction of  $x$  to  $S$ , i.e.,  $x|_S = (x_{s_1}, \dots, x_{s_n})$  where  $s_1 < \dots < s_n$  are the elements of  $S$ .

**Definition C.9.** For  $\phi: \{-1, 1\}^n \rightarrow \mathbb{R}$ , the  $(N, n, \phi)$ -pattern matrix  $M$  is given by

$$M = [\phi(x|_S \oplus w)]_{x \in \{-1, 1\}^N, (S, w) \in \mathcal{P}(N, n) \times \{-1, 1\}^n}.$$

Note that  $M$  is a matrix with  $2^N$  rows and  $(N/n)^n 2^n$  columns.

When  $\phi$  is partial function  $\{-1, 1\}^n \rightarrow \{-1, 0, 1\}$ , the  $(N, n, \phi)$ -pattern matrix  $M$  can be viewed as a communication problem as follows: Alice and Bob aim to evaluate the  $\phi(u)$  for some “hidden” input  $u$  to  $\phi$ . Alice gets a sequence of bits  $x \in \{-1, 1\}^N$ , while Bob gets a set of coordinates  $S = \{s_1, \dots, s_n\}$  and a shift  $w \in \{-1, 1\}^n$ . The hidden input  $u$  is simply  $x|_S \oplus w$ .

Note that  $M_{x,y}$  is defined (i.e.,  $M_{x,y} \neq 0$ ) if and only if  $\phi(u)$  is defined ( $\phi(u) \neq 0$ ). This is the reason we represent a partial function by a function  $\{-1, 1\}^n \rightarrow \{-1, 0, 1\}$ .

### C.1.6 Symmetrization

We introduce the notion of symmetrization in this subsection, which is one of the key technical ingredients in our proof.

**Definition C.10.** Let  $T: \{-1, 1\}^k \rightarrow D$ , where  $D$  is a finite subset of  $\mathbb{R}^n$  for some  $n \in \mathbb{N}$ . The map  $T$  is *degree non-increasing* if for every polynomial  $p: \{-1, 1\}^k \rightarrow \mathbb{R}$ , there exists a polynomial  $q: D \rightarrow \mathbb{R}$  with  $\deg(q) \leq \deg(p)$  such that

$$q(T(x)) = \mathbb{E}_{y \text{ s.t. } T(y)=T(x)} [p(y)]$$

for every  $x \in \{-1, 1\}^k$ . We say that a degree non-increasing map  $T$  *symmetrizes* a function  $f: \{-1, 1\}^k \rightarrow \mathbb{R}$  if  $f(x) = f(y)$  whenever  $T(x) = T(y)$ , and in this case we say that  $T$  is a symmetrization for  $f$ .

For any function  $\psi: \{-1, 1\}^k \rightarrow \mathbb{R}$ , a symmetrization  $T: \{-1, 1\}^k \rightarrow D$  for  $\psi$  induces a symmetrization function  $\tilde{\psi}: D \rightarrow \mathbb{R}$  defined as  $\tilde{\psi}(z) := \mathbb{E}_{x \in T^{-1}(z)} [\psi(x)]$  (if  $T^{-1}(z)$  is empty, we let  $\tilde{\psi}(z)$  to be 0). It will also be convenient to define an “unnormalized” version  $\hat{\psi}$  of  $\tilde{\psi}$ , defined via  $\hat{\psi}(z) := \sum_{x \in T^{-1}(z)} \psi(x)$ .

Observe that if  $\mu$  is a distribution on  $\{-1, 1\}^k$ , then  $\hat{\mu}$  is a distribution on  $D$ .

Let  $T: \{-1, 1\}^k \rightarrow D$  be a degree non-increasing map. A function  $\hat{\psi}: D \rightarrow \mathbb{R}$  naturally induces an un-symmetrized function  $\psi: \{-1, 1\}^k \rightarrow \mathbb{R}$  by setting  $\psi(x) = \frac{1}{T^{-1}(z)} \hat{\psi}(z)$  where  $z = T(x)$ . That is,  $\psi$  spreads the mass of  $\hat{\psi}(z)$  out evenly over points  $x \in T^{-1}(z)$ . Observe that, for any  $\hat{\psi}$  and any degree non-increasing map  $T$ , the induced function  $\psi$  is symmetrized by  $T$ .



### C.1.7 Dual Objects

A key technical ingredient in Bun and Thaler's [BT16] methodology for proving sign-rank lower bounds is the notion of a *dual object* for a Boolean function  $f$ , which is roughly a dual witness  $\psi$  for the high one-sided approximate degree of  $f$ , that satisfies additional metric properties. We introduce a related definition below. The difference between our definition of a dual object and Bun and Thaler's is that our definition only requires  $\psi$  to witness the high approximate degree (rather than one-sided approximate degree) of  $f$ .

**Definition C.11** (Dual Object). Let  $f : \{-1, 1\}^k \rightarrow \{-1, 0, 1\}$  be a partial function, and let  $T : \{-1, 1\}^k \rightarrow D$  be a degree non-increasing symmetrization for  $f$ . Let  $\hat{\psi} : D \rightarrow \mathbb{R}$  be any function, and let  $\psi$  be the associated function on  $\{-1, 1\}^k$  induced by  $T$ . We say that  $\hat{\psi}$  is a  $(d, \varepsilon, \eta)$ -dual object for  $f$  (with respect to  $T$ ) if:

- $$\sum_{x \in D_f} \psi(x)f(x) - \sum_{x \in \{-1, 1\}^k \setminus D_f} |\psi(x)| > \varepsilon \quad (50)$$

- $$\|\psi\|_1 = 1 \quad (51)$$

- $$\psi \text{ has pure high degree at least } d \quad (52)$$

- $$\hat{\psi}(z_+) \geq \eta \text{ for some } z_+ \text{ satisfying } \hat{f}(z_+) = 1 \quad (53)$$

**Remark C.12.** *The first three Conditions (50), (51) and (52) together are equivalent to requiring that  $\psi$  is a dual witness for  $\deg_\varepsilon(f) > d$  as per Theorem C.6. Condition (53) is an additional metric property that is crucial for the construction of a smooth orthogonalizing distribution of  $\text{GapMaj}(f)$ .*

## C.2 The Key Technical Ingredient

We now define a class of functions  $\mathcal{C}_{d,a}$ , and establish that for appropriate values of  $d$  and  $a$ , for any function  $f \in \mathcal{C}_{d,a}$ , it holds that the pattern matrix of  $\text{GapMaj}(f)$  has large sign-rank.

**Definition C.13.** Let  $f : \{-1, 1\}^k \rightarrow \{-1, 0, 1\}$  be a partial function and let  $d, a > 0$ . Then  $f$  is in class  $\mathcal{C}_{d,a}$  if there exists a symmetrization  $T : \{-1, 1\}^k \rightarrow D$  for  $f$  such that:

- There exists a  $(d, \varepsilon, \varepsilon/2)$ -dual object for  $f$  with respect to  $T$ , such that  $\frac{\varepsilon}{1 - \varepsilon} > a$ . (54)

- $f$  evaluates to **FALSE** (i.e.  $f(x) = 1$ ) for all but at most a  $2^{-d}$  fraction of inputs in  $\{-1, 1\}^k$ . (55)

Now we are ready to state the key technical claim that we use (cf. Section C.3 below) to separate  $\text{NISZK}^{\text{cc}}$  and  $\text{UPP}^{\text{cc}}$ .

**Theorem C.14.** *Let  $\varepsilon \in (0, 0.1)$ , consider a partial function  $f : \{-1, 1\}^k \rightarrow \{-1, 0, 1\} \in \mathcal{C}_{m, 40/\varepsilon}$  such that  $\varepsilon \cdot m > 50$ . Let  $F := \text{GapMaj}_{m, 1-\varepsilon}(f)$  and  $n := m \cdot k$ . Then the  $(2^{36+6 \log \varepsilon^{-1}} \cdot n, n, F)$ -pattern matrix  $M$  has sign-rank of  $\exp(\Omega(\varepsilon \cdot m))$ .*

### C.2.1 Proof for Theorem C.14

We need the following theorem for lower bounding sign-rank, which is implicit in [RS10, Theorem 1.1].

**Theorem C.15** (Implicit in [RS10, Theorem 1.1]). *Let  $h: \{-1, 1\}^n \rightarrow \{-1, 0, 1\}$  be a Boolean function and  $\alpha > 1$  be a real number. Suppose there exists a  $d$ -orthogonalizing distribution  $\mu$  for  $h$  such that  $\mu(x) \geq 2^{-\alpha d} 2^{-n}$  for all but a  $2^{-\Omega(d)}$  fraction of inputs  $x \in \{-1, 1\}^n$ . Then the  $(2^{3\alpha} n, n, h)$ -pattern matrix  $M$  has sign rank  $\exp(\Omega(d))$ .*

For a partial function  $f: \{-1, 1\}^n \rightarrow \{-1, 0, 1\}$  and its gapped majority version  $h_m := \text{GapMaj}_{m, 1-\varepsilon}(f)$ , we use  $Z(h_m)$  to denote the set

$$\{x = (x_1, x_2, \dots, x_m) \in \{-1, 1\}^{nm} : f(x_i) = 1 \text{ for all } x_i\}.$$

That is,  $Z(h_m)$  is the set of inputs to  $h_m$  such that all copies of  $f$  evaluates to FALSE.

The following theorem asserts the existence of the  $d$ -orthogonalizing distribution for  $\text{GapMaj}(f)$  that is needed to apply Theorem C.15.

**Theorem C.16.** *Let  $\varepsilon \in (0, 0.1)$ ,  $m$  be an integer such that  $\varepsilon \cdot m > 50$ , and  $f$  be partial function  $f: \{-1, 1\}^k \rightarrow \{-1, 0, 1\}$  with a  $(d_1, \varepsilon_2, \eta)$ -dual object (with respect to symmetrization  $T: \{-1, 1\}^k \rightarrow D$ ) such that  $\frac{\varepsilon_2}{1 - \varepsilon_2} > 40/\varepsilon$ . Let  $h_m := \text{GapMaj}_{m, 1-\varepsilon}(f)$  and  $d = \min\{d_1, \varepsilon/4 \cdot m\}$ . Then there exists a  $d$ -orthogonalizing distribution  $\mu: \{-1, 1\}^{mk} \rightarrow [0, 1]$  for  $h_m$  such that*

$$\mu(x) \geq 2^{-2d} \cdot \binom{m}{d}^{-2} \cdot (2\eta)^m 2^{-mk}$$

for all  $x \in Z(h_m)$ .

Before proving Theorem C.16, we show that combining it with Theorem C.15 implies Theorem C.14.

*Proof of Theorem C.14.* By Condition (54) of  $f \in \mathcal{C}_{m, 40/\varepsilon}$ ,  $f$  has a  $(m, \varepsilon_2, \varepsilon_2/2)$ -dual object with respect to  $T$  such that  $\frac{\varepsilon_2}{1 - \varepsilon_2} > 40/\varepsilon$ . Applying Theorem C.16, for  $d = \min\{m, \varepsilon/4 \cdot m\} = \varepsilon/4 \cdot m$ , there exists a  $d$ -orthogonalizing distribution  $\mu: \{-1, 1\}^{mk} \rightarrow [0, 1]$  for  $F$  such that  $\mu(x) > 2^{-2d} \cdot \binom{m}{d}^{-2} \cdot \varepsilon_2^m 2^{-mk}$  for all  $x \in Z(F)$ .

By the inequality  $\binom{a}{b} \leq \left(\frac{e \cdot a}{b}\right)^b$ , we have

$$\binom{m}{d} \leq (4e/\varepsilon)^d \leq 2^{(4+\log \varepsilon^{-1})d}. \quad (56)$$

Since  $\frac{\varepsilon_2}{1 - \varepsilon_2} > 40/\varepsilon$ , we have  $1 - \varepsilon_2 < \frac{\varepsilon_2}{40/\varepsilon} \leq \frac{\varepsilon}{40}$ , hence  $\varepsilon_2 > 1 - \varepsilon/40$ . Therefore,

$$\varepsilon_2^m \geq (1 - \varepsilon/40)^m \geq 4^{-\varepsilon m/40} \geq 2^{-d/5}, \quad (57)$$

where the second inequality holds by the inequality  $(1 - x)^{1/x} \geq 1/4$  for all  $x \in (0, 0.5)$ .

Putting Inequalities (56) and (57) together, we have for all  $x \in Z(F)$ ,

$$\mu(x) > 2^{-mk} \cdot 2^{-(2+2(4+\log \varepsilon^{-1})+1/5)d} > 2^{-mk} \cdot 2^{-(12+2 \log \varepsilon^{-1})d}. \quad (58)$$

And by Condition (55) of  $f \in \mathcal{C}_{m,40/\varepsilon}$ ,  $f(x) = 1$  for all but at most a  $2^{-m}$  fraction of inputs in  $\{-1, 1\}^k$ . Hence, by a union bound, there is at most a  $m \cdot 2^{-m} \leq 2^{-m/2} \leq 2^{-d}$  fraction of inputs do not belong to  $Z(\text{GapMaj}_{d,1-\varepsilon}(f)) = Z(F)$ .

By the above fact and Inequality (58), we conclude that  $\mu$  is a  $d$ -orthogonalizing for  $F$  such that  $\mu(x) \geq 2^{-mk} \cdot 2^{-(12+2 \log \varepsilon^{-1})d}$  for all but a  $2^{-d}$  fraction of inputs in  $\{-1, 1\}^{mk}$ . Therefore, invoking Theorem C.15, the  $(2^{36+6 \log \varepsilon^{-1}} \cdot n, n, F)$ -pattern matrix  $M$  has sign-rank  $\exp(\Omega(d)) = \exp(\Omega(\varepsilon \cdot m))$ .  $\square$

## C.2.2 Proof of Theorem C.16

**Additional Notation.** Let  $f : \{-1, 1\}^k \rightarrow \{-1, 0, 1\}$  be as in the statement of Theorem C.16, and let  $T : \{-1, 1\}^k \rightarrow D$  be the symmetrization for  $f$  associated with the assumed  $(d_1, \varepsilon_2, \eta)$ -dual object for  $f$ . Define  $T^m : \{-1, 1\}^{mk} \rightarrow D^m$  by  $T^m(x_1, \dots, x_m) := (T(x_1), \dots, T(x_m))$ . Since  $T$  is degree non-increasing, it is easy to see that  $T^m$  is also degree non-increasing. Moreover,  $T^m$  is a symmetrization for  $h_m$ . The map  $T^m$  induces a symmetrized version  $\tilde{h}_m : D^M \rightarrow \mathbb{R}$  of  $h_m$  given by  $\tilde{h}_m = \text{GapMaj}_{m,1-\varepsilon}(\tilde{f})$ .

Throughout the proof, we let  $c \in \tilde{f}^{-1}(1)$  denote the point on which the dual object  $\hat{\psi}$  for  $f$  has  $\hat{\psi}(c) \geq \eta$  (cf. Condition (53) within Definition C.11).

**Proof Outline.** Our proof follows roughly the same steps as in [BT16]. Let  $Z^+ := T^m(Z(h_m)) \subseteq D^m$ . At a high level, our proof will produce, for every  $z \in Z^+$ , a  $d$ -orthogonalizing distribution  $\mu_z$  that is targeted to  $z$ , in the sense that

$$\hat{\mu}_z(z) \geq 2^{-O(d)} \cdot \binom{m}{d}^{-2} \cdot (2\eta)^m.$$

Since the property of  $d$ -orthogonalization is preserved under averaging, we construct the final distribution by a convex combination of these constructed distributions  $\mu_z$ 's so that it places the required amount of probability mass on each input  $x \in (T^m)^{-1}(Z^+) = Z(h_m)$ . The goal therefore becomes to construct these targeted distributions  $\mu_z$ . We do this in two stages.

**Stage 1.** In the first stage (see Claim 4 below), we construct distributions  $\mu_z$  for every  $z$  belonging to a highly structured subset  $G \subset Z^+$  that we now describe. The set  $G$  consists of inputs in  $Z^+$  for which  $c$  is repeated many times (specifically, at least  $(1 - \varepsilon/4) \cdot m$  times).

**Stage 2.** In the second stage (see Claim 5 below), we show that given the family of distributions  $\{\mu_z : z \in G\}$  constructed in Stage 1, we can construct appropriate distributions  $\mu_z$  for  $z$  belonging to the entire set  $Z^+$ .

We begin Stage 1 with a lemma.

**Lemma C.17.** *Let  $\ell = (1 - \varepsilon/4) \cdot m$ , and let  $f$ ,  $T$ , and  $T^\ell$  be as above. Consider the partial function  $g_\ell : \{-1, 1\}^{k\ell} \rightarrow \{-1, 0, 1\}$  defined as  $g_\ell := \text{GapMaj}_{\ell,1-2\varepsilon/3}(f)$ . There exists a function  $\psi : \{-1, 1\}^{k\ell} \rightarrow [0, 1]$  symmetrized by  $T^\ell$  with the following properties.*

- $\psi$  is a dual witness for  $\deg_{\pm}(g_\ell) > d$  as per Theorem C.7, where  $d = \min\{\varepsilon/4 \cdot m, d_1\}$ . (59)

- $\|\psi\|_1 = 1$ . (60)

- $\hat{\psi}(\underbrace{c, \dots, c}_{\ell \text{ times}}) \geq (2\eta)^\ell / 6$ . (61)

*Proof.* The proof of this lemma is just analyzing the dual witness constructed for  $\text{GapMaj}_{\ell, 1-2\varepsilon/3}(f)$  in Theorem 3.4. The analysis is as follows.

**Properties of the dual witness constructed in Theorem 3.4.** We formalize some properties of the dual witness constructed by Theorem 3.4 below. The original theorem deals with partial functions with signature  $\{0, 1\}^M \rightarrow \{0, 1\}$ , but with the transformation described in Remark C.5, it is straightforward to obtain from it the following result for partial functions with signature  $\{-1, 1\}^M \rightarrow \{-1, 0, 1\}$ .

**Proposition C.18** (Implicit in Theorem 3.4). *Let  $f : \{-1, 1\}^M \rightarrow \{-1, 0, 1\}$  be a partial function,  $n$  be a sufficiently large integer,  $d$  be an integer,  $\varepsilon \in (0.5, 1)$  and  $\varepsilon_2 \in (0.98, 1)$  be two constants. Let  $a = \frac{\varepsilon_2}{1 - \varepsilon_2}$ ,*

$$N = \min \left( d, \left( 1 - \left( 1 + \frac{10}{a} \right) \cdot \varepsilon \right) \cdot n - 4 \right) \text{ and } F := \text{GapMaj}_{n, \varepsilon}(f).$$

*Suppose  $\widetilde{\text{deg}}_{\varepsilon_2}(f) > d$ , and let  $\mu$  be a dual witness to this fact as per Theorem C.6. Define  $\mu_+(x) := \max\{0, \mu(x)\}$  and  $\mu_-(x) := -\min\{0, \mu(x)\}$  to be two non-negative real functions on  $\{-1, 1\}^M$  (analogous to Lemma 3.1).*

*Then there exists a function  $\psi_{\text{old}} : \{-1, 1\}^{n \cdot M} \rightarrow \mathbb{R}$  such that*

- $\psi_{\text{old}}$  takes non-zero values only on the domain of  $F$ . (62)

- $\psi_{\text{old}}(x) \in \left[ \frac{1}{2} \cdot \prod_i \mu_+(x_i), \frac{3}{2} \cdot \prod_i \mu_+(x_i) \right]$  when  $F(x) = \text{FALSE}$ . (63)

- $-\psi_{\text{old}}(x) \in \left[ \frac{1}{2} \cdot \prod_i \mu_-(x_i), \frac{3}{2} \cdot \prod_i \mu_-(x_i) \right]$  when  $F(x) = \text{TRUE}$ . (64)

- $\psi_{\text{old}}$  has pure high degree  $N$ . (65)

Now, let  $\hat{\varphi}$  be the  $(d_1, \varepsilon_2, \eta)$ -dual object with respect to  $T$ , and  $\varphi$  be the associated function from  $\{-1, 1\}^k \rightarrow \mathbb{R}$ . By Remark C.12,  $\varphi$  is a dual witness for  $\widetilde{\text{deg}}_{\varepsilon_2}(f) > d_1$ . And by assumption,  $\frac{\varepsilon_2}{1 - \varepsilon_2} > 40/\varepsilon > 400$ , which means  $\varepsilon_2 > 0.98$ . So we can invoke Proposition C.18 to construct the function  $\psi_{\text{old}}$  for  $F := g_\ell = \text{GapMaj}_{\ell, 1-2\varepsilon/3}(f)$ .

**Verification of Conditions (59)-(61).** We simply set  $\psi := \psi_{\text{old}} / \|\psi_{\text{old}}\|_1$ . Conditions (60) follows immediately from the definition of  $\psi$ .

To check Condition (59), note that  $\psi_{\text{old}}$  has pure degree

$$\begin{aligned} & \min \left( d_1, \left( 1 - \left( 1 + \frac{10}{40/\varepsilon} \right) \cdot (1 - 2\varepsilon/3) \right) \cdot \ell - 4 \right) \\ & \geq \min \left( d_1, \left( 1 - 1 - \frac{\varepsilon}{4} + \frac{2\varepsilon}{3} + \frac{\varepsilon^2}{6} \right) \cdot (1 - \varepsilon/4) \cdot m - 4 \right) \\ & \geq \min(d_1, \varepsilon/4 \cdot m). \end{aligned} \quad (\varepsilon \cdot m \geq 50 \text{ and } \varepsilon < 0.1)$$

Together with Properties (62), (63) and (64) (recall that -1 represents TRUE and 1 represents FALSE), this implies that  $\psi$  satisfies Condition (59).

Finally, we verify Condition (61). Let  $z_c = \underbrace{(c, \dots, c)}_{\ell \text{ times}}$ . Since  $\tilde{f}(c) = \text{FALSE}$ , we have  $\tilde{g}_\ell(z_c) = \text{FALSE}$  and therefore  $F(x) = \text{FALSE}$  for  $x$  such that  $T^\ell(x) = z_c$ . So we have

$$\begin{aligned}
\hat{\psi}_{\text{old}}(z_c) &= \sum_{x \in (T^\ell)^{-1}(z_c)} \psi_{\text{old}}(x) \\
&\geq \frac{1}{2} \cdot \sum_{x \in (T^\ell)^{-1}(z_c)} \prod_{i=1}^{\ell} \mu_+(x_i) && \text{(Condition (63))} \\
&= \frac{1}{2} \cdot \prod_{i=1}^{\ell} \sum_{x \in T^{-1}(c)} \mu_+(x) \\
&\geq \frac{1}{2} \cdot \prod_{i=1}^{\ell} \hat{\varphi}(c) \geq \frac{1}{2} \cdot \eta^\ell,
\end{aligned}$$

where the second last inequality holds since  $\hat{\varphi}(c) = \sum_{x \in T^{-1}(c)} \varphi(x)$  and  $\varphi(x) \leq \mu_+(x)$  as  $\mu_+(x) := \max\{0, \mu(x)\}$ , and the last inequality is due to Condition (53) from the definition of a  $(d_1, \varepsilon_2, \eta)$ -dual object.

Also, by Properties (63) and (64), combined with the fact that  $\|\mu_+\|_1 = \|\mu_-\|_1 = \frac{1}{2}$ , we have

$$\|\psi_{\text{old}}\|_1 \leq \frac{3}{2} \cdot \left( \|\mu_+\|_1^\ell + \|\mu_-\|_1^\ell \right) = \frac{3}{2} \cdot 2^{-\ell} \cdot 2 = 3 \cdot 2^{-\ell}.$$

Putting them together, we have

$$\hat{\psi}(z_c) \geq \frac{\eta^\ell/2}{\|\psi_{\text{old}}\|_1} = (2\eta)^\ell/6.$$

This establishes Condition (61) and completes the proof of Lemma C.17.  $\square$

With the above lemma, we are now ready to complete Stage 1 by showing that for every input  $w \in D^m$  that is close in Hamming distance to the special point  $z_c = \underbrace{(c, \dots, c)}_{m \text{ times}}$ , there is an orthogonalizing distribution

for  $h_m$  that places substantial weight on  $w$ .

Let  $G$  denote the set of inputs in  $Z^+$  that take the value  $c$  on at least  $(1 - \varepsilon/4) \cdot m$  coordinates. That is,

$$G = \{z \in Z^+ : \sum_{i=1}^m \mathbb{1}_{z_i=c} \geq (1 - \varepsilon/4) \cdot m\}.$$

**Claim 4.** *Let  $G$  be as above. For every  $w = (w_1, \dots, w_m) \in G$ , there exists a  $d$ -orthogonalizing distribution  $\nu_w : \{-1, 1\}^{km} \rightarrow [0, 1]$  for  $h_m$  such that  $\nu_w$  is symmetrized by  $T^m$  and  $\hat{\nu}_w(w) \geq (2\eta)^m/6$ .*

*Proof.* Let  $\ell = (1 - \varepsilon/4) \cdot m$ ,  $I = \{i_1, \dots, i_\ell\}$  denote the first  $\ell$  coordinates on which  $w$  takes the value  $c$ .

Then we define the distribution  $\hat{\nu}_w$  by

$$\hat{\nu}_w(z) = \begin{cases} |\hat{\psi}(z_{i_1}, \dots, z_{i_\ell})| & \text{if } z_i = w_i \text{ for all } i \notin I \\ 0 & \text{otherwise} \end{cases}$$

where  $\hat{\psi}$  is the function from Lemma C.17 for  $g_\ell$ . It is immediate from the definition that  $\hat{\nu}_w$  is a distribution on  $D^m$ , and hence  $\nu_w$  is a distribution on  $\{-1, 1\}^{km}$ . Moreover,  $\hat{\nu}_w(w) \geq (2\eta)^\ell/6 \geq (2\eta)^m/6$ .

To show that  $\nu_w$  is  $d$ -orthogonalizing, let  $p_1, \dots, p_m$  be polynomials over  $\{-1, 1\}^k$  whose degrees sum to at most  $d$ . Let  $\tilde{p}_1, \dots, \tilde{p}_m: D \rightarrow \mathbb{R}$  denote polynomials satisfying the property that for all  $i$  and all  $z$  in the image of  $T$ ,  $\tilde{p}_i(z) := \mathbb{E}_{x \in T^{-1}(z)}[p_i(x)]$ . (Since  $T$  is degree non-increasing, there exist such  $\tilde{p}_i$ 's whose degrees sum to at most  $d$ , but we will not make use of this property in this proof).

Observe that:

$$\begin{aligned}
\sum_{x=(x_1, \dots, x_m) \in \{-1, 1\}^{km}} \nu_w(x) h_m(x) \prod_{i=1}^m p_i(x_i) &= \sum_{z=(z_1, \dots, z_m) \in D^m} \hat{\nu}_w(z) \tilde{h}_m(z) \prod_{i=1}^m \tilde{p}_i(z_i) \\
&= \sum_{\substack{z=(z_1, \dots, z_m) \in D^m \text{ s.t.} \\ \forall i \notin I \ z_i = w_i}} |\hat{\psi}(z_{i_1}, \dots, z_{i_\ell})| \cdot \tilde{h}_m(z) \prod_{i=1}^m \tilde{p}_i(z_i) \\
&= \sum_{\substack{z=(z_1, \dots, z_m) \in D^m \text{ s.t.} \\ \forall i \notin I \ z_i = w_i}} \hat{\psi}(z_{i_1}, \dots, z_{i_\ell}) \prod_{i=1}^m \tilde{p}_i(z_i) \\
&= \left( \prod_{i \notin I} \tilde{p}_i(w_i) \right) \sum_{z=(z_1, \dots, z_\ell) \in D^\ell} \hat{\psi}(z) \prod_{i \in I} \tilde{p}_i(z_i) \\
&= \left( \prod_{i \notin I} \tilde{p}_i(w_i) \right) \sum_{x=(x_1, \dots, x_\ell) \in \{-1, 1\}^{k \cdot \ell}} \psi(x) \prod_{i \in I} p_i(x_i) \\
&= 0.
\end{aligned}$$

Here, the second equality holds by definition of  $\hat{\nu}_w$ , and the final equality holds because  $\psi$  has pure high degree at least  $d$ , and  $\prod_{i \in I} p_i(x)$  is a polynomial of total degree at most  $d$ . To see the second inequality holds,

suppose  $\hat{\psi}(z_{i_1}, \dots, z_{i_\ell}) > 0$ , as  $\psi$  agrees in sign with  $g_\ell$ , we must have  $\tilde{g}_\ell(z_{i_1}, \dots, z_{i_\ell}) = \text{FALSE}$ . Recall that  $g_\ell := \text{GapMaj}_{\ell, 1-2\varepsilon/3}(f)$ . This means that there are at least  $(1 - 2\varepsilon/3) \cdot \ell = (1 - 2\varepsilon/3)(1 - \varepsilon/4) \cdot m \geq (1 - \varepsilon) \cdot m$  copies of  $f$  that evaluate to **FALSE**. Hence  $\tilde{h}_m(z)$  itself must be **FALSE**. So  $|\hat{\psi}(z_{i_1}, \dots, z_{i_\ell})| \cdot \tilde{h}_m(z) = \hat{\psi}(z_{i_1}, \dots, z_{i_\ell})$ . Similarly, when  $\hat{\psi}(z_{i_1}, \dots, z_{i_\ell}) < 0$ , again as  $\psi$  agrees in sign with  $g_\ell$ , we must have  $\tilde{g}_\ell(z_{i_1}, \dots, z_{i_\ell}) = \text{TRUE}$ . So there are at least  $(1 - 2\varepsilon/3) \cdot \ell = (1 - 2\varepsilon/3)(1 - \varepsilon/4) \cdot m \geq (1 - \varepsilon) \cdot m$  copies of  $f$  evaluate to **TRUE**. Hence  $\tilde{h}_m(z)$  must be **TRUE** as well. So  $|\hat{\psi}(z_{i_1}, \dots, z_{i_\ell})| \cdot \tilde{h}_m(z) = -\hat{\psi}(z_{i_1}, \dots, z_{i_\ell}) \cdot -1 = \hat{\psi}(z_{i_1}, \dots, z_{i_\ell})$ . Putting them together, we can see  $|\hat{\psi}(z_{i_1}, \dots, z_{i_\ell})| \cdot \tilde{h}_m(z) = \hat{\psi}(z_{i_1}, \dots, z_{i_\ell})$  for all  $z$  appearing in the summation, so the second inequality holds.  $\square$

**Stage 2.** Now we move to Stage 2. With the distributions constructed for  $w \in G$ , for any point  $v \in Z^+$ , we construct a  $d$ -orthogonalizing distribution that puts significant weight on it.

**Claim 5.** *Let  $G$  be as before, and suppose that for every  $w \in G$  there exists a  $d$ -orthogonalizing distribution  $\nu_w : \{-1, 1\}^{km} \rightarrow [0, 1]$  for  $h_m$  that is symmetrized by  $T^m$ , and satisfies  $\hat{\nu}_w(w) \geq \delta$ . Then for every  $v \in (Z^+ \setminus G)$ , there exists a  $d$ -orthogonalizing distribution  $\rho_v$  for  $h_m$  that is symmetrized by  $T^m$ , and*

$$\hat{\rho}_v(v) \geq 6\delta \cdot 2^{-2d} \cdot \binom{m}{d}^{-2}.$$

The main technical ingredient in the proof of Claim 5 is the construction of a function  $\phi : \{0, 1\}^m \rightarrow \mathbb{R}$  of pure high degree  $d$  for which  $\phi(1^m)$  is ‘‘large’’. This can be viewed as a dual formulation of a bound on the growth of low-degree polynomials. The construction of  $\phi$  appears as part of the proof of such a bound in [RS10].

**Remark C.19.** We choose to state Lemma C.20 below for a function  $\phi : \{0, 1\}^m \rightarrow \mathbb{R}$ , rather than applying our usual convention of working with functions over  $\{-1, 1\}^m$ , because it makes various statements in the proof of Claim 5 cleaner. To clarify the terminology below, we recall that a function  $\phi : \{0, 1\}^m \rightarrow \mathbb{R}$  has pure high degree  $d$  if  $\sum_{x \in \{0, 1\}^m} \phi(x) \cdot p(x) = 0$  for every polynomial  $p : \{0, 1\}^m \rightarrow \mathbb{R}$  of degree at most  $d$ . The Hamming weight function  $|\cdot| : \{0, 1\}^m \rightarrow [m]$  counts the number of 1's in its input, i.e.  $|s| = s_1 + s_2 + \dots + s_m$ .

**Lemma C.20** (cf. [RS10, Proof of Lemma 3.2]). *Let  $d$  be an integer with  $0 \leq d \leq m - 1$ . Then there exists a function  $\phi : \{0, 1\}^m \rightarrow \mathbb{R}$  such that*

- $\phi(1^m) = 1$  (66)

- $\phi(x) = 0$  for all  $d < |x| < m$  (67)

- $\phi$  has pure high degree at least  $d$  (68)

- $\sum_{|x| \leq d} |\phi(x)| \leq 2^d \binom{m}{d}$  (69)

*Proof of Claim 5.* Fix  $v \in (Z^+ \setminus G)$ . Define an auxiliary function  $\hat{\phi}_v : D^m \rightarrow [0, 1]$  as follows. For any  $z = (z_1, \dots, z_m)$ , let

$$\hat{\phi}_v(z) := \sum_{\substack{s \in \{0, 1\}^m \text{ s.t.} \\ \forall i \ z_i = (1-s_i)c + s_i v_i}} \phi(s),$$

where  $\phi$  is as in Lemma C.20, with  $d$  set as in the conclusion of Claim 4 (observe that if there is some  $z_i$  such that  $z_i \neq c$  and  $z_i \neq v_i$ , then  $\hat{\phi}_v(z) = 0$ ).

Letting  $\phi_v$  denote the function on  $\{-1, 1\}^{km}$  induced from  $\hat{\phi}_v$  by  $T^m$ , we record some properties of  $\phi_v$  and  $\hat{\phi}_v$ .

- $\hat{\phi}_v(v) = \phi(1^m) = 1$  (70)

- $\text{supp } \hat{\phi}_v \subset G \cup \{v\}$  (71)

- $\phi_v$  has pure high degree at least  $d$  (72)

- $\|\phi_v\|_1 \leq 2^d \binom{m}{d} + 1$  (73)

- $\hat{\phi}_v$  is supported on at most  $d \binom{m}{d}$  points in  $D^m$  (74)

**Verifying Conditions (70)-(74).** For  $s \in \{0, 1\}^m$ , we define

$$\tau(s) := ((1 - s_1)c + s_1 v_1, \dots, (1 - s_m)c + s_m v_m).$$

Then we can see  $\hat{\phi}_v(z) = \sum_{s \in \tau^{-1}(z)} \phi(s)$ .

To see Condition (70) holds, note that  $v \notin G$ , so there are strictly larger than  $\varepsilon/4 \cdot m = d$  coordinates  $v_i$  in  $v$  satisfies  $v_i \neq c$ . Which means for  $\tau(s) = v$ ,  $|s| > d$ . So by Condition (67), the only  $s \in \tau^{-1}(v)$  satisfying  $\phi(s) \neq 0$  is  $1^m$ , and we have  $\hat{\phi}_v(v) = \phi(1^m) = 1$ .

To verify Conditions (71) and (74), note that when  $\hat{\phi}_v(z) > 0$  for  $z \neq v$ , it means there exists some  $s$  with  $|s| \leq d$ , such that  $z = \tau(s)$ . By the definition of  $\tau$ , it means  $z$  takes the value  $c$  on at least  $m - |s| \geq m - d = (1 - \varepsilon/4) \cdot m$  coordinates, therefore  $z \in G$ . Moreover, one can see there are at most  $\sum_{i=0}^d \binom{m}{i} \leq d \cdot \binom{m}{d}$  such  $z$ 's.

For Condition (72), it is enough to show that if  $p_1, \dots, p_m$  are polynomials over  $\{-1, 1\}^k$  whose degrees sum to at most  $d$ , then  $\sum_{x=(x_1, \dots, x_m) \in \{-1, 1\}^{km}} \phi_v(x) \prod_{i=1}^m p_i(x_i) = 0$ . To establish this, let  $\tilde{p}_1, \dots, \tilde{p}_m: D \rightarrow \mathbb{R}$  denote polynomials satisfying  $\deg(\tilde{p}_i) \leq \deg(p_i)$ , and such that for all  $i$  and all  $z_i$  in the image of  $T$ ,  $\tilde{p}_i(z_i) := \mathbb{E}_{x \in T^{-1}(z_i)}[p_i(x_i)]$ . Such polynomials are guaranteed to exist, since  $T$  is degree non-increasing. Then:

$$\begin{aligned} \sum_{x=(x_1, \dots, x_m) \in \{-1, 1\}^{km}} \phi_v(x) \prod_{i=1}^m p_i(x_i) &= \sum_{z=(z_1, \dots, z_m) \in D^m} \hat{\phi}_v(z) \prod_{i=1}^m \tilde{p}_i(z_i) \\ &= \sum_{z=(z_1, \dots, z_m) \in D^m} \left( \sum_{\substack{s \in \{0, 1\}^m \text{ s.t.} \\ \forall i \ z_i = (1-s_i)c + s_i v_i}} \phi(s) \right) \prod_{i=1}^m \tilde{p}_i(z_i) \\ &= \sum_{s \in \{0, 1\}^m} \phi(s) \prod_{i=1}^m \tilde{p}_i((1-s_i)c + s_i v_i) \\ &= 0, \end{aligned}$$

To see that the final equality holds, recall that that degrees of the polynomials  $\tilde{p}_i$  sum to at most  $d$ . Hence,  $p(s_1, \dots, s_m) := \prod_{i=1}^m \tilde{p}_i((1-s_i)c + s_i v_i)$  is a polynomial of degree strictly at most  $d$  over  $\{0, 1\}^m$ . The final equality then follows from the fact that  $\phi$  has pure high degree at least  $d$ .

To establish Condition (73), we check that

$$\sum_{z \in D^m, z \neq v} |\hat{\phi}_v(z)| \leq \sum_{s \in \{0, 1\}^m, s \neq 1^m} |\phi(s)| \leq 2^d \binom{m}{d},$$

where the final inequality holds by Condition (69).

**Construction and analysis of  $\rho_v$ .** Up to normalization, the function  $\phi_v \cdot h_m$  has all of the properties that we need to establish Claim 5, except that there are locations where it may be negative. We obtain our desired orthogonalizing distribution  $\rho_v$  by adding correction terms to  $\hat{\phi}_v$  in the locations where  $\hat{\phi}_v$  may disagree with  $\tilde{h}_m$  in sign. These correction terms are derived from the distributions  $\hat{\nu}_w$  whose existence are hypothesized in the statement of Claim 5. We start by defining

$$\hat{P}_v(z) = \frac{\delta}{2^d \binom{m}{d} + 1} \tilde{h}_m(z) \hat{\phi}_v(z) + \sum_{w \in (\text{supp } \hat{\phi}_v \setminus \{v\})} \hat{\nu}_w(z). \quad (75)$$

Observe that each  $w$  appearing in the sum on the right hand side of (75) is in the set  $G$ , owing to Condition (71). This guarantees that each term  $\hat{\nu}_w$  in the sum is well-defined.

Now we check that  $\hat{P}_v$  is nonnegative. Since each term  $\hat{\nu}_w$  appearing in the sum on the right hand side of (75) is a distribution (and hence non-negative), it suffices to check that  $\hat{P}_v(z) \geq 0$  for each point



$z \in \text{supp } \hat{\phi}_v$ . On each such point with  $z \neq v$ , Condition (73) guarantees that  $\frac{\delta}{2^d \binom{m}{d} + 1} \tilde{h}_m(z) \hat{\phi}_v(z) \geq -\delta$ .

Moreover, the contribution of the sum is at least  $\hat{\nu}_z(z) \geq \delta$  by hypothesis. Hence,  $\hat{P}_v$  is a non-negative function.

Next, we check that normalizing  $\hat{P}_v$  yields a distribution  $\hat{\rho}_v := \hat{P}_v / \|\hat{P}_v\|_1$  for which  $\hat{\rho}_v(v) \geq 6\delta \cdot 2^{-2d} \cdot \binom{m}{d}^{-2}$  as required. By construction,  $\hat{P}_v(v) = \delta / \left(2^d \binom{m}{d} + 1\right)$ . Moreover, Conditions (70), (73), and (74) together show that  $\|\hat{P}_v\|_1 \leq \delta + d \binom{m}{d} \leq 2d \binom{m}{d}$ . Hence,

$$\hat{P}_v(v) \geq \delta / \left(2d \binom{m}{d} \cdot \left(2^d \binom{m}{d} + 1\right)\right) \geq 6\delta / \left(2^{2d} \cdot \binom{m}{d}^2\right),$$

as  $d = \varepsilon/4 \cdot m \geq 10$  by assumption.

Finally, we must check that  $\rho_v = P_v / \|P_v\|_1$  is  $d$ -orthogonalizing for  $h_m$ . To see this, observe that  $P_v \cdot h_m$  is a linear combination of the functions  $\phi_w$  and  $\nu_w \cdot h_m$  for  $w \in (\text{supp } \hat{\phi}_v \setminus \{v\})$ . Moreover, each of these functions has pure high degree at least  $d$  ( $\phi_w$  does so by Condition (72), while  $\nu_w \cdot h_m$  does by the fact that  $\nu_w$  is  $d$ -orthogonalizing for  $h_m$ ). By linearity, it follows that  $P_v \cdot h_m$  has pure high degree at least  $d$ , so  $\rho_v$  is  $d$ -orthogonalizing for  $h_m$  as desired.

This completes the proof of Claim 5.  $\square$

At last we are ready to conclude the proof of Theorem C.16. By Claim 4, for every  $w \in G$  there exists a  $d$ -orthogonalizing distribution  $\nu_w : \{-1, 1\}^{km} \rightarrow [0, 1]$  for  $h_m$  that is symmetrized by  $T^m$ , with  $\hat{\nu}_w(w) \geq (2\eta)^m / 6$ . Thus, by Claim 5, it is also true that for every  $v \in (Z^+ \setminus G)$ , there is a  $d$ -orthogonalizing distribution  $\rho_v : \{-1, 1\}^{km} \rightarrow [0, 1]$  that is symmetrized by  $T^m$ , with  $\hat{\rho}_v(v) \geq (2\eta)^m \cdot 2^{-2d} \cdot \binom{m}{d}^{-2}$ . Now, for each element  $z \in Z^+$ , we define its weight,  $W_z = |(T^m)^{-1}(z)|$ . Consider the following distribution:

$$\hat{\mu}(z) = \left(\sum_{z \in Z^+} W_z\right)^{-1} \cdot \left(\sum_{w \in G} W_w \cdot \hat{\nu}_w(z) + \sum_{v \in (Z^+ \setminus G)} W_v \cdot \hat{\rho}_v(z)\right).$$

We verify that the (un-symmetrized) distribution  $\mu : (\{-1, 1\}^k)^m \rightarrow [0, 1]$  satisfies our requirements. As  $\hat{\mu}$  is a convex combination of  $d$ -orthogonalizing distributions for  $\tilde{h}_m$ , it is itself a  $d$ -orthogonalizing distribution for  $\tilde{h}_m$ , therefore  $\mu$  is a  $d$ -orthogonalizing distribution for  $h_m$ . Now for each  $x \in Z(h_m)$ , let  $w = T^m(x)$ , we have

$$\begin{aligned} \mu(x) &\geq \left(\sum_{z \in Z^+} W_z\right)^{-1} \cdot \frac{1}{W_w} \cdot W_w \cdot (2\eta)^m \cdot 2^{-2d} \cdot \binom{m}{d}^{-2} \\ &= |Z(h_m)|^{-1} (2\eta)^m \cdot 2^{-2d} \cdot \binom{m}{d}^{-2} \\ &\geq 2^{-mk} \cdot (2\eta)^m \cdot 2^{-2d} \cdot \binom{m}{d}^{-2}. \end{aligned}$$

This completes the proof.

### C.3 Exhibiting A Problem Separating $\text{NISZK}^{\text{cc}}$ From $\text{UPP}^{\text{cc}}$

Now we are ready to prove the communication complexity classes separation  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$ . In order to utilize our lifting Theorem C.14, we have to choose a partial function  $f$  that satisfies: (1) it has an efficient  $\text{NISZK}$  protocol, so that  $\text{GapMaj}(f)$  also is in  $\text{NISZK}$ ; (2) it belongs to our partial function class  $\mathcal{C}_{d,a}$  for appropriate choices of  $d$  and  $a$  (cf. Definition C.13). However, it turns out that the  $\text{Col}$  function, which we used in the query complexity case, does not satisfy the second condition in the definition of  $\mathcal{C}_{d,a}$ . In fact,  $\mathcal{C}_{d,a}$  requires the function evaluates to **FALSE** nearly everywhere, but  $\text{Col}$  is undefined on most inputs, as a random function is neither a permutation nor  $k$ -to-1.

To address this issue, we use the PTP problem (cf. Definition 2.15) instead of  $\text{Col}$ . We have the following lemma, showing PTP is the function we want.

**Lemma C.21.**  $\text{PTP}_n \in \mathcal{C}_{d,a}$  for any  $a > 1$  and  $d = \Omega(\sqrt[3]{n}/a)$ .

We defer the proof of Lemma C.21 to Appendix D, restricting ourselves here to a brief sketch as follows. Bun and Thaler [BT16] gave a primal condition that implies the existence of a suitable dual object. The existence of dual object required by  $\mathcal{C}_{d,a}$  (namely, Condition (54)) can be easily proved by combining this primal condition with a simple modification of Kutin’s proof for the approximate degree of  $\text{Col}$ . For Condition (55) of  $\mathcal{C}_{d,a}$ , it suffices to use a Chernoff bound to show a random function from  $[n] \rightarrow [n]$  is far from any permutation. The details of the proof can be found in Appendix D.

Now we are ready to prove the communication complexity class separation  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$ .

**Theorem.** (Restatement of Theorem 6.1)  $\text{NISZK}^{\text{cc}} \not\subseteq \text{UPP}^{\text{cc}}$ .

*Proof.* Let  $k$  be a sufficiently large integer, and  $\varepsilon = 1/10 \log k$ , then by Lemma C.21,  $\text{PTP}_k \in \mathcal{C}_{d,40/\varepsilon}$  for some  $d = \Omega(\sqrt[3]{k}/\log k)$ . Then we define  $F := \text{GapMaj}_{d,1-\varepsilon}(\text{PTP}_k)$ ,  $n := d \cdot k$ , and  $M$  be the  $(N, n, F)$ -pattern matrix for  $N = 2^{36+6 \log \varepsilon^{-1}} \cdot n$ . Invoking Theorem C.14, we have  $\text{UPP}^{\text{cc}}(M) = \Omega(d \cdot \varepsilon) = \Omega(\sqrt[3]{k}/\log^2 k) = \Omega(\sqrt[4]{n}/\log^2 n)$ . Recall that in the communication problem (cf. Section C.1.5), Alice gets an input of length  $N = \text{polylog}(n) \cdot n$ , while Bob gets an input of length at most  $n \cdot \ln N + n = \text{polylog}(n) \cdot n$ . Therefore, we conclude that  $M$  is not in  $\text{UPP}^{\text{cc}}$  by the definition of  $\text{UPP}^{\text{cc}}$  in Section C.1.3.

**Showing it is sufficient to construct an  $\text{NISZK}$  protocol for  $F$ .** Next we show  $M$  is in  $\text{NISZK}^{\text{cc}}$ . We claim that it is enough to show that  $F$  admits an efficient  $\text{NISZK}^{\text{dt}}$  protocol  $P$ . First we recall some notation from Section C.1.5. In the communication problem corresponding to  $M$ , Alice gets a sequence of bits  $x \in \{-1, 1\}^N$ , while Bob gets a set of coordinates  $S = \{s_1, \dots, s_n\}$  and a shift  $w \in \{-1, 1\}^n$ . They together want to compute  $F(u)$  for  $u = x|_S \oplus w$ . Alice and Bob can simulate  $P$  to solve  $M$  as follows: Alice interacts with the prover as if she were the verifier in  $P$ . Whenever she needs the value of  $u_i$ , she asks Bob to send her the value of  $s_i$  and  $w_i$ . Then she knows  $u_i = x_{s_i} \oplus w_i$ . The correctness of this protocol follows directly from the correctness of  $P$ , and it only incurs a logarithmic overhead in the running time. Therefore, if  $F$  admits an efficient  $\text{NISZK}$  protocol, it can be transformed to an efficient  $\text{NISZK}^{\text{cc}}$  protocol for  $M$ .

**Reduction to EA.** Now we prove that  $F$  has an efficient  $\text{NISZK}^{\text{dt}}$  protocol by reducing it to EA, which is  $\text{NISZK}$ -complete (cf. Definition C.1). Given an input  $x \in \{-1, 1\}^{d \cdot k}$  to  $F$ , let  $x = (x_1, \dots, x_d)$ , where  $x_i$  denotes the input to the  $i$ -th copy of  $\text{PTP}_k$ . By the definition of  $\text{PTP}_k$ , we further interpret  $x_i$  as a function  $f_i : [k] \rightarrow [k]$ .

We define the distribution  $\mathcal{D}(x)$  as follows: pick  $i \in [d]$  and  $x \in [k]$  at uniformly random, and then output the pair  $(i, f_i(x))$ . For a function  $f : [k] \rightarrow [k]$ , let  $\mathcal{D}_f$  be the distribution obtained by outputting

$f(x)$  for a uniformly randomly chosen  $x \in [k]$ . Then we can express  $\mathcal{D}(x)$  as  $\mathcal{D}(x) := \frac{1}{d} \cdot \sum_{i=1}^d \{i\} \times \mathcal{D}_{f_i}$ .

Note when  $f$  is a permutation, i.e.  $\text{PTP}(f) = \text{TRUE}$ , we have  $H(\mathcal{D}_f) = \log(k)$ . And when  $\text{PTP}(f) = \text{FALSE}$ , by the definition of  $\text{PTP}_k$ , the size of the support of the distribution  $\mathcal{D}_f$  is at most  $7/8 \cdot k$ , therefore  $H(\mathcal{D}_f) \leq \log(k \cdot 7/8) \leq \log k - 0.18$ .

Also, let the output of  $\mathcal{D}(x)$  be the random variable pair  $(X, Y)$ , note  $Y$  depends on  $X$ , we have

$$H(\mathcal{D}(x)) = H(X, Y) = H(X) + H(Y|X) = \log d + \frac{1}{d} \cdot \sum_{i=1}^d H(\mathcal{D}_{f_i}).$$

With the above observation, we can bound  $H(\mathcal{D}(x))$  easily: when  $F(x) = \text{TRUE}$ ,

$$H(\mathcal{D}(x)) \geq \log d + (1 - \varepsilon) \cdot \log k \geq \log d + \left(1 - \frac{1}{10 \log k}\right) \cdot \log k \geq \log n - \frac{1}{10},$$

as there is at least a  $1 - \varepsilon$  fraction of  $f_i$ 's satisfy  $H(\mathcal{D}_{f_i}) = \log k$  and  $\log d + \log k = \log dk = \log n$ ; when  $F(x) = \text{FALSE}$ ,

$$H(\mathcal{D}(x)) \leq \log d + \varepsilon \cdot \log k + (1 - \varepsilon) \cdot (\log k - 0.18) \leq \log n - 0.15,$$

as there is at least a  $1 - \varepsilon$  fraction of  $f_i$ 's satisfy  $H(\mathcal{D}_{f_i}) \leq \log k - 0.18$ , and  $(1 - \varepsilon) \cdot 0.18 \geq 0.15$  when  $k$  is sufficiently large.

Finally, we take the reduction to be  $A(x) = \mathcal{D}(x)^{\otimes 50}$ , i.e., a sample from  $A(x)$  is a sequence of 50 i.i.d. samples from  $\mathcal{D}(x)$ . Then we can see when  $F(x) = \text{TRUE}$ ,  $H(A(x)) \geq 50 \left(\log n - \frac{1}{10}\right) \geq 50 \log n - 5$ , and when  $F(x) = \text{FALSE}$ ,  $H(A(x)) \leq 50(\log n - 0.15) \leq 50 \log n - 7.5$ . Therefore, the pair  $(A(x), 50 \log n - 6.25)$  is a valid reduction to EA and this completes the proof.  $\square$

## D PTP is in $\mathcal{C}_{d,a}$

In order to show  $\text{PTP} \in \mathcal{C}_{d,a}$ , we need to prove the existence of a suitable dual object, and show that nearly all inputs to  $\text{PTP}_n$  evaluate to **FALSE**.

### D.1 Nearly All Inputs to PTP Evaluate to **FALSE**

We begin with the second condition, which is relatively easy. It is tantamount to verify that nearly every function from  $[n] \rightarrow [n]$  is far from any permutation, which can in turn be proved by a simple application of a Chernoff bound.

**Lemma D.1.** *With probability at least  $1 - 2^{-\Omega(n)}$ , a random function  $f$  from  $[n] \rightarrow [n]$  satisfies  $\text{PTP}_n(f) = \text{FALSE}$ .*

*Proof.* For each  $i$  in  $[n]$ , we define the random variable  $x_i$  to be the indicator of whether  $f^{-1}(i) \neq \emptyset$ , and we let  $X := \frac{1}{n} \cdot \sum_{i=1}^n x_i$ . Then we have  $\Pr[x_i = 0] = \left(1 - \frac{1}{n}\right)^n \approx e^{-1}$ . Since  $x_i$  takes values in  $\{0, 1\}$ , we have  $\mathbb{E}[X] = \mathbb{E}[x_i] \approx 1 - e^{-1}$ .

Although the  $x_i$ 's are not independent, by [DR96], they are negatively associated, which means that we can still apply a Chernoff bound to obtain the following concentration result:

$$\Pr[X > 0.65] \leq e^{-\Omega(n)},$$

as  $1 - e^{-1} \sim 0.63 < 0.65$ . Note that for a function  $f$  with  $X \leq 0.65$ , we have  $\text{PTP}_n(f) = \text{FALSE}$ , as it must differ on at least  $0.35 \cdot n > n/8$  coordinates with any permutation. This completes the proof.  $\square$

## D.2 A Primal Condition

In order to show the existence of a suitable dual object for the PTP problem, we introduce a sufficient *primal* condition that was given in [BT16]. The original statement from [BT16] only considers total functions. But it is easy to observe that the proof in [BT16] makes no use of the fact that the function is total; hence the original proof works for partial functions as well.

**Definition D.2.** Let  $T : \{-1, 1\}^k \rightarrow D$  be a symmetrization for a partial function  $f : \{-1, 1\}^k \rightarrow \{-1, 0, 1\}$ . Let  $V = T^{-1}(\tilde{V})$  for some  $\tilde{V} \subseteq \tilde{f}^{-1}(1)$ . We say that  $p : \{-1, 1\}^k \rightarrow \mathbb{R}$  is a weak  $\varepsilon$ -error one-sided approximation to  $f$  under the promise that the input  $x$  is in  $V \cup f^{-1}(-1)$  (with respect to  $T$ ) if the following holds. Define  $q : \{-1, 1\}^k \rightarrow \mathbb{R}$  by  $q(x) := \mathbb{E}_{y:T(y)=T(x)}[p(y)]$ . Then  $q$  satisfies the following three properties:

- $q(x) \leq -1 + \varepsilon$  for all  $x \in f^{-1}(-1)$ .
- $|q(x) - 1| \leq \varepsilon$  for all  $x \in V$ .
- $|q(x)| \leq 1 + \varepsilon$  for all  $x \in \{-1, 1\}^k$  such that  $f(x) \neq 0$  and  $x \notin (f^{-1}(-1) \cup V)$ .

**Theorem D.3** (Essentially Theorem B.1 in [BT16]). *Let  $T$  be a symmetrization for  $f$ . Let  $V = T^{-1}(z_+)$  for some  $z_+ \in \tilde{f}^{-1}(1)$ . If there does not exist a weak  $2\eta$ -error, degree- $d$  one-sided approximation to  $f$  under the promise that the input is in  $V \cup f^{-1}(-1)$ , then there exists a function  $\hat{\psi} : D \rightarrow \mathbb{R}$  with  $\psi$  being the associated function on  $\{-1, 1\}^k$  induced by  $T$ , such that*

$$\bullet \quad \langle \psi, f \rangle \geq \varepsilon \text{ and } \psi \text{ only takes non-zero values on } D_f. \quad (76)$$

$$\bullet \quad \|\psi\|_1 = 1 \quad (77)$$

$$\bullet \quad \psi \text{ has pure high degree at least } d \quad (78)$$

$$\bullet \quad f(x) = -1 \implies \psi(x) < 0 \quad (79)$$

$$\bullet \quad \hat{\psi}(z_+) \geq \eta \text{ for some } z_+ \in D \text{ satisfying } \tilde{f}(z_+) = 1 \quad (80)$$

Observe that the conditions in the above theorem are indeed strictly stronger than our requirements for a  $(d, \varepsilon, \eta)$ -dual object.<sup>12</sup> So we have the following corollary.

**Corollary D.4.** *Let  $T$  be a symmetrization for  $f$ . Let  $V = T^{-1}(z_+)$  for some  $z_+ \in \tilde{f}^{-1}(1)$ . If there does not exist a weak  $2\eta$ -error, degree- $d$  one-sided approximation to  $f$  under the promise that the input is in  $V \cup f^{-1}(-1)$ , then  $f$  has a  $(d, 2\eta, \eta)$ -dual object.*

## D.3 Existence of a Suitable Dual Object for PTP

We begin by defining a natural symmetrization  $T : \{-1, 1\}^M \rightarrow D$  for  $\text{PTP}_n$ , which is also used in [BT16].

Let  $x \in \{-1, 1\}^M$  be an input to  $\text{PTP}_n$ , and let  $f$  be the corresponding function from  $[n] \rightarrow [n]$ . We define  $T_1(x) := (|f^{-1}(1)|, |f^{-1}(2)|, \dots, |f^{-1}(n)|) \in \mathbb{R}^n$ . And for any vector  $v \in \mathbb{R}^n$ , we define  $T_2(v)$  be the vector in  $\mathbb{R}^n$  which sorts the coordinates in  $v$  in increasing order. Our final symmetrization is defined as  $T := T_2 \circ T_1$ , that is, first count the number of occurrences of each value in the image, and then sort them in ascending order. It is easy to see that  $T$  symmetrizes PTP, and by Lemma C.2 in [BT16], it is a degree non-increasing map.

Let  $z_+$  be the point in the symmetrized domain  $D$  representing all the 2-to-1 inputs and  $V = T^{-1}(z_+)$ . We prove the following lemma in this subsection.

<sup>12</sup>In our definition of a  $(d, \varepsilon, \eta)$ -dual object, we don't require  $\psi$  to be zero outside of  $D_f$ , and Condition (79) is not demanded as well.

**Lemma D.5.** For all  $\varepsilon \in (0, 1)$ , any weak  $\varepsilon$ -error, degree- $d$  one-sided approximation to  $\text{PTP}_n$  under the promise that the input is in  $V \cup f^{-1}(-1)$ , must have  $d = \Omega\left((1 - \varepsilon) \cdot n^{1/3}\right)$ .

Our proof is a simple modification of Kutin's lower bound for approximate degree of Col [Kut05]. We use a more sophisticated version of a lemma from Paturi [Pat92], as we want an explicit dependence on the approximate error, rather than treating it as a small constant.

Following Kutin [Kut05], we define a special collection of functions which are  $a$ -to-1 on one part of the domain and  $b$ -to-1 on the other part. We call a triple of numbers  $(m, a, b)$  *valid* if  $a|m$  and  $b|(n - m)$ . For each valid triple  $(m, a, b)$ , we define

$$g_{m,a,b}(i) = \begin{cases} \lceil i/a \rceil & \text{if } 1 \leq i \leq m \\ n - \lfloor (n - i)/b \rfloor & \text{if } m < i \leq n. \end{cases}$$

and  $R_{m,a,b} := T^{-1}(T(g_{m,a,b}))$ .

We have the following important lemma from [Kut05].

**Lemma D.6** (Lemma 2.2 in [Kut05]). Let  $P(x)$  be a degree- $d$  polynomial in  $\{-1, 1\}^M$ . For a valid triple  $(m, a, b)$ , define  $Q(m, a, b)$  by

$$Q(m, a, b) = \mathbb{E}_{y: y \in R_{m,a,b}}[P(y)].$$

Then  $Q(m, a, b)$  is a degree- $d$  polynomial in  $m, a, b$ .

Now, suppose there is a weak  $\varepsilon$ -error,  $d$ -degree one-sided approximation  $p$  to  $f$  under the promise that the input is in  $V \cup f^{-1}(-1)$ . We are going to show that  $d = \Omega\left((1 - \varepsilon) \cdot n^{1/3}\right)$ .

By Lemma D.6, let  $c_1 = 1 - \varepsilon$ , then there is a degree- $d$  polynomial  $Q(m, a, b)$  such that

- $Q(m, 1, 1) \leq -1 + \varepsilon = -c_1$  for any  $m$ .
- $Q(m, 2, 2) \in [c_1, 2 - c_1]$  for any  $2|m$ .
- $Q(m, a, b) \in [-2 + c_1, 2 - c_1]$  for any valid  $(m, a, b)$  such that  $\text{PTP}(g_{m,a,b}) = \text{FALSE}$ .

We need the following lemma by Paturi [Pat92].

**Lemma D.7** (Paturi [Pat92]). Let  $a, b$  be two reals and  $q : \mathbb{R} \rightarrow \mathbb{R}$  be a univariate polynomial such that  $|q(j)| \leq \delta$  for all integers  $j \in [a, b]$ , and suppose that  $|q(\lceil x \rceil) - q(x)| \geq c \cdot \delta$  for some  $x \in [a, b]$  and a real  $c \in (0, 1)$ . Then  $\deg(q) = \Omega\left(c \cdot \sqrt{(x - a + 1)(b - x + 1)}\right)$ .

Now we prove Lemma D.5 by showing the polynomial  $Q$  must have degree at least  $d = \Omega\left((1 - \varepsilon) \cdot n^{1/3}\right) = \Omega\left(c_1 \cdot n^{1/3}\right)$ . The following proof basically mimics the original proof in [Kut05].

*Proof of Lemma D.5.* Let  $M = n/2$ , depending on the value of  $Q(M, 1, 2)$ , there are two cases.

- $Q(M, 1, 2) \geq 0$  : Let  $g(x) = Q(M, 1, 2x)$  and  $k$  be the least positive integer such that  $|g(k)| \geq 2$ . Then we have  $|g(x)| \leq c_1$  for all positive integers  $< k$ , and  $g(1) - g(1/2) \geq c$  by assumption. Hence by Theorem D.7, we have

$$d = \Omega(c_1 \cdot \sqrt{k}).$$

Now, let  $c = 2k$  and consider the polynomial  $h(i) = Q(n - ci, 1, c)$ . For any integer  $i$  with  $\lceil n/4c \rceil \leq i \leq \lfloor n/c \rfloor$ , the triple  $(n - ci, 1, c)$  is valid, and it is easy to see  $\text{PTP}$  evaluates to **FALSE** on  $g_{n-ci,1,c}$  for

all those  $i$ 's, as at least  $n/4$  inputs belong to the  $c$ -to-1 part and  $c \geq 2$ . Hence we have  $|h(i)| \leq 2 - c_1$  for  $i$  in that range. But  $\left| h\left(\frac{n}{2c}\right) \right| = |Q(M, 1, c)| = |g(k)| \geq 2$ . Therefore, by Theorem D.7, we have

$$d = \Omega(c_1 \cdot n/c) = \Omega(c_1 \cdot n/k).$$

Putting them together, we have  $d^3 = \Omega(c_1^3 \cdot n/k \cdot k) = \Omega(c_1^3 \cdot n)$ , which means  $d = \Omega(c_1 \cdot n^{1/3})$ .

- $Q(M, 1, 2) < 0$  : We let  $g(x) = Q(M, 2x, 2)$  and  $k$  be the least positive integer such that  $|g(k)| \geq 2$ . Then we have  $g(1) - g(1/2) \geq c_1$ . So again by Theorem D.7, we have  $d = \Omega(c_1 \cdot \sqrt{k})$ .

Then, let  $c = 2k$  and  $h(i) = Q(ci, c, 2)$ . For any integer  $i$  with  $0 \leq i \leq \lfloor n/c \rfloor$ , the triple  $(ci, c, 2)$  is valid (both  $n$  and  $c$  is even), and clearly  $\text{PTP}(g_{ci,c,2}) = \text{FALSE}$  for those  $i$ 's. Hence we have  $|h(i)| \leq 2 - c_1$ . But  $\left| h\left(\frac{n}{2c}\right) \right| = |g(k)| \geq 2$ . Again by Theorem D.7, we have  $d = \Omega(c_1 \cdot n/k)$ .

Similarly, we also have  $d = \Omega(c_1 \cdot n^{1/3})$  in this case. This completes the proof. □

#### D.4 Proof for Lemma C.21

Finally, we prove Lemma C.21.

*Proof of Lemma C.21.* Let  $\eta = \frac{1 - 1/2a}{2}$ , by Lemma D.5 and Corollary D.4, there exists a  $(d, 2\eta, \eta)$ -dual object for  $\text{PTP}_n$  with respect to symmetrization  $T$ , for some  $d = \Omega((1 - 2\eta) \cdot n^{1/3}) = \Omega(n^{1/3}/a)$ . Note  $\frac{2\eta}{1 - 2\eta} > a$ , hence this dual object satisfies Condition (54) of  $\mathcal{C}_{d,a}$ .

And by Lemma D.1, the Condition (55) of  $\mathcal{C}_{d,a}$  follows immediately. □

## E A Weaker Polarization Lower Bound Using Fourier Analysis

Here we show that, if one only cares about black box polarization in the restricted form proposed by Holenstein and Renner [HR05], then one can prove a lower bound against polarization directly using Fourier analysis alone. This may help the readers understand what's going on in the proof. But please note this result is subsumed by our oracle separation between SZK and PP.

**Definition E.1.** An  $(n, \ell, m)$ -special polarizer is a pair of joint distributions over pairs of strings,  $(S^0, R^0)$  and  $(S^1, R^1)$ , where  $S^0$  and  $S^1$  are over  $\{0, 1\}^n$ , and  $R^0$  and  $R^1$  are over  $\{0, 1\}^\ell$ .

For any distributions  $D_0$  and  $D_1$ , we define the polarized distributions  $\widehat{D}_0$  and  $\widehat{D}_1$  resulting from this polarizer as:

$$\widehat{D}_b = (D_{S_1^b}, \dots, D_{S_n^b}, R^b)$$

The polarizer then provides the following guarantees:

$$\begin{aligned} \|D_0 - D_1\| > 2/3 &\implies \|\widehat{D}_0 - \widehat{D}_1\| > 1 - 2^{-m} \\ \|D_0 - D_1\| < 1/3 &\implies \|\widehat{D}_0 - \widehat{D}_1\| < 2^{-m} \end{aligned}$$

An  $(n, \ell)$ -pseudo polarizer is the same, except it doesn't provide the above guarantees.

It is to be noted that the technique for polarizing distance between distributions from [SV03] is a special polarizer. Note also that any  $(n, \ell, m)$ -special polarizer is an  $(n, \ell)$ -pseudo polarizer.

Consider distributions over  $\{0, 1\}^k$ . If there existed a polynomial-time computable  $(n, \ell, m)$ -special polarizer such that  $nk + \ell < 2m$ , then Theorem 7.1 implies that deciding whether pairs of such distributions are close or far can be done in PP. If such a polarizer existed for every  $k$ , then this would imply that SZK is contained in PP because of the completeness of the Statistical Distance problem [SV03]. We rule out this approach of showing such a containment with the following theorem.

**Theorem E.2.** *For any  $(n, \ell, m)$ -special polarizer,  $n = \Omega(m)$ .*

Theorem E.2 follows immediately from the following two lemmas. For any  $\alpha \in [0, 1]$  and bit  $b$ , denote by  $D_b^\alpha$  the distribution over  $\{0, 1\}$  that is equal to  $b$  with probability  $(1 + \alpha)/2$ . It is easy to see that  $\|D_0^\alpha - D_1^\alpha\| = \alpha$ . We denote by  $(\hat{D}_0^\alpha, \hat{D}_1^\alpha)$  the distributions that result from applying the special polarizer in the relevant context to  $(D_0^\alpha, D_1^\alpha)$  and by  $(\tilde{D}_0^\alpha, \tilde{D}_1^\alpha)$  the distributions resulting from the pseudo-polarizer.

**Lemma E.3.** *For any  $(n, \ell)$ -pseudo polarizer and any  $\alpha, \beta \in (0, 1)$  such that  $\alpha > \beta$ ,*

$$\frac{\|\tilde{D}_0^\alpha - \tilde{D}_1^\alpha\|}{\|\tilde{D}_0^\beta - \tilde{D}_1^\beta\|} \leq 2^{(n+\ell)/2} \left(\frac{\alpha}{\beta}\right)^n$$

*Proof.* Throughout the proof, we use the symbols for distributions interchangeably with the symbols for vectors representing their mass functions. For each  $\alpha \in (0, 1)$ , we define the following matrix:

$$B_\alpha = \begin{pmatrix} \frac{1+\alpha}{2} & \frac{1-\alpha}{2} \\ \frac{1-\alpha}{2} & \frac{1+\alpha}{2} \end{pmatrix}$$

Consider any distribution  $p$  over  $\{0, 1\}$ . The distribution obtained by selecting a bit  $b$  according to  $p$  and then sampling  $D_b^\alpha$  is given by  $B_\alpha p$ . This can be extended to the case when  $p$  is over  $\{0, 1\}^n$  – if  $x$  is drawn according to  $p$ , the distribution of  $(D_{x_1}^\alpha, \dots, D_{x_n}^\alpha)$  is given by  $B_\alpha^{\otimes n} p$ .

Further, if  $p_0$  happens to be the distribution of  $(S^0, R^0)$  from an  $(n, \ell, m)$  special polarizer, then  $\tilde{D}_0^\alpha$ , when the polarizer is applied to  $(D_0^\alpha, D_1^\alpha)$ , is given by  $(B_\alpha^{\otimes n} \otimes I^{\otimes \ell})p_0$ , where  $I$  is the  $2 \times 2$  identity matrix. Similarly,  $\tilde{D}_1^\alpha$  would be  $(B_\alpha^{\otimes n} \otimes I^{\otimes \ell})p_1$ . Let  $C_\alpha = (B_\alpha^{\otimes n} \otimes I^{\otimes \ell})$ . We then have:

$$\|\tilde{D}_0^\alpha - \tilde{D}_1^\alpha\| = \frac{1}{2} \|C_\alpha(p_1 - p_0)\|_1$$

Both  $B_\alpha$  and  $I$  have the vectors  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$  as eigenvectors. The corresponding eigenvalues are 1 and  $\alpha$  for  $B_\alpha$ , and both 1 for  $I$ . This implies that the eigenvectors of  $B$  are all possible tensor products of these eigenvectors, and the eigenvalue of such a resulting vector is simply the products of the eigenvalues of the vectors that were tensored.

In different terms, the eigenvectors are  $(\chi_{T_1} \otimes \chi_{T_2})$  for any  $T_1 \subseteq [n]$  and  $T_2 \subseteq [\ell]$ , which are the characters of  $\mathbb{F}_2^{n+\ell}$ , and the eigenvalue of this vector would be  $\alpha^{|T_1|}$ . Since these vectors form a basis, we can write  $p_0 = \sum_{T_1, T_2} \hat{p}_{0, (T_1, T_2)} (\chi_{T_1} \otimes \chi_{T_2})$ .

Using the standard relationships between  $L_1$  and  $L_2$  norms, we have the following inequalities for any

$\alpha, \beta \in (0, 1)$  such that  $\alpha > \beta$ :

$$\begin{aligned}
\frac{\|\widehat{D}_0^\alpha - \widehat{D}_1^\alpha\|}{\|\widehat{D}_0^\beta - \widehat{D}_1^\beta\|} &= \frac{\|C_\alpha(p_1 - p_0)\|_1}{\|C_\beta(p_1 - p_0)\|_1} \\
&\leq 2^{(n+\ell)/2} \frac{\|C_\alpha(p_1 - p_0)\|_2}{\|C_\beta(p_1 - p_0)\|_2} \\
&= 2^{(n+\ell)/2} \frac{\|C_\alpha \sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2}{\|C_\beta \sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2} \\
&= 2^{(n+\ell)/2} \frac{\|\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \alpha^{|T_1|} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2}{\|\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \beta^{|T_1|} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2} \\
&= 2^{(n+\ell)/2} \left( \frac{\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \alpha^{2|T_1|} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})^2}{\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \beta^{2|T_1|} (\widehat{p}_{1,(T_1, T_2)} - \widehat{p}_{0,(T_1, T_2)})^2} \right)^{1/2} \\
&\leq 2^{(n+\ell)/2} \left( \frac{\alpha}{\beta} \right)^n
\end{aligned}$$

where the last inequality follows from the readily verified fact that for any sequences of positive real numbers  $\{a_i\}$ ,  $\{b_i\}$ , and  $\{c_i\}$ ,  $\frac{\sum_i c_i a_i}{\sum_i c_i b_i}$  is at most  $\max_i \frac{a_i}{b_i}$ .  $\square$

**Lemma E.4.** For any  $(n, \ell)$ -pseudo polarizer and any  $\alpha, \beta \in (0, 1)$  such that  $\alpha > \beta$ , there is an  $(n, 1)$ -pseudo polarizer such that:

$$\frac{\|\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha\|}{\|\widetilde{D}_0^\beta - \widetilde{D}_1^\beta\|} \geq \frac{\|\widehat{D}_0^\alpha - \widehat{D}_1^\alpha\|}{\|\widehat{D}_0^\beta - \widehat{D}_1^\beta\|}$$

*Proof.* The lemma follows from the following two easily verified facts about Total Variation distance of joint distributions.

**Fact 1.** For random variables  $X, Y$  and  $Y'$ ,

$$\|(X, Y) - (X, Y')\| = \sum_x \Pr[X = x] \|Y_{|X=x} - Y'_{|X=x}\|$$

**Fact 2.** For random variables  $X_0$  and  $X_1$  and a uniformly distributed bit  $B$ ,

$$\|(B, X_B) - (\overline{B}, X_B)\| = \|X_0 - X_1\|$$

For convenience, we write the resulting distributions from a polarizer as  $\widehat{D}_0^\alpha = (D_{S_0}^\alpha, R^0)$ , etc., which is indeed the structure that these distributions have. From the above two facts, we have the following for a uniformly distributed bit  $B$ :

$$\begin{aligned}
\frac{\|\widehat{D}_0^\alpha - \widehat{D}_1^\alpha\|}{\|\widehat{D}_0^\beta - \widehat{D}_1^\beta\|} &= \frac{\|(D_{S_0}^\alpha, R^0) - (D_{S_1}^\alpha, R^1)\|}{\|(D_{S_0}^\beta, R^0) - (D_{S_1}^\beta, R^1)\|} \\
&= \frac{\|(B, D_{S_B}^\alpha, R^B) - (\overline{B}, D_{S_B}^\alpha, R^B)\|}{\|(B, D_{S_B}^\beta, R^B) - (\overline{B}, D_{S_B}^\beta, R^B)\|} \\
&= \frac{\sum_r \Pr[R_B = r] \|(B, D_{S_B}^\alpha)_{|R^B=r} - (\overline{B}, D_{S_B}^\alpha)_{|R^B=r}\|}{\sum_r \Pr[R_B = r] \|(B, D_{S_B}^\beta)_{|R^B=r} - (\overline{B}, D_{S_B}^\beta)_{|R^B=r}\|} \\
&\leq \max_r \frac{\|(B, D_{S_B}^\alpha)_{|R^B=r} - (\overline{B}, D_{S_B}^\alpha)_{|R^B=r}\|}{\|(B, D_{S_B}^\beta)_{|R^B=r} - (\overline{B}, D_{S_B}^\beta)_{|R^B=r}\|}
\end{aligned}$$



where the last inequality is from the same argument about sequences of positive numbers as the one at the end of the proof of Lemma E.3.

This proves what we need, as for any  $r$ ,  $((B, D_{SB})|_{R_B=r}, (\overline{B}, D_{SB})|_{R_B=r})$  is an  $(n, 1)$ -pseudo polarizer.  $\square$

*Proof of Theorem E.2.* For any  $(n, \ell, m)$ -special polarizer we have the following when  $\alpha = 2/3$  and  $\beta = 1/3$ :

$$\frac{\|\widehat{D}_0^\alpha - \widehat{D}_1^\alpha\|}{\|\widehat{D}_0^\beta - \widehat{D}_1^\beta\|} \geq \frac{1 - 2^{-m}}{2^{-m}} = 2^m - 1$$

Lemmas E.3 and E.4 imply that there is an  $(n, 1)$ -pseudo polarizer such that:

$$\frac{\|\widehat{D}_0^\alpha - \widehat{D}_1^\alpha\|}{\|\widehat{D}_0^\beta - \widehat{D}_1^\beta\|} \leq \frac{\|\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha\|}{\|\widetilde{D}_0^\beta - \widetilde{D}_1^\beta\|} \leq 2^{(n+1)/2} \left(\frac{\alpha}{\beta}\right)^n = 2^{(3n+1)/2}$$

The above two inequalities tell us that  $n = \Omega(m)$ .  $\square$

## F Improved Polarization Places SZK in $\text{BPP}_{\text{path}}$

Here we show that if the Polarization Lemma of Sahai and Vadhan were strengthened in a black box manner, it would imply  $\text{SZK} \subseteq \text{BPP}_{\text{path}}$ . This immediately gives that the Polarization Lemma cannot be strengthened in this manner.

### F.1 Proof of Theorem 7.2

To prove the theorem, suppose that the statistical difference problem is SZK-hard for distributions on  $N$  bits which are either  $\varepsilon$ -close or  $(1 - \varepsilon)$ -far, where  $\varepsilon = o(2^{-2N/3})$ . We will give a  $\text{BPP}_{\text{path}}$  algorithm to solve this problem, using the characterization that  $\text{BPP}_{\text{path}} = \text{postBPP}$ . The algorithm is inspired by Aaronson, Bouland, Fitzsimon, and Lee's proof that  $\text{SZK} \subseteq \text{naCQP}$  given in [ABFL16]. We thank Tomoyuki Morimae and Harumichi Nishimura for helpful discussions on this topic.

The algorithm is as follows: flip three coins  $b_1, b_2, b_3$ , and draw independent samples  $y_1, y_2, y_3$  from the distributions  $D_{b_1}, D_{b_2}, D_{b_3}$ , respectively. Postselect on the condition that  $y_1 = y_2 = y_3$ . Output that the distributions are far apart if  $b_1 = b_2 = b_3$ , and otherwise output that the distributions are close.

If  $\varepsilon = 0$ , then clearly this algorithm is correct. In the case the distributions are far apart, they have disjoint support, which implies the values  $b_i$  must be identical, so in this case the algorithm has zero probability of error. In the case the distributions are close, they are identical, so the string  $b_1 b_2 b_3$  is uniformly random after postselection, so the algorithm errs with probability  $1/4$ . Note that the correctness of this algorithm in the case  $\varepsilon = 0$  doesn't tell us anything new in structural complexity, because in the  $\varepsilon = 0$  case, the problem is in NP (as a witness to the fact the distributions are identical, simply provide  $x_0, x_1$  such that  $P_0(x_0) = P_1(x_1)$ ), and hence is obviously in  $\text{BPP}_{\text{path}}$  and in PP as well.

We now claim that if  $\varepsilon = o(2^{-2N/3})$ , then this algorithm still works. Note that our choice of  $\varepsilon$  is asymptotically tight for our algorithm; if  $\varepsilon = \Omega(2^{-2N/3})$ , then there is a simple counterexample which foils the algorithm<sup>13</sup>. To show that the algorithm works, we'll show two things. First, if the distributions are

<sup>13</sup>Let  $D_0$  be a uniform distribution, and let  $D_1$  be the distribution which places an  $\varepsilon$  amount of weight on a single item  $x$ , while the remaining weight is spread uniformly on the remaining elements. These distributions are  $\varepsilon$ -close in total variation distance, but one can easily show that this algorithm will yield the string  $\hat{b} = 111$  with high probability, and hence the algorithm will incorrectly identify them as being far apart. The reason this counterexample works is that postselecting the distributions on seeing the same outcome  $y_1 = y_2 = y_3$  heavily skews the distributions towards more likely  $y_i$  outputs, and in this example we will almost always have  $y = x$ , and hence will almost always output  $\hat{b} = 111$ .

$\varepsilon$ -close for this small  $\varepsilon$ , then we'll show that as  $n \rightarrow \infty$ , then  $\hat{b}$ 's value approaches the uniform distribution over all 8 possible output strings. Therefore for sufficiently large  $n$ , the algorithm is correct. On the other hand, if the distributions are  $1 - \varepsilon$ -far, we'll show the algorithm is correct with high probability.

Let's first handle the case in which the distributions are  $\varepsilon$ -close. Let  $\hat{b} \in \{0, 1\}^n$  be the random variable corresponding to the output of  $b_1 b_2 b_3$ . Let  $D_b(y)$  denote the probability that distribution  $D_b$  outputs  $y$ . Let  $S$  be the event that  $y_1 = y_2 = y_3$ , and let  $S(y)$  be the event that  $y_1 = y_2 = y_3 = y$ . By Bayes' rule, we have that

$$\begin{aligned} \Pr[\hat{b} = b_1 b_2 b_3 | S] &= \frac{\Pr[S | \hat{b} = b_1 b_2 b_3] \Pr[\hat{b} = b_1 b_2 b_3]}{\Pr[S]} \\ &= \frac{\sum_{y \in \{0,1\}^n} \Pr[S(y) | \hat{b} = b_1 b_2 b_3] \frac{1}{8}}{\sum_{y \in \{0,1\}^n} \Pr[S(y)]} \\ &= \frac{\sum_{y \in \{0,1\}^n} D_1(y)^{w(\hat{b})} D_0(y)^{3-w(\hat{b})} \frac{1}{8}}{\sum_{y \in \{0,1\}^n} \Pr[S(y)]} \end{aligned}$$

where  $w(\hat{b})$  is the Hamming weight of  $\hat{b}$ .

Hence we have that

$$\frac{\Pr[\hat{b} = b_1 b_2 b_3 | S]}{\Pr[\hat{b}' = b'_1 b'_2 b'_3 | S]} = \frac{\sum_{y \in \{0,1\}^n} D_1(y)^{w(\hat{b})} D_0(y)^{3-w(\hat{b})}}{\sum_{y \in \{0,1\}^n} D_1(y)^{w(\hat{b}')} D_0(y)^{3-w(\hat{b}')}}$$

We'll now show that as  $n \rightarrow \infty$ , the ratio of the probabilities between each string tends to 1. Therefore for sufficiently large  $n$ , the strings  $\hat{b}$  can be made arbitrarily close to equiprobable, so the algorithm works. We'll break into three cases, showing that the strings  $\hat{b} = 111$  and  $000$ ,  $100$ , and  $110$  become equiprobable as  $n \rightarrow \infty$ . Since the probability of obtaining a string  $\hat{b}$  is only a function of its hamming weight, this will imply all eight possible outcomes for  $\hat{b}$  become equiprobable for large  $n$ , and hence the error probability of the algorithm approaches  $1/4$  as  $n \rightarrow \infty$ .

### Case 1: 111 and 000

Let's consider the extremal case, where  $\hat{b} = 111$  or  $\hat{b} = 000$ . Let  $\delta_y = |D_1(y) - D_0(y)|$ , so  $\sum_y \delta_y \leq \varepsilon$ , and furthermore that  $D_0(y) \leq D_1(y) + \delta_y$  and  $D_1(y) \leq D_0(y) + \delta_y$ . Therefore we have that

$$\frac{\Pr[\hat{b} = 111 | S]}{\Pr[\hat{b} = 000 | S]} = \frac{\sum_{y \in \{0,1\}^n} D_1(y)^3}{\sum_{y \in \{0,1\}^n} D_0(y)^3} \quad (81)$$

$$\leq \frac{\sum_{y \in \{0,1\}^n} (D_0(y) + \delta_y)^3}{\sum_{y \in \{0,1\}^n} D_0(y)^3} \quad (82)$$

$$= \frac{\sum_{y \in \{0,1\}^n} D_0(y)^3 + 3D_0(y)^2 \delta_y + 3D_0(y) \delta_y^2 + \delta_y^3}{\sum_{y \in \{0,1\}^n} D_0(y)^3} \quad (83)$$

$$= 1 + 3 \frac{\langle \delta, D_0^2 \rangle}{\langle D_0, D_0^2 \rangle} + 3 \frac{\langle \delta^2, D_0 \rangle}{\langle D_0^2, D_0 \rangle} + \frac{|\delta^3|_1}{|D_0^3|_1} \quad (84)$$

$$\leq 1 + 3 \frac{\varepsilon \max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle} + 3 \frac{\varepsilon^2 \max_y D_0(y)}{\langle D_0^2, D_0 \rangle} + \frac{\varepsilon^3}{|D_0^3|_1} \quad (85)$$

$$\leq 1 + 3 \frac{\varepsilon \max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle} + 3 \frac{\varepsilon^2 \max_y D_0(y)}{\langle D_0^2, D_0 \rangle} + \frac{2^{-3cn}}{2^{-2n}} \quad (86)$$

where on line 84 we expressed these sums as inner products, on line 85 we used the fact the sums in the denominators are maximized when the weight of  $\delta$  is placed on a single item, line 86 follows from the fact the denominator is minimized by the uniform distribution. We now need to bound the terms  $\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle}$  and  $\frac{\max_y D_0(y)}{\langle D_0, D_0^2 \rangle}$  as a function of the universe size  $N = 2^n$ . One can easily show that the first is upper bounded by  $\Theta(N^{2/3})$ , and the second is upper bounded by  $\Theta(N^{4/3})$ .  
 To see this, let  $k = \max_y D_0(y)$ , so  $2^{-n} \leq k \leq 1$ . Then we have that

$$\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle} \leq \frac{k^2}{k^3 + \frac{(1-k)^3}{(N-1)^2}}$$

because given  $k$ , the denominator is minimized by spreading the remaining probability mass evenly over the remaining  $N - 1$  elements. By taking the derivative of this as a function of  $k$  and setting it equal to zero, we see that the maximum occurs at a solution to the equation  $k((-5 - (N - 1)^2)k^3 + 12k^2 - 9k + 2) = 0$ . As  $N \rightarrow \infty$  the real roots of this equation are 0 and  $\Theta(N^{-2/3})$  (plus two complex roots), and one can easily show the first is a minimum while the second is the maximum. Hence this quantity is maximized when  $k = \Theta(N^{-2/3})$ , which implies the quantity is upper bounded by

$$\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle} \leq \frac{N^{-4/3}}{N^{-2} + \frac{(1-N^{-2/3})^3}{(N-1)^2}} = \frac{N^{-2/3}}{\Theta(N^{-2})} = \Theta(N^{2/3})$$

A similar proof shows that the second quantity is upper bounded by  $\Theta(N^{4/3})$ .

Therefore we have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 000|S]} \leq 1 + 3 * 2^{-cn} 2^{2n/3} + 3 * 2^{-2cn} 2^{4n/3} + \frac{2^{-3cn}}{2^{-2n}} \quad (87)$$

$$\leq 1 + o(1) \quad (88)$$

since we have  $c > 2/3$ . Note that the identical proof holds for the case where  $D_0$  and  $D_1$  are switched,

therefore we have that  $\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 111|S]} \leq 1 + o(1)$  as well. Hence we have

$$1 - o(1) \leq \frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 000|S]} \leq 1 + o(1)$$

So as  $n \rightarrow \infty$ , these strings become equiprobable.

**Case 2: 111 and 100** We have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 100|S]} = \frac{\sum_{y \in \{0,1\}^n} D_1(y)^3}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} \quad (89)$$

$$\leq \frac{\sum_{y \in \{0,1\}^n} D_1(y)(D_0(y)^2 + 2D_0(y)\delta_y + \delta(y)^2)}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} \quad (90)$$

$$= 1 + 2 \frac{\sum_{y \in \{0,1\}^n} D_1(y)D_0(y)\delta_y}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} + \frac{\sum_{y \in \{0,1\}^n} D_1(y)\delta_y^2}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} \quad (91)$$

$$= 1 + 2 \frac{\langle \delta_y, D_0 D_1 \rangle}{\langle D_0, D_0 D_1 \rangle} + \frac{\langle \delta_y^2, D_1 \rangle}{\langle D_0^2, D_1 \rangle} \quad (92)$$

$$\leq 1 + 2 \frac{\varepsilon \max_y D_0(y) D_1(y)}{\langle D_0, D_0 D_1 \rangle} + \frac{\varepsilon^2 \max_y D_1(y)}{\langle D_0^2, D_1 \rangle} \quad (93)$$

$$\leq 1 + 2\varepsilon \frac{\max_y D_0(y)^2 + \delta_y D_0(y)}{\langle D_0, D_0 D_1 \rangle} + \varepsilon^2 \frac{\max_y D_1(y)}{\langle D_0^2, D_1 \rangle} \quad (94)$$

$$\leq 1 + 2\varepsilon \frac{\max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle - \varepsilon \max_y D_0(y)^2} + 2\varepsilon^2 \frac{\max_y D_0(y)}{\langle D_0, D_0 D_1 \rangle - \varepsilon \max_y D_0(y)^2} \quad (95)$$

$$+ \varepsilon^2 \frac{\max_y D_1(y)}{\langle D_0, D_0^2 \rangle - \varepsilon \max_y D_0(y)^2} \quad (96)$$

Where line 96 comes from the fact that  $D_1(y) \geq D_0(y) - \delta_y$  for all  $y$ . We now show that this is upper bounded by  $1+o(1)$ , by showing that the term  $\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2 \rangle - \varepsilon \max_y D_0(y)^2}$ , the term  $\frac{\max_y D_0(y)}{\langle D_0, D_0 D_1 \rangle - \varepsilon \max_y D_0(y)^2}$  and the term  $\frac{\max_y D_1(y)}{\langle D_0, D_0^2 \rangle - \varepsilon \max_y D_0(y)^2}$  are upper bounded by  $O(2^{2n/3})$ ,  $O(2^{4n/3})$  and  $O(2^{4n/3})$ , respectively. This, combined with the fact that  $\varepsilon = O(2^{-cn})$  for  $c > 2/3$ , implies that  $\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 100|S]} \leq 1 + o(1)$  as desired.

For the first term, let  $k = \max_y D_0(y)$ . The this term is upper bounded by

$$\frac{k^2}{k^3 - \frac{k^3}{(N-1)^2} - \varepsilon k^2}$$

because the denominator is minimized by spreading the remaining probability mass evenly over the remaining  $N - 1$  elements. One can easily show this function is maximized by setting  $k = \Theta(N^{-2/3})$ . Indeed, taking the derivative of this equation and setting it equal to 0, one can see that the extreme values of  $k$  satisfy  $k(-2(N-1)^2 k^3 - 3k + 2) = 0$ . Hence the optimal value of  $k$  satisfies  $k = \Theta(N^{-2/3})$ . For this value of  $k$ , the term evaluates to  $O(N^{2/3})$ .

For the second term, if we let  $k$  be defined as above, then by the same reasoning we have that the term is upper bounded by

$$\frac{k}{k^3 - \frac{k^3}{(N-1)^2} - \varepsilon k^2}$$

Again by a similar proof, one can easily show the function is maximized by setting  $k = O(N^{-2/3})$ , which implies the term is upper bounded by  $O(2^{4n/3})$ .

For the third term, let  $k = \max_y D_1(y)$ . By a similar argument as above, we have that

$$\frac{\max_y D_1(y)}{\langle D_0^2, D_1 \rangle} \leq \frac{\max_y D_1(y)}{\langle D_1^2, D_1 \rangle - 2\langle \delta, D_1^2 \rangle + \langle \delta^2, D_1 \rangle} \leq \frac{k}{k^3 - \frac{(1-k)^3}{(N-1)^2} - 2\epsilon k^2 + \epsilon^2 k}$$

One can show that this term is maximized by setting  $k = \Theta(N^{-2/3})$ , and therefore this term is upper bounded by  $O(N^{4/3})$ . Indeed, taking the derivative of this quantity with respect to  $k$  and setting it equal to zero, one can see that the maximum value of  $k$  satisfies  $(-2(N-1)^2 + 2)k^3 + (2\epsilon - 3)k^2 + 1 = 0$ , which implies the maximum value satisfies  $k = \Theta(N^{-2/3})$ .

We've now shown that  $\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 100|S]} \leq 1 + o(1)$ . Now consider the opposite ratio. By the same reasoning as before, we have that

$$\frac{\Pr[\hat{b} = 100|S]}{\Pr[\hat{b} = 111|S]} = \frac{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (97)$$

$$\leq 1 + 2 \frac{\langle \delta, D_1^2 \rangle}{\sum_{y \in \{0,1\}^n} D_1(y)^3} + \frac{\langle \delta^2, D_1 \rangle}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (98)$$

$$\leq 1 + 2\epsilon \frac{\max_y D_1(y)}{\sum_{y \in \{0,1\}^n} D_1(y)^3} + \epsilon^2 \frac{\max_y D_1(y)^2}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (99)$$

$$\leq 1 + o(1) \quad (100)$$

Where on line 100 we used the fact that we previously upper bounded these terms when handling Case 1. Hence as  $n \rightarrow \infty$  the strings  $\hat{b} = 111$  and  $\hat{b} = 100$  become equiprobable.

**Case 3: 111 and 110** We have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 110|S]} = \frac{\sum_{y \in \{0,1\}^n} D_1(y)^3}{\sum_{y \in \{0,1\}^n} D_0(y) D_1(y)^2} \quad (101)$$

$$\leq 1 + \frac{\sum_{y \in \{0,1\}^n} \delta_y D_1(y)^2}{\sum_{y \in \{0,1\}^n} D_0(y) D_1(y)^2} \quad (102)$$

$$= 1 + \frac{\langle \delta_y, D_1^2 \rangle}{\langle D_0, D_1^2 \rangle} \quad (103)$$

$$\leq 1 + \frac{\epsilon \max_y D_1(y)^2}{\langle D_0, D_1^2 \rangle} \quad (104)$$

$$\leq 1 + \frac{\epsilon \max_y D_1(y)^2}{\langle D_1, D_1^2 \rangle - \langle \delta_y, D_1^2 \rangle} \quad (105)$$

$$\leq 1 + \frac{\epsilon \max_y D_1(y)^2}{\langle D_1, D_1^2 \rangle - \epsilon \max_y D_1(y)^2} \quad (106)$$

$$\leq 1 + o(1) \quad (107)$$

Where on line 105 we used the fact that  $D_0(y) \geq D_1(y) - \delta(y)$ , on line 106 we used that fact that the numerator is minimized if all the mass of  $\delta_y$  is placed on the maximum likelihood event of  $D_1$ , and on line 107 we used the fact that this is the same as the first term we bounded in Case 2.

Now consider the opposite ratio. We have that

$$\frac{\Pr[\hat{b} = 110|S]}{\Pr[\hat{b} = 111|S]} = \frac{\sum_{y \in \{0,1\}^n} D_0(y) D_1(y)^2}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (108)$$

$$\leq \frac{\sum_{y \in \{0,1\}^n} (D_1(y) + \delta(y)) D_1(y)^2}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (109)$$

$$= 1 + \frac{\sum_{y \in \{0,1\}^n} \delta(y) D_1(y)^2}{\sum_{y \in \{0,1\}^n} D_1(y)^3} \quad (110)$$

$$\leq 1 + o(1) \quad (111)$$

Where the last line follows from our previous arguments in Case 1. Hence we have that  $1 - o(1) \leq \frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 110|S]} \leq 1 + o(1)$  as desired.

Hence we have shown  $1 - o(1) \leq \frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = x|S]} \leq 1 + o(1)$  for any three-bit string  $x$ . Hence all strings are equiprobable, so in the case the distributions are  $\varepsilon$ -close, the algorithm's error probability tends to  $1/4$  as  $n \rightarrow \infty$ , and hence the algorithm is correct in this case.

To complete the proof, we now show that the probability of error is low then the distributions are  $1 - \varepsilon$  far apart in total variation distance.

Suppose the distributions are  $1 - \varepsilon$  far apart in total variation distance. By the definition of total variation distance, there must exist some event  $T \subseteq \{0, 1\}^n$  for which  $|D_0(T) - D_1(T)| \geq 1 - \varepsilon$ , where the notation  $D_0(T)$  indicates the probability that  $D_0$  outputs an element of the set  $T$ , i.e.  $D_0(T) = \sum_{y \in T} D_0(y)$ . Without loss of generality we have that  $D_0(T) - D_1(T) \geq 1 - \varepsilon$ , which implies  $D_1(\bar{T}) - D_0(\bar{T}) \geq 1 - \varepsilon$ . Since  $D_0$  and  $D_1$  are probability distributions, this implies  $D_0(T) \geq 1 - \varepsilon$  and  $D_1(T) \leq \varepsilon$ , and likewise  $D_1(\bar{T}) \geq 1 - \varepsilon$  and  $D_0(\bar{T}) \leq \varepsilon$ . In other words  $D_0$  has almost all its probability mass in  $T$  and  $D_1$  has almost all its probability mass in  $\bar{T}$ .

We'll now show that under these distributions, one will almost certainly see the output  $\hat{b} = 000$  or  $\hat{b} = 111$ . As before, we'll show this by proving that for large  $n$ , the strings  $\hat{b} = 000$  or  $\hat{b} = 111$  are far more likely than  $\hat{b} = 001$  or  $\hat{b} = 011$ , which implies the algorithm almost always outputs the correct answer.

Let  $k_0 = \max_{y \in T} D_0(y)$  and let  $k_1 = \max_{y \in \bar{T}} D_1(y)$ . Suppose without loss of generality that  $k_0 \geq k_1$  (otherwise exchange  $D_0$  and  $D_1$  in the argument). Then we have that

$$\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 100|S]} = \frac{\sum_{y \in \{0,1\}^n} D_0(y)^3}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} \quad (112)$$

$$= \frac{\langle D_0^2, D_0 \rangle}{\langle D_0^2, D_1 \rangle} \quad (113)$$

$$\geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon k_0^2 + \varepsilon^2 k_1} \quad (114)$$

$$\geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon k_0^2 + \varepsilon^2 k_0} \quad (115)$$

where line 112 follows from the same arguments as the previous section, and line 114 follows because the numerator is minimized by placing the uniform distribution on all elements other than the element

responsible for  $k_0$ , and the denominator is maximized if all the weight that  $D_0$  has on  $\bar{T}$  is placed on the element of maximal weight under  $D_1$ , and vice versa. Line 115 follows from the fact that  $k_0 \geq k_1$

Now we show that this quantity is  $\omega(1)$ , i.e. it approaches infinity as  $n \rightarrow \infty$ . Suppose by contradiction that there exists a constant  $c > 1$  which is an upper bound for this quantity. Since  $\varepsilon = o(N^{-2/3})$ , there exists an  $n_0$  such that for all  $n > n_0$ ,  $\varepsilon < \frac{1}{2c}N^{-2/3}$ . We claim that for all  $n > n_0$ , this quantity is greater than  $c$ , which is a contradiction.

To see this, we break into three cases.

**Case 1:**  $k_0 \geq N^{-2/3}$

In this case, the numerator is at least  $k_0^3$ , while the denominator is at most  $\frac{1}{10c}N^{-2/3}k_0^2 + \frac{1}{100c^2}N^{-4/3}k_0 \leq \frac{1}{2c}k_0^3 + \frac{1}{4c^2}k_0^3 < \frac{1}{c}k_0^3$ , where the last step follows from the fact that  $\frac{1}{2c} + \frac{1}{4c^2} < \frac{1}{c}$  for any  $c > 1$ . Therefore the quantity on line 115 is strictly greater than  $\frac{k_0^3}{\frac{1}{c}k_0^3} = c$  as desired.

**Case 2:**  $k_0 \leq N^{-2/3}$

In this case the numerator is at least  $\frac{(1-k_0)^3}{(N-1)^2}$  which is  $\geq 0.75N^{-2}$  for sufficiently large  $n$ , while the denominator is  $\varepsilon k_0^2 + \varepsilon^2 k_0 \leq \frac{1}{2c}N^{-2} + \frac{1}{4c^2}N^{-2} < \frac{3}{4c}N^{-2}$  for  $c > 1$ , which follows from our upper bounds on  $\varepsilon$  and  $k_0$ . Hence the quantity on line 115 is strictly greater than  $c$  as desired.

Therefore we have shown that and  $n \rightarrow \infty$ , the string  $\hat{b} = 000$  (or  $\hat{b} = 111$ , if  $k_0 < k_1$ ) is much more likely to occur than  $\hat{b} = 001$ .

A similar proof holds to show that the string  $\hat{b} = 000$  (or  $\hat{b} = 111$ ) is more likely to occur than  $\hat{b} = 110$ ; indeed by the same arguments as above, assuming  $k_0 \geq k_1$ , we have

$$\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 110|S]} \geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon^2 k_0 + \varepsilon k_1} \geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon^2 k_0 + \varepsilon k_0} \geq \omega(1)$$

Hence the string  $\hat{b} = 000$  is far more likely to occur than the strings  $\hat{b} = 001$  or  $\hat{b} = 011$  (assuming  $k_0 > k_1$ , otherwise the string  $\hat{b} = 111$  is more likely to occur than 001 or 011), and hence the algorithm errs with probability  $o(1)$  when the distributions are  $1 - \varepsilon$  far apart. This completes the proof.