Placing Conditional Disclosure of Secrets in the Communication Complexity Universe*

Benny Applebaum† Prashant Nalini Vasudevan‡

November 26, 2018

Abstract

In the *conditional disclosure of secrets* (CDS) problem (Gertner et al., J. Comput. Syst. Sci., 2000) Alice and Bob, who hold n-bit inputs x and y respectively, wish to release a common secret z to Carol (who knows both x and y) if and only if the input (x, y) satisfies some predefined predicate f. Alice and Bob are allowed to send a single message to Carol which may depend on their inputs and some shared randomness, and the goal is to minimize the communication complexity while providing information-theoretic security.

Despite the growing interest in this model, very few lower-bounds are known. In this paper, we relate the CDS complexity of a predicate f to its communication complexity under various communication games. For several basic predicates our results yield tight, or almost tight, lower-bounds of $\Omega(n)$ or $\Omega(n^{1-\epsilon})$, providing an exponential improvement over previous logarithmic lower-bounds.

We also define new communication complexity classes that correspond to different variants of the CDS model and study the relations between them and their complements. Notably, we show that allowing for imperfect correctness can significantly reduce communication – a seemingly new phenomenon in the context of information-theoretic cryptography. Finally, our results show that proving explicit super-logarithmic lower-bounds for imperfect CDS protocols is a necessary step towards proving explicit lower-bounds against the class AM, or even AM $\cap$ co-AM – a well known open problem in the theory of communication complexity. Thus imperfect CDS forms a new minimal class which is placed just beyond the boundaries of the "civilized" part of the communication complexity world for which explicit lower-bounds are known.

*An extended abstract of this paper appears in the proceedings of the 10th Innovations in Theoretical Computer Science (ITCS) 2019.

†Tel Aviv University, Tel Aviv, Israel. Email: bennyap@post.tau.ac.il. Supported by the European Union's Horizon 2020 Programme (ERC-StG-2014-2020) under grant agreement no. 639813 ERC-CLC, by an ICRC grant and by the Check Point Institute for Information Security.

‡UC Berkeley, Berkeley, USA. Email: prashvas@berkeley.edu. This work was done in part while the author was visiting Tel Aviv University. Supported in part by NSF Grant CNS-1350619, and by the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Research Office under contracts W911NF-15-C-0226 and W911NF-15-C-0236.

1 Introduction

Understanding the communication complexity of information-theoretically secure protocols is a fundamental research problem. Despite much effort, we have very little understanding of the communication complexity of even simple cryptographic tasks, and for most models, there are exponentially large gaps between the best known upper-bounds and the best known lower-bounds. In an attempt to simplify the problem, one may try to focus on the most basic settings with a minimal non-trivial number of players (say two or three) and the simplest possible communication pattern (e.g., single message protocols). Different cryptographic tasks have been studied in this minimal setting, including secure computation [FKN94], and non-interactive zero-knowledge proofs [GPW15]. In this paper we will focus on what seems to be the simplest task in this model: *Conditional Disclosure of Secrets* (CDS) [GIKM00].[1]

Conditional Disclosure of Secrets. Consider a pair of computationally unbounded parties, Alice and Bob, each holding an input, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively, to some public predicate $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Alice and Bob also hold a joint secret z (say a single bit) and have access to a joint source of randomness $r \xleftarrow{R} \mathcal{R}$. The parties wish to disclose the secret z to a third party, Carol, if and only if the predicate $f(x,y)$ evaluates to 1. To this end, Alice and Bob should each send a single message $a = a(x,z;r)$ and $b = b(y,z;r)$ to Carol. Based on the transcript (a,b) and the inputs (x,y), Carol should be able to recover the secret z if and only if $f(x,y) = 1$. (Note that Carol is assumed to know x and y.) That is, we require two properties:

- *Correctness*: There exists a deterministic decoder algorithm Dec that recovers z from (x,y,a,b) with high probability whenever x,y is a 1-input (i.e., $f(x,y) = 1$);
- *Privacy*: For every fixed 0-input (x,y) (for which the predicate evaluates to 0), regardless of the value of the secret z, the joint distribution of the transcript (a,b), induced by a choice of the shared randomness, is statistically close (up to some small deviation error) to some canonical distribution $\mathsf{Sim}(x,y)$.

The main complexity measure of CDS protocols is their communication complexity which is taken to be the total bit-length of the messages a and b. (See Figure 1 for a schematic view and Section 4 for formal definitions.)

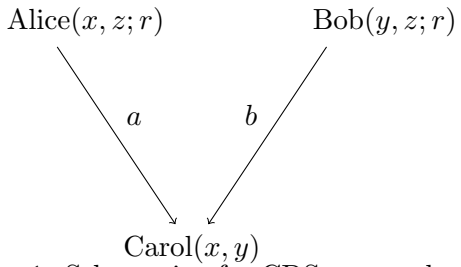

Figure 1: Schematic of a CDS protocol.

[1] While we do not wish to define the notions from [FKN94] and [GPW15], let us just mention that the complexity of a function in these two models upper-bounds the complexity in the CDS model [GIKM00, AR16]. In this sense, CDS may be considered as being simpler.

Apart from being a natural basic notion, CDS has turned out to be a useful primitive with various applications in the context of private information retrieval (PIR) [GIKM00], secure multiparty computation [AIR01, IKP10], secret sharing schemes [BD91, CSGV93, SS97, LVW17b, BKN18, AA18, LV18], and attribute-based encryption [Att14, Wee14]. Correspondingly, the communication complexity of CDS was extensively studied in the last few years.

Upper bounds. On the positive side, it is known that the CDS complexity of a predicate f is at most linear in the formula complexity of f [GIKM00]. This result was extended to other (presumably stronger) computational models such as (arithmetic) branching programs [IW14], and (arithmetic) span programs [AR16]. The latter paper also shows that the CDS complexity of f is at most linear in the complexity of any zero-information Arthur Merlin (ZAM) protocol for f. (The ZAM model, introduced by [GPW15], adds a zero-knowledge property to the standard AM communication complexity model.)[2] In a recent breakthrough, Liu, Vaikuntanathan and Wee [LVW17a] showed that the CDS complexity of any predicate $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ over n-bit inputs is at most $2^{\widetilde{O}(\sqrt{n})}$, improving over the exponential upper-bound of $O(2^{n/2})$ from [BIKK14]. Applebaum et al. [AARV17] showed that when the secret is very long (exponential in the size of the domain of the predicate) the overhead per each bit of z can be reduced to $O(n)$; a constant-rate solution (in which the total communication is $O(|z|)$) was recently given in [AA18].

The quest for lower bounds. On the lower-bound front much less is known. While we have tight lower bounds for restricted forms of CDS (e.g., when the computations are restricted to linear functions [GKW15, BFMP17, BP18]), only few, relatively weak, lower-bounds are known for general CDS. It is important to note that an insecure solution to the problem has a communication cost of 1 bit! (Let Alice send the secret in the clear regardless of her input.) Hence, any super-constant lower-bound is, in a sense, non-trivial. Indeed, unlike the case of standard communication games for which communication lower-bounds are based on the correctness properties of the protocol, the challenge here is to somehow capture the additional cost of *privacy*.

The first super-constant lower-bound was proved by Gay et al [GKW15].

Theorem 1 ([GKW15])**.** *For every predicate $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$\mathrm{CDS}(f) \geq \Omega(\log(\mathrm{R}_{A\to B}(f) + \mathrm{R}_{B\to A}(f))),$$

where $\mathrm{R}_{A\to B}(f)$ denotes the one-way randomized communication complexity of f, and $\mathrm{CDS}(f)$ denotes the minimal communication complexity of a CDS protocol for f with privacy and correctness error of 0.1.[3]

For n-bits predicates, Theorem 1 leads, at best, to a logarithmic lower-bound of $\Omega(\log n)$. Applebaum et al [AARV17] showed that this bound is essentially tight: There are (partial) functions whose randomized communication complexity is exponentially larger than their CDS complexity. They also proved a linear n-bit lower-bound for a random (non-explicit) n-bit predicate $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. An explicit version of this result was proved by [AHMS18].

[2] The theorem of [AR16] actually relates the communication and randomness complexity of CDS for f to the randomness and communication complexity of a ZAM protocol for the complement of f. However, using our results in this paper, specifically Lemma 1, one can conclude that the CDS communication of f is at most linear in the ZAM communication of f.

[3] The theorem was originally proved for perfect CDS, however, the proof generalizes to the imperfect case (see [AARV17]).

Theorem 2 ([AHMS18]). *For every non-degenerate predicate*[4] $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ *whose largest 0-monochromatic rectangle is of size at most* L*,*

$$\mathrm{pCDS}(f) \geq \log \frac{|f^{-1}(0)|}{L} - \log \frac{|\mathcal{X} \times \mathcal{Y}|}{|f^{-1}(0)|} - 1 = 2 \log |f^{-1}(0)| - \log |\mathcal{X}| - \log |\mathcal{Y}| - \log L - 1,$$

where $\mathrm{pCDS}(f)$ *denotes the minimal communication complexity of a CDS protocol for* f *with perfect privacy and perfect correctness.*

The theorem is effective for predicates whose communication matrix is rich in zeroes, and at the same time avoids large zero-monochromatic rectangles. In particular, for mod-2 inner product over n-bit inputs, we get a tight lower-bound of $n - O(1)$ and for Set-Intersection a lower-bound of $\Omega(n)$. Unfortunately, the theorem is not robust to errors, leaving the imperfect CDS complexity of these predicates wide open. Moreover, for many basic predicates the theorem does not even give logarithmic bounds either due to the lack of many zeroes (e.g., the Not-Equal predicate) or due to the existence of huge zero-rectangles (e.g., the Greater-Than predicate).

This paper. Theorems 1 and 2 provide a very partial picture, and fall short of proving meaningful and robust lower-bounds for many basic predicates, such as Not-equal, Greater-Than, Intersection, and Index.[5] We believe that a full understanding of these simple cases is necessary for the more ambitious goal of proving stronger lower bounds. Our goal in this paper is to remedy the situation by providing new lower-bound techniques. Specifically, we enrich our lower-bound toolbox by relating the CDS complexity of a function to its communication complexity under various communication games. Our results provide simple, yet effective, ways to leverage privacy to construct communication protocols. They lead to new lower-bounds for perfect and imperfect CDS protocols, and allow us to establish new results regarding the relations between different variants of the CDS model.

2 Our Contribution

2.1 Perfectly-correct CDS and coNP Games

Our first theorem relates the complexity of any perfectly-correct CDS protocol for f to the non-deterministic communication complexity of f's complement.

Theorem 3. *For every predicate* $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$*,*

$$\mathrm{pcCDS}(f) \geq \Omega(\mathrm{coNP}(f)) - O(\log(n)),$$

where n *denotes the total input length of* f*, and* $\mathrm{pcCDS}(f)$ *denotes the minimal communication complexity of a CDS protocol for* f *with perfect correctness and privacy error of* 0.1*.*

Proof idea. To prove the theorem, we first show that the coNP complexity is upper-bounded by the randomness complexity of the CDS, and then prove that one can always assume that the randomness complexity is comparable to the communication complexity via a new sparsification lemma (similar to that of Newman [New91]; see Section 4.2). The first part relies on the following

[4] A predicate is non-degenerate if for every fixing of $x \in \mathcal{X}$ the residual function $f(x, \cdot)$ is not the constant zero function.

[5] Apart of being basic examples, these predicates are motivated by some of the applications of CDS.

simple observation: In order to convince Alice and Bob that $f(x,y)$ evaluates to zero it suffices to prove that the joint distribution of the CDS messages for zero-secret, $(a(x,z=0;r),b(y,z=0;r))$, induced by a random choice of r, and the joint distribution of the messages for one-secret $(a(x,z=1;r),b(y,z=1;r))$, are *not disjoint*. A prover can prove this statement by sending to Alice and Bob a pair of strings r_0 and r_1 for which $(a(x,z=0;r_0),b(y,z=0;r_0))$ equals to $(a(x,z=1;r_1),b(y,z=1;r_1))$. (See Section 5.)

Despite its simplicity, this theorem is quite powerful. In particular, ignoring the constants in the Omega-notation and the logarithmic loss, the bound provided by Theorem 3 subsumes the lower-bound of Theorem 2 from [AHMS18]. Indeed, the latter lower-bound is at most the logarithm of the ratio between the zero-mass of f and its largest zero-monochromatic rectangle – a quantity that cannot be larger than the non-deterministic communication complexity of the complement of f (i.e., coNP(f)). Moreover, our new theorem can be applied to predicates that have only few zero entries or to predicates with huge zero-rectangles, for which Theorem 2 becomes meaningless. For example, by plugging-in classical coNP lower-bounds, we settle the complexity of the not-equal predicate with respect to perfectly correct CDS protocols.

Corollary 1. *Let* $\mathsf{NEQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *denote the not-equal predicate which evaluates to* 1 *if and only if* $x \neq y$*. Then,*

$$\text{pcCDS}(\mathsf{NEQ}_n) \geq \Omega(n).$$

Similar tight linear lower-bounds can be obtained for the pcCDS complexity of the Greater-Than predicate, the Set-Intersection predicate, and the Inner-Product predicate. Previously, we had no super-logarithmic lower bounds that tolerate privacy error. (As already mentioned, for Greater-Than and NEQ_n, we did not have such bounds even for perfect CDS protocols.)

pcCDS **is not closed under complement.** Interestingly, the *equality* function EQ_n has a very succinct perfect CDS protocol: Use the shared randomness to sample a pair-wise independent hash function $h : \{0,1\}^n \to \{0,1\}$, and let Alice output $h(x)$ and Bob output $h(y) \oplus z$. The protocol has a minimal communication complexity of 2 and randomness complexity of $O(n)$. (The latter can be reduced to $O(\log n)$ by using an almost pair-wise independent hash function and settling for a constant privacy error.) This yields a strong separation between the complexity of a predicate and its complement with respect to perfectly-correct perfectly-private CDS protocols (pCDS).

Corollary 2. $\text{pCDS}(\mathsf{EQ}_n) = 2$ *whereas* $\text{pCDS}(\mathsf{NEQ}_n) \geq \text{pcCDS}(\mathsf{NEQ}_n) \geq \Omega(n)$*. In particular, the classes* pCDS *and* pcCDS *are not closed under complement.*[6]

Transformations from CDS protocols for f to its complement were studied in [AARV17]. The resulting protocols either introduce a privacy error or suffer from a communication overhead that grows polynomially with the randomness complexity of the original protocol. The NEQ_n example shows that at least one of these losses is inherent.

The benefit of decoding errors. The results of [AARV17] (together with our randomness sparsification lemma) show that imperfect CDS is closed under complement. This general result leads to a polylogarithmic CDS protocol for NEQ_n with imperfect privacy and imperfect correctness,

[6] We follow the standard communication complexity terminology and write pCDS to denote the class of predicates that admit a pCDS protocol whose complexity is polylogarithmic in the input length. A similar convention will be used throughout the paper for all other variants of the CDS model.

providing a surprising separation between general imperfect CDS protocols and ones which have perfect correctness. In fact, it is not hard to directly design a CDS protocol for NEQ_n with *constant communication*, *perfect privacy*, and constant correctness error. (See Appendix A for a more general statement.) This leads to the following stronger separation.

Corollary 3. *There is an n-bit predicate f for which* $\mathrm{pcCDS}(f) = \Omega(n)$ *and* $\mathrm{ppCDS}(f) = O(1)$*, where* $\mathrm{ppCDS}(f)$ *denotes the minimal communication complexity of a CDS protocol for f with perfect privacy and correctness error of* 0.1*. In particular,*

$$\mathrm{ppCDS} \not\subseteq \mathrm{pcCDS}.$$

As pointed to us by Hoteck Wee, Corollary 3 provides a rare example for an information-theoretic secure protocol that can significantly benefit from a small correctness error. This phenomena seems new in the context of information-theoretic secure cryptography, and is worth further exploration.[7]

2.2 Perfectly-Private CDS and PP Games

Our next goal is to lower-bound the complexity of CDS protocols with correctness errors. We begin with the case of perfectly private protocols.

Theorem 4. *For every predicate $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$\mathrm{ppCDS}(f) \geq \Omega(\mathrm{PP}(f)) - O(\log(n)),$$

where n denotes the total input length of f, and $\mathrm{ppCDS}(f)$ *denotes the minimal communication complexity of a CDS protocol for f with perfect privacy and correctness error of* 0.1*.*

The complexity measure $\mathrm{PP}(f)$ essentially corresponds to the sum of the communication complexity and number of private random bits used by a communication protocol that computes f correctly with probability more than $1/2$, where shared randomness is not allowed. (See Definition 3 for a formal definition.) The discrepancy method implies that the PP complexity of the mod-2 inner-product predicate IP_n is $\Omega(n)$ (cf. [KN97]) and so we get the following.

Corollary 4. *Let $\mathsf{IP}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ denote the inner-product predicate on n-bit inputs. Then,*

$$\mathrm{ppCDS}(\mathsf{IP}_n) \geq \Omega(n).$$

This is the first linear lower-bound on CDS with imperfect correctness. (Previous arguments fail to achieve such a result even for a non-explicit predicate.)

Proof idea. In order to prove Theorem 4, we turn a ppCDS protocol into a PP protocol. Loosely speaking, the idea is to construct a randomized protocol that accepts the input (x, y) based on collisions between random CDS transcripts that correspond to a zero-secret and random CDS transcripts that correspond to a one-secret. This idea, which was employed in the query setting by [BCH+17], leads to the desired result. (See Section 6.)

[7] Compare this, for example, to Shannon's classical lower-bound for perfectly-secure one-time symmetric encryption [Sha49] in which a constant decryption error has a minor effect on the key/ciphertext length [Dod12].

2.3 Imperfect CDS, Interactive Proofs, and Zero Knowledge

We move on to the most general case of imperfect CDS protocols with both constant privacy error and correctness error. We show that the complexity of such protocols is at least polynomial in the AM communication complexity of f. (The latter class is the communication complexity analogue of Arthur-Merlin proofs; see Definition 5.)

Theorem 5. *There exists some universal constant $c > 0$, such that for any Boolean function f it holds that*

$$\mathrm{CDS}(f) \geq \mathrm{AM}(f)^c - \mathrm{polylog}(n),$$

where n denotes the total input length of f, and $\mathrm{CDS}(f)$ denotes the minimal communication complexity of a CDS protocol for f with correctness and privacy errors of 0.1.

Since (imperfect) CDS is closed under complement (by [AARV17, Theorem 2] and Lemma 1), it holds that $\mathrm{CDS}(\bar{f}) \leq \mathsf{poly}(\mathrm{CDS}(f))$, and so we conclude the following.

Corollary 5. *There exists some universal constant $c > 0$, such that for any Boolean function f it holds that*

$$\mathrm{CDS}(f) \geq \max(\mathrm{AM}(f), \mathrm{coAM}(f))^c - \mathrm{polylog}(n),$$

where n denotes the total input length of f.

Explicit CDS lower-bounds? Corollary 5 can be used to show that the CDS complexity of most n-bit predicates must be at least polynomial in n, even when the protocol is imperfect. Unfortunately, it falls short of providing explicit lower-bounds; Finding an explicit function outside $\mathrm{AM} \cap \mathrm{coAM}$ is a central open problem in the theory of communication complexity. In fact, $\mathrm{AM} \cap \mathrm{coAM}$ forms a minimal class for which no explicit lower-bounds are known [GPW18]. Corollary 5 places CDS as a weaker (and perhaps more accessible) target for explicit lower-bounds.

Proof idea. To prove Theorem 5 we show that a CDS protocol can be transformed into a constant-round private-coins interactive-proof. Then, we note that, just like in the computational setting, such interactive proofs can be converted to an AM protocol with polynomial overhead [Bab85, GS89].[8] The first step is obtained by imitating the standard interactive proof of Graph Nonisomorphism [GMW91]. Indeed, the AM protocol constructed in Theorem 5 turns out to satisfy a *statistical zero-knowledge* property; That is, the view of Alice and Bob can be simulated via a low complexity 2-party randomized protocol. (See Section 8 for details.)

CDS **vs.** SZK. Recall that, by definition, a CDS protocol yields a (distributed mapping) from the input (x, y) and the secret z to a distribution D_z over the transcript (a, b) such that the distributions, D_0 and D_1, are either statistically-close or statistically-far depending on the value of $f(x, y)$. This resembles the Statistical Difference problem [SV03], which is known to be complete for the computational complexity class SZK (consisting of problems that have interactive proofs that are statistically zero-knowledge). One may therefore hope to prove that in the communication complexity setting CDS complexity is characterized by SZK complexity. As already mentioned, Theorem 5

[8] This reduction has a polynomial dependency in the randomness. In order to avoid such an overhead in the final statement, we prove a randomness sparsification lemma (Lemma 9) for constant-round interactive protocols. This requires some care due to the use of private coins.

actually shows that CDS $\subseteq$ SZK, however, we do not know whether the reverse direction holds. Roughly speaking, such a result faces two obstacles. Firstly, the completeness result from [SV03] has an overhead that depends on the randomness complexity of the protocol, and we do not know how to get rid of this dependency. (In particular, it is not clear how to prove a proper sparsification lemma for SZK without sacrificing the zero-knowledge property.) Secondly, even if the randomness complexity is small, we do not know how to obtain a CDS protocol without allowing some interaction between Alice and Bob. Indeed, in Section 9, we show that SZK$'$ $\subseteq$ CDS$'$ where the "prime" version of SZK charges randomness towards the total complexity and the "prime" version of CDS allows short interaction between Alice and Bob. The problem of proving that SZK $\subseteq$ CDS (and therefore SZK = CDS) remains as an interesting open problem.

The results described so far are summarised in Fig. 2, which shows the relationship between perfect and imperfect CDS and various measures from communication complexity. In Table 1, we list the current state of knowledge of the various CDS complexities of a number of commonly studied predicates. (See Section 3.)

2.4 Asymmetry in CDS and One-Way Communication

We shift gears, and turn to study the communication tradeoffs between Alice's and Bob's messages. Suppose that Alice's message is restricted to a short string of length t_A. Can we prove that Bob's message must be very long? We prove such tradeoffs based on the one-way randomized communication complexity of f.

Theorem 6. *In any perfectly correct* 0.1*-private* CDS *protocol for f in which Alice and Bob communicate t_A and t_B bits respectively and the total input length of the function is n, it holds that:*

$$2^{t_A}(t_A + t_B + \log n) \geq \Omega(\mathrm{R}_{B\to A}(f)).$$

(In fact, the result holds even if one considers one-way randomized protocols that err only over zero inputs. See Section 7.) Recall that Theorem 1 (which is from [GKW15]) shows that the total communication complexity $t_A + t_B$ is at least logarithmic in $(\mathrm{R}_{A\to B}(f) + \mathrm{R}_{B\to A}(f))$, which is tight for some predicates [AARV17]. Theorem 6 provides a more accurate picture. If the total communication complexity is dominated by t_A, then one gets a logarithmic bound, similar to Theorem 1; however, when t_A is small (e.g., constant), we get a strong linear lower-bound of

$$t_B = \Omega(\mathrm{R}_{B\to A}(f)) - O(\log n).$$

In fact, when $\mathrm{R}_{B\to A}(f) = \Omega(n)$, for any constant $\alpha < 1$ if $t_A \leq \alpha \log n$ then

$$t_B = \Omega(n^{1-\alpha}).$$

Concretely, consider the Index_n predicate in which Bob holds an n-bit database $x \in \{0,1\}^n$ and Alice holds an index $i \in [n]$ (encoded as a string of length $\log n$) and the output is the i-th bit of x. Since $\mathrm{R}_{B\to A}(\mathsf{Index}_n) = \Omega(n)$ [KNR99] we get:

Corollary 6. *In any perfectly correct* 0.1*-private* CDS *protocol for* Index_n *in which Alice communicates at most $\alpha \log n + O(1)$ bits for some constant $0 \leq \alpha < 1$, the database owner, Bob, must communicate at least $\Omega(n^{1-\alpha})$ bits.*

Similar results can be obtained for predicates like Greater-Than, Set-Disjointness and Set-Intersection, based on classical lower-bounds for randomized one-way communication complexity (cf. [MNSW98, KNR99]).

The Index_n predicate plays an important role in CDS constructions and applications. First, it is complete for CDS in the sense that any n-bit predicate can be reduced to Index_N for $N = 2^n$. Indeed, the best known general CDS protocols were obtained by improving the pCDS complexity of Index [LVW17a]. In addition, a CDS for the index function can be viewed as a one-time version of the well-studied notion of *Broadcast Encryption*, and the lower-bound of Corollary 6 becomes especially appealing under this framework. Details follow.

Broadcast Encryption [FN93]. Suppose that we have a single sender and n receivers. The sender has a private encryption key r and each receiver $i \in [n]$ has its own private decryption key k_i. All the keys were collectively generated and distributed in an offline phase. In an online phase, the sender gets a message z together with a public list of authorized users $y \subseteq [n]$, represented by an n-bit characteristic vector. The sender should broadcast a ciphertext $b = b(y, z; r)$ to all the receivers (who also know y) so that an *authorized* receiver will be able to decrypt the ciphertext, and an unauthorized (computationally unbounded) receiver will learn nothing about the message z. The goal is to minimize the length of the ciphertext b, and the length of the keys k_i.

Information-theoretic one-time secure Broadcast Encryption turns to be equivalent to the CDS problem with respect to the Index_n predicate: Identify the ciphertext with Bob's message $b = b(y, z; r)$ and the i-th key with Alice's message $a(i; r)$.[9] The problem can be solved with n-bit ciphertext and 1-bit keys, and with 1-bit ciphertext and n-bit keys. In fact, [GKW15] showed that one can smoothly get any tradeoff as long as the product of the ciphertext length and the key length is n. Corollary 6 shows that when the key-length is sub-logarithmic the ciphertext must be almost linear, confirming a conjecture of Wee [Wee18].

Proof idea (of Theorem 6). The idea is to let Bob send to Alice a pair of random strings r_0 and r_1 that are mapped to the same Bob's message b under the zero-secret and under the one-secret respectively. Alice then uses the string r_z and the secret z to compute a corresponding message a_z, and accepts if the zero message a_0 equals to the one message a_1. Perfect correctness guarantees that Alice will never err on 0-inputs. We further show that, when $f(x, y) = 1$, Alice accepts with probability which is at least inverse-exponential in her message length (up to a loss that is proportional to the privacy error of the protocol). See Section 7 for details.

3 Conclusion and Open Questions

In this paper we studied the relations between CDS protocols and standard communication complexity games. We established new connections between CDS communication complexity (with perfect and imperfect privacy and correctness) to well-known communication complexity measures for non-deterministic protocols, randomized unbounded-error protocols, and one-way protocols. This leads to new CDS bounds for various simple functions. These results are summarized in Fig. 2 and Table 1.

We end by listing the immediate interesting questions left open following our work.

[9] Here we assume that we have a CDS in which only Bob holds the secret. However, any CDS can be transformed into this form with an additional communication cost of $O(|z|) = O(1)$.

1. Prove an explicit polynomial lower-bound on (imperfect) CDS complexity. (A natural candidate would be Inner-Product.)

2. Our current ppCDS lower-bounds are based on PP complexity, which corresponds to discrepancy. Can we derive such bounds on weaker, easier-to-establish, properties? In particular, can we prove non-trivial ppCDS lower-bounds for predicates that have low randomized bounded-error communication complexity like Greater-Than?

3. Unlike all the other communication complexity measures considered here, CDS complexity is not necessarily upper-bounded by the length of the inputs. But we have no super-linear (or even linear with a large constant factor) lower-bounds for even perfect CDS protocols. Can any of the existing lower-bound techniques from communication complexity be used to obtain such bounds?

4. If not, can this difficulty be explained, perhaps by relating the problem of proving such lower bounds for CDS to more well-studied problems that are still unsolved?

5. Following the paradigm of lifting query complexity lower bounds to the communication setting, is there a natural query complexity measure that can be lifted to CDS complexity?

6. One simple predicate that has eluded all our bounds is Set-Disjointness, for which the best (imperfect) CDS protocol we know has $O(n)$ complexity, and the best lower bound we can prove, even for perfect CDS, is $\Omega(\log(n))$. Are either of these tight?

Predicate	pCDS	pcCDS	ppCDS	CDS
Equality	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$
Non-Equality	$\Theta(n)$	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$
Inner-Product	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$	$O(n)$ & $\Omega(\log n)$
Greater-Than	$\Theta(n)$	$\Theta(n)$	$O(n)$ & $\Omega(\log n)$	$O(n)$ & $\Omega(\log n)$
Set-Intersection	$\Theta(n)$	$\Theta(n)$	$O(n)$ & $\Omega(\log n)$	$O(n)$ & $\Omega(\log n)$
Set-Disjointness	$O(n)$ & $\Omega(\log n)$	$O(n)$ & $\Omega(\log n)$	$O(n)$ & $\Omega(\log n)$	$O(n)$ & $\Omega(\log n)$

Table 1: The CDS complexity of some simple functions. By definition, an upper-bound in the leftmost column (pCDS) implies an upper-bound in all other columns, and a lower-bound in the rightmost column (CDS) implies a lower-bound in all other columns. All the linear upper-bounds for pCDS follow from the fact that all of these predicates can be computed by a linear-size formula. The logarithmic lower-bounds for CDS follow from Theorem 1 (and the fact that the corresponding predicates have linear randomized one-way communication complexity.) The linear lower-bounds for pcCDS and ppCDS follow from Theorems 3 and 4 respectively.

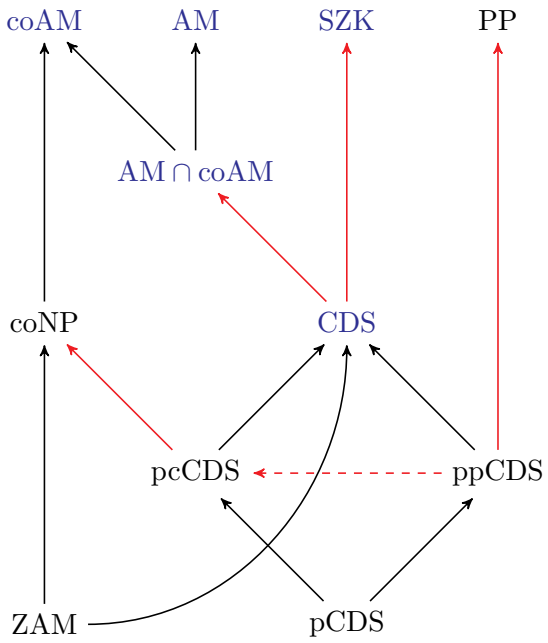

Figure 2: As is standard, we use the name of a complexity measure to also denote the class of functions with polylog(n) complexity under the measure. For classes C_1 and C_2, a solid arrow $C_1 \to C_2$ indicates that $C_1 \subseteq C_2$, and a dashed arrow $C_1 \dashrightarrow C_2$ indicates that $C_1 \not\subseteq C_2$. Red arrows indicate new results from this paper. Blue text indicates classes for which explicit bounds are not known.

4 Formal Setup and a Sparsification Lemma

For a finite set A we write $a \xleftarrow{R} A$ to denote a random variable which is sampled uniformly from A. The *statistical distance* between two discrete random variables, X and Y, denoted by $\Delta(X;Y)$ is defined by $\Delta(X;Y) := \frac{1}{2}\sum_z |\Pr[X=z] - \Pr[Y=z]|$. We will also use statistical distance for probability distributions, where for a probability distribution D the value $\Pr[D=z]$ is defined to be $D(z)$.

4.1 CDS Definitions

Definition 1 (CDS)**.** Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Let $\mathsf{F}_A : \mathcal{X} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}_A$ and $\mathsf{F}_B : \mathcal{Y} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}_B$ be deterministic encoding algorithms, where $\mathcal{Z}$ is the *secret domain*. Then, the pair $(\mathsf{F}_A, \mathsf{F}_B)$ is a CDS scheme for f with *correctness error* c and *privacy error* s if the function $\mathsf{F}(x,y,z,r) = (\mathsf{F}_A(x,z,r), \mathsf{F}_B(y,z,r))$ that corresponds to the joint computation of F_A and F_B on a common z and r, satisfies the following properties:

1. (c-Correctness) There exists a deterministic algorithm Dec, called a *decoder*, such that for every 1-input (x,y) of f and any secret $z \in \mathcal{Z}$ we have that:
$$\Pr_{r \xleftarrow{R} \mathcal{R}}[\mathsf{Dec}(x,y,\mathsf{F}(x,y,z,r)) \neq z] \leq c$$
2. (s-Privacy) There exists a randomized simulator Sim such that for every 0-input (x,y) of f, every secret $z \in \mathcal{Z}$, and uniformly chosen randomness $r \xleftarrow{R} \mathcal{R}$ the following holds:
$$\Delta\left(\mathsf{Sim}(x,y)\ ;\ \mathsf{F}(x,y,z,r)\right) \leq s.$$

The *communication complexity* of the CDS protocol is $(\log|\mathcal{T}_A| + \log|\mathcal{T}_B|)$ and its *randomness complexity* is $\log|\mathcal{R}|$. If c and s are zeros, such a CDS scheme is called *perfect*.

Unless stated otherwise, we restrict our attention to the case of single-bit secrets, i.e., $\mathcal{Z} = \{0,1\}$. By default, we let $\mathcal{X} = \{0,1\}^{n_A}$, $\mathcal{Y} = \{0,1\}^{n_B}$, $\mathcal{R} = \{0,1\}^{\rho}$, $\mathcal{T}_A = \{0,1\}^{t_A}$, and $\mathcal{T}_B = \{0,1\}^{t_B}$ for positive integers n_A, n_B, s, ρ, t_A, and t_B. Further, we denote $\mathcal{T}_A \times \mathcal{T}_B$ by $\mathcal{T}$; that is, the joint function $\mathsf{F} = (\mathsf{F}_A, \mathsf{F}_B)$ is of the form $\mathsf{F} : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}$, and by default $\mathcal{T} = \{0,1\}^t$, where $t = (t_A + t_B)$.

We denote by $\mathrm{CDS}(f)$ the least communication complexity of any CDS protocol for f with correctness and privacy errors at most 0.1 (with single-bit secrets). We further denote by pCDS, pcCDS, and ppCDS the same for perfect, perfectly correct, and perfectly private CDS protocols (where the unspecified error is at most 0.1). Our choice of the constant 0.1 is arbitrary up to the requirement that $(1-c)^2 > s$. Sometimes, we also denote c-correct s-private CDS by $(c,s)-\mathrm{CDS}$.

All of our discussion, while stated here for total functions for simplicity, extends in a straightforward manner to partial functions. We refer the reader to the book by Kushilevitz and Nisan [KN97] for the definitions of standard communication complexity measures that we use but do not define.

Remark 4.1 (CDS encoding). A function $\mathsf{F} : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}$ that satisfies *c-correctness* and *s-privacy* properties defined above, is referred to as a *CDS encoding*. Using this terminology, our notion of CDS scheme corresponds to a CDS encoding that can be decomposed into an x-part and a y-part. This view allows us to separate the semantics of CDS from its syntax.

4.2 Randomness Sparsification for CDS

Lemma 1. *Suppose there is a* CDS *protocol for f with communication t and randomness ρ, and with correctness error c and privacy error s. Then, for any $\epsilon > 0$, there is a $(c+\epsilon, s+\epsilon)$-*CDS *protocol for f with communication t and randomness*

$$2t + \log t + \log n + 2\log(1/\epsilon),$$

where $n = \log|\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}|$ is the total bit-length of the input and the secret. Further, if the former protocol is perfectly correct, then so is the latter.[10]

This lemma is similar to the classic theorem of Newman for communication complexity [New91]. It can also be proved in the same manner as that theorem, by sub-sampling random strings (see, for instance, [IK04, Lemma 1]). We describe the proof using the terminology of non-Boolean PRGs [DI06]. This proof leads to efficient sparsification in some restricted cases (e.g., when the CDS protocol is efficiently computable), and may therefore find future applications.

Definition 2 (non-Boolean PRG)**.** Let $\mathcal{D} = \{D\}$ be a class of functions from $\mathcal{R}$ to $\mathcal{T}$ and let $\epsilon > 0$ be an error parameter. We say that $G : \mathcal{L} \to \mathcal{R}$ is a *non-Boolean PRG* with error ϵ against $\mathcal{D}$ (in short $(\mathcal{D}, \epsilon)$-nbPRG) if for every $D \in \mathcal{D}$, G ϵ-fools D, i.e., the statistical distance between $D(G(U_{\mathcal{L}}))$ and $D(U_{\mathcal{R}})$ is at most ϵ, where $U_{\mathcal{X}}$ stands for the uniform distribution over the finite set $\mathcal{X}$.

Claim 1. *Let $\mathsf{F}_A : \mathcal{X} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}$ and $\mathsf{F}_B : \mathcal{Y} \times \mathcal{Z} \times \mathcal{R} \to \mathcal{T}$ be a* CDS *protocol with correctness error c and privacy error s for a function f. Let $\mathcal{D}$ be the class of functions from $\mathcal{R}$ to $\mathcal{T}$ obtained by restricting the joint* CDS *computation $\mathsf{F}(x, y, z, \cdot)$ for all possible $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and secret $z \in \mathcal{Z}$. Let $G : \mathcal{L} \to \mathcal{R}$ be a $(\mathcal{D}, \epsilon)$-nbPRG. Then, the* CDS *protocol $\mathsf{F}'_A : \mathcal{X} \times \mathcal{Z} \times \mathcal{L} \to \mathcal{T}$ and $\mathsf{F}'_B : \mathcal{Y} \times \mathcal{Z} \times \mathcal{L} \to \mathcal{T}$ defined via*

$$\mathsf{F}'_A(x, z, r') = \mathsf{F}_A(x, z, G(r')) \qquad \text{and} \qquad \mathsf{F}'_B(y, z, r') = \mathsf{F}_B(y, z, G(r')),$$

computes f with correctness error of $c + \epsilon$ and privacy error of $s + \epsilon$. Moreover, if the original protocol is perfectly correct then so is the new protocol.

Proof. Denote by F and by F' the joint computation of the original CDS and the new CDS, respectively. For every x, y, z, it holds that the distributions

$$\mathsf{F}(x, y, z, r) \qquad \text{and} \qquad \mathsf{F}'(x, y, z, r'), \qquad \text{where } r \xleftarrow{R} \mathcal{R}, r' \xleftarrow{R} \mathcal{R}',$$

are ϵ-close. Hence, the original simulator and the original decoder can be used for F' with an error larger by at most ϵ. To see that perfect correctness is preserved, observe that, for every x, y, z, the support of $\mathsf{F}'(x, y, z, r')$ is always a subset of the support of $\mathsf{F}(x, y, z, r)$. □

The above claim can be used to reduce the randomness complexity of low-complexity CDS while preserving computational efficiency by relying on a suitable family of nbPRGs. Notable examples include the case where the CDS encoding functions have low algebraic degree (e.g., linear) or low

[10] A slightly better bound (with t instead of $2t + \log t$) can be achieved with more careful arguments, in particular by using Sanov's theorem to directly bound deviation in statistical distance rather than looking at the difference in each probability. However, this only improves our bounds by constant factors.

computational complexity (e.g., polynomial-time computable). For both cases explicit constructions of nbPRGs exist in the literature. In this paper efficiency is not a major concern and we instantiate Claim 1 with a general-purpose inefficient nbPRG whose existence follows easily from the probabilistic method.

Claim 2. *For every $\rho, t \in \mathbb{N}$, family $\mathcal{D}$ of functions from $\{0,1\}^\rho$ to $\{0,1\}^t$, and $\epsilon > 0$, a random function $G : \{0,1\}^\ell \to \{0,1\}^\rho$ with $\ell = 2t + \log t + \log(\log|\mathcal{D}|) + 2\log(1/\epsilon)$ is with probability at least $2/3$ a $(\mathcal{D}, \epsilon)$-nbPRG.*

The claim follows from standard calculation and is given here for completeness.

Proof. Choose a random mapping $G : \{0,1\}^\ell \to \{0,1\}^\rho$ and let S be the image of G. We view S as a set of $L = 2^\ell$ randomly chosen strings from $\{0,1\}^\rho$. We show that G is $(\mathcal{D}, \epsilon)$-nbPRG with probability $2/3$. Fix $D \in \mathcal{D}$ and v in the image of D, and let

$$p_v = \Pr[D(U_\rho) = v] = |D^{-1}(v)|/2^\rho.$$

Let q_v be the fraction of elements in S that hit the set $D^{-1}(v)$. By an additive Chernoff bound, the probability that $|q_v - p_v| > \epsilon/2^t$ is at most $2e^{-2\epsilon^2 L/2^{2t}}$. Applying union-bound over all r's, we get that except with probability $2e^{-2\epsilon^2 L/2^{2t}}2^t$, the generator G ϵ-fools D. By applying another union-bound over all D's, we get that G fails to be $(\mathcal{D}, \epsilon)$-nbPRG with probability at most $2e^{-2\epsilon^2 L/2^{2t}}2^t|\mathcal{D}|$ which is smaller than $1/3$ for $\ell = 2t + \log(\log|\mathcal{D}| + t) + 2\log(1/\epsilon) \leq 2t + \log t + \log(\log|\mathcal{D}|) + 2\log(1/\epsilon)$. □

Lemma 1 follows by combining Claims 1 and 2.

5 Bounding Perfectly Correct CDS by coNP Complexity

In this section we bound the communication complexity of perfectly correct CDS protocols in terms of coNP Complexity. We need the following lemma.

Lemma 2. *In any perfectly correct* CDS *protocol for f with s-privacy (for some $s < 1/2$) that has randomness complexity ρ,*

$$2\rho \geq \text{coNP}(f)$$

Moreover if the CDS *is also perfectly private, then $\rho \geq \text{coNP}(f)$.*

Theorem 3, the bound on perfectly correct CDS protocols, follows from Lemma 2 and Lemma 1 (which allows us to bound ρ in terms of the CDS communication complexity t), by setting the error ϵ in the latter to be a very small constant.

Proof of Lemma 2. Let $\mathsf{F} = (\mathsf{F}_A, \mathsf{F}_B)$ be a perfectly correct CDS protocol for f with less than $1/2$ privacy error that has randomness complexity ρ. The coNP witness, for any (x, y) such that $f(x, y) = 0$, is a pair of randomness strings $r_0, r_1 \in \mathcal{R}$ such that $\mathsf{F}(x, y, 0, r_0) = \mathsf{F}(x, y, 1, r_1)$. The length of this witness is at most 2ρ.

Given a candidate witness (r_0, r_1), Alice accepts if $\mathsf{F}_A(x, 0, r_0) = \mathsf{F}_A(x, 1, r_1)$, and Bob accepts if $\mathsf{F}_B(y, 0, r_0) = \mathsf{F}_B(y, 1, r_1)$. This way, there is no interaction necessary, and they both accept if and only if $\mathsf{F}(x, y, 0, r_0) = \mathsf{F}(x, y, 1, r_1)$.

Such a witness (r_0, r_1) exists when $f(x, y) = 0$, as the distance between $\mathsf{F}(x, y, 0, r)$ and $\mathsf{F}(x, y, 1, r)$ is at most $2s$, which is less than 1, and so there has to be an intersection in the range of $\mathsf{F}(x, y, 0, \cdot)$ and $\mathsf{F}(x, y, 1, \cdot)$. No such (r_0, r_1) exists when $f(x, y) = 1$ due to perfect correctness, as in this case $\mathsf{F}(x, y, 0, \cdot)$ and $\mathsf{F}(x, y, 1, \cdot)$ have disjoint ranges. Thus, the coNP complexity of f is at most 2ρ.

To prove the moreover part note that, in the case of perfect privacy, we have the additional guarantee that for any $r_0 \in \mathcal{R}$, there is some $r_1 \in \mathcal{R}$ such that $\mathsf{F}(x, y, 0, r_0) = \mathsf{F}(x, y, 1, r_1)$. Thus, we can replace the string r_0 with some fixed canonical string (e.g., the all zero string) and reduce the witness size to ρ by having just r_1 as the witness. □

6 Bounding Perfectly Private CDS by PP Complexiy

The PP complexity of f, denoted $\mathrm{PP}(f)$, corresponds to the sum of the communication complexity and number of private random bits (represented by the $\log(1/\alpha)$ term in the definition below) used by a communication protocol that computes f correctly with probability more than $1/2$. Note that here we do not allow shared randomness.

Definition 3 (PP)**.** A randomised communication protocol Π is a PP protocol for f if it computes f correctly with probability more than $1/2$. Suppose Π uses t bits of communication; let α be the largest number such that $\min_{(x,y)} \Pr[\Pi(x, y) = f(x, y)] \geq 1/2 + \alpha$. The cost of Π is $(t + \log(1/\alpha))$, and the PP complexity of f is the minimum cost of any such protocol for f.

Recall that Theorem 4 asserts that the complexity of a perfectly private CDS protocol is lower-bounded by the PP communication complexity of f. The proof is based on the following lemma, which is proven after the proof of the theorem below. For the purposes of this section, we define the output $\Pi(x, y)$ of an interactive communication protocol Π on input (x, y) to be 1 if both parties output 1, and 0 otherwise – this is without loss of generality if there is no bound on the number of rounds of communication allowed, as each party can communicate its decision to the other with one additional bit.

Lemma 3. *In any c-correct s-private* CDS *protocol for f that communicates at most t bits and has randomness complexity ρ, for any $\eta > 0$ such that $(2s^2 + \eta) < (1 - 2c)^2/2^{t+1}$, it holds that*

$$2(t + \rho) + \log(1/\eta) + O(1) \geq \mathrm{PP}(f).$$

Proof of Theorem 4. Given a perfectly private CDS protocol for f with correctness error 0.1, communication t, we first apply Lemma 1 with approximation error $\epsilon = 0.01/2^{t/2}$ to get another CDS protocol for f that has privacy error $s = \epsilon$, correctness error $c = (0.1 + \epsilon)$, communication t, and randomness complexity:

$$\rho = 2t + \log t + \log n + 2\log(1/\epsilon) \leq 4t + \log(n) + O(1)$$

To this latter CDS protocol, we apply Lemma 3 with the gap $\eta = 0.01/2^t$ (noting that the condition $(2s^2 + \eta) < (1 - 2c)^2/2^t$ is indeed satisfied) to get:

$$\begin{aligned}
\mathrm{PP}(f) &\leq 2(t + \rho) + \log(1/\eta) + O(1) \\
&\leq 2(t + \rho) + t + O(1) \\
&= 3t + 2\rho + O(1) \\
&\leq 11t + 2\log n + O(1)
\end{aligned}$$

This proves the theorem. □

In order to prove Lemma 3, we need to construct an appropriate PP protocol for f from a given CDS protocol. In order to do this, we make use of the following randomized protocol that accepts with probability related to the L_2 distance between distributions. This is adapted from a similar protocol in the query setting from [BCH+17].

Lemma 4. *There exists a two-party randomised protocol Π that, given black-box access to two two-party protocols D_0 and D_1, has the following properties. Suppose that the protocol D_0 (resp., D_1) uses ρ bits of private randomness and t bits of communication, and at the end of these protocols one party obtains a sample from a probability distribution denoted, by abuse of notation, by D_0 (resp., D_1). Then, the protocol Π uses $(2\rho+2)$ bits of private randomness and $(2t+2)$ bits of communication, and the protocol Π accepts (outputs 1) with probability $1/2 + \|D_0 - D_1\|_2^2/8$.*

Proof of Lemma 4. In the protocol we construct, Alice first picks two random bits and sends them to Bob. These bits are used to decide upon one of the following subsequent tasks to perform. In each case, Bob always accepts.

1. With probability 1/4, the parties sample a and b from D_0. Alice accepts with probability 1 if $a = b$, and with probability 1/2 otherwise.
2. With probability 1/4, they do the same with D_1.
3. With the remaining probability 1/2, they pick one sample a from D_0 and one sample b from D_1. Alice rejects if $a = b$, and accepts with probability 1/2 otherwise.

The probability that Alice accepts is calculated as follows:

$$\begin{aligned}\Pr\left[\text{Alice accepts}\right] = \frac{1}{4}\left(\sum_a D_0(a)^2 + \sum_{a\neq b}\frac{D_0(a)D_0(b)}{2}\right) + \frac{1}{4}\left(\sum_a D_1(a)^2 + \sum_{a\neq b}\frac{D_1(a)D_1(b)}{2}\right) \\ + \frac{1}{2}\left(\sum_{a\neq b}\frac{D_0(a)D_1(b)}{2}\right)\end{aligned}$$

Rearranging the terms in the first two parentheses and adding and subtracting $\sum_a D_0(a)D_1(a)/2$ to the last, we have:

$$\begin{aligned}\Pr\left[\text{Alice accepts}\right] &= \frac{1}{4}\left(\sum_a \frac{D_0(a)^2}{2} + \sum_{a,b}\frac{D_0(a)D_0(b)}{2}\right) + \frac{1}{4}\left(\sum_a \frac{D_1(a)^2}{2} + \sum_{a,b}\frac{D_1(a)D_1(b)}{2}\right) \\ &\quad + \frac{1}{2}\left(\sum_{a,b}\frac{D_0(a)D_1(b)}{2} - \sum_a \frac{D_0(a)D_1(a)}{2}\right) \\ &= \frac{1}{4}\left(\sum_a \frac{D_0(a)^2}{2} + \frac{1}{2}\right) + \frac{1}{4}\left(\sum_a \frac{D_1(a)^2}{2} + \frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{2} - \sum_a \frac{D_0(a)D_1(a)}{2}\right) \\ &= \frac{1}{2} + \frac{1}{8}\sum_a \left(D_0(a)^2 + D_1(a)^2 - 2D_0(a)D_1(a)\right) \\ &= \frac{1}{2} + \frac{\|D_0 - D_1\|_2^2}{8}\end{aligned}$$

This proves the lemma. □

We finish by proving Lemma 3.

Proof of Lemma 3. Given a c-correct s-private CDS protocol $\mathsf{F} = (\mathsf{F}_A, \mathsf{F}_B)$ for f, we wish to constuct a PP protocol for f. We first construct a protool that,given input (x, y) and a bit b, samples the distribution $\mathsf{F}(x, y, b, \cdot)$ with communication $(t + \rho)$. This sampling protocol is as follows:

1. Alice picks $r \xleftarrow{R} \mathcal{R}$ and computes $\mathsf{F}_A(x, b, r)$.
2. Alice sends r to Bob.
3. Bob computes $\mathsf{F}_B(y, b, r)$ and sends it to Alice.

At the end of the above protocol, Alice has a sample from $\mathsf{F}(x, y, 0, \cdot)$ or $\mathsf{F}(x, y, 1, \cdot)$, according to the input b – call these distributions $\mathsf{F}_0^{x,y}$ and $\mathsf{F}_1^{x,y}$ for convenience. In other words, for any (x, y), the distributions $\mathsf{F}_0^{x,y}$ and $\mathsf{F}_1^{x,y}$ can be sampled by protocols with communication $(t + \rho)$.

Invoking Lemma 4, we then have a randomised protocol – call it Π – with communication complexity $2(t + \rho) + 2$ that, given input (x, y), accepts with probability $1/2 + \|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_2^2/8$. We will now show that this probability is above or below a certain threshold depending on $f(x, y)$.

By the relationships between the L_1 and L_2 norms, and noting that both distributions are supported in a domain of size 2^t, we have:

$$\frac{\|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_1^2}{2^t} \leq \|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_2^2 \leq \|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_1^2 \tag{1}$$

If $f(x, y) = 0$, the privacy of the CDS protocol guarantees that:

$$\|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_1 = 2\,\Delta\left(\mathsf{F}_0^{x,y}; \mathsf{F}_1^{x,y}\right) \leq 2\left(\Delta\left(\mathsf{Sim}(x, y); \mathsf{F}_0^{x,y}\right) + \Delta\left(\mathsf{Sim}(x, y); \mathsf{F}_1^{x,y}\right)\right) \leq 4s \tag{2}$$

where the first inequality follows from the triangle inequality for statistical distance.

Similarly, if $f(x, y) = 1$, the correctness of the CDS protocol guarantees that:

$$\begin{aligned}\|\mathsf{F}_0^{x,y} - \mathsf{F}_1^{x,y}\|_1 &= 2\,\Delta\left(\mathsf{F}_0^{x,y}; \mathsf{F}_1^{x,y}\right) \\ &\geq 2\left(\Pr\left[\mathsf{Dec}(x, y, \mathsf{F}_1^{x,y}) = 1\right] - \Pr\left[\mathsf{Dec}(x, y, \mathsf{F}_0^{x,y}) = 1\right]\right) \\ &\geq 2(1 - 2c)\end{aligned} \tag{3}$$

where the first inequality follows from the fact that the statistical distance is the greatest advantage any algorithm could have in distinguishing between two distributions.

Putting together Eqs. (1) and (2), the protocol Π accepts with probability at most $1/2 + 2s^2$ (which we denote p_0) if $f(x, y) = 0$. And if $f(x, y) = 1$, by Eqs. (1) and (3), it accepts with probability at least $1/2 + (1 - 2c)^2/2^{t+1}$ (which we denote p_1).

Finally, we recenter the probabilities of acceptance so that the protocol accepts with probability more than $1/2$ if $f(x, y) = 1$ and less than $1/2$ if $f(x, y) = 0$. We make use of the fact that, by our hypothesis, $p_0 + \eta < p_1$. The final protocol would ideally be as follows:

1. With probability $1/(p_0 + p_1)$, run protocol Π on the given input (x, y).
2. Otherwise, reject.

It may be verified that if $f(x, y) = 1$, the above protocol accepts with probability more than $1/2 + \eta/4$ (nearly $1/2 + \eta/2$, in fact), and if $f(x, y) = 0$ with probability less than $1/2 - \eta/4$.

However, doing something with probability exactly $1/(p_0 + p_1)$ would require (unless p_0 and p_1 are well-behaved) a large, possibly unbounded, number of random bits. It is sufficient for our

purposes, however, to approximate this probability up to an error somewhat smaller than $\eta/4$. With $(\log(1/\eta) + O(1))$ random bits, we can approximate it to, say, $\eta/8$, and the probabilities of acceptance would still be bounded away from $1/2$ by $\Omega(\eta)$.

Thus, this is a PP protocol for f with complexity at most $(2(t+\rho) + \log(1/\eta) + O(1))$. This proves the lemma. □

7 Asymmetry in CDS and One-Way Communication

In this section, we show that the trade-offs that can be achieved between the communication of the two parties in a CDS protocol is limited in certain ways by the one-way communication complexity of the predicate involved.

Denote by $\mathrm{R}_{B\to A}(f)$ the one-way randomized communication complexity of $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ where Bob (who gets inputs from $\mathcal{Y}$) sends the message; $\mathrm{R}_{A\to B}(f)$ is defined similarly. For concreteness, we do not allow shared randomness between Alice and Bob, though allowing this would only affect our bounds by a factor of 2. We prove the following stronger version of Theorem 6 from the introduction.

Theorem 7. *In any perfectly correct s-private* CDS *protocol for f in which Alice and Bob communicate t_A and t_B bits respectively and the total input length of the function is n, if $s < 1/8 - \epsilon$ for some ϵ, it holds that:*

$$\frac{2^{t_A}(t_A + t_B + \log n + \log(1/\epsilon))}{(1 - 8(s+\epsilon))^2} \geq \Omega(\mathrm{R}_{B\to A}(f)).$$

In particular, in our default setting where the soundness error is at most 0.1*, we have:*

$$2^{t_A}(t_A + t_B + \log n + O(1)) \geq \Omega(\mathrm{R}_{B\to A}(f)).$$

Theorem 7 follows immediately from Lemma 1 from Section 4.2 and Lemma 5 below, which proves a similar bound, but involves the randomness complexity of the CDS protocol.

Lemma 5. *In any perfectly correct* CDS *protocol for f with s-privacy for some $s < 1/8$,*

$$2^{t_A}\rho/(1-8s)^2 \geq \Omega(\mathrm{R}_{B\to A}(f))$$

Proof of Lemma 5. Given a CDS protocol $(\mathsf{F}_A, \mathsf{F}_B)$, consider the following one-way communication protocol:

1. Bob (who has input y) picks a random $r \in \mathcal{R}$ and computes $b \xleftarrow{R} \mathsf{F}_B(y, 0, r)$.
2. Bob samples random $r' \in \mathcal{R}$ such that $\mathsf{F}_B(y,0,r) = \mathsf{F}_B(y,1,r') = b$. If no such r' exists, set $r' = \perp$.
3. Bob sends (r, r') to Alice.
4. Alice accepts if either $r' = \perp$ or $\mathsf{F}_A(x,0,r) \neq \mathsf{F}_A(x,1,r')$.

Note that Alice rejects if and only if $\mathsf{F}(x,y,0,r) = \mathsf{F}(x,y,1,r')$. Due to perfect correctness, if $f(x,y) = 1$, such r and r' do not exist, and so Alice always accepts. We show that if $f(x,y) = 0$, Alice rejects with some noticeable probability.

Fix some (x,y) such that $f(x,y) = 0$. For convenience, let $\mathsf{F}_A(x,1,\perp)$ be $\perp$. For $\sigma \in \{0,1\}$, denote the random variable $(\mathsf{F}_A(x,0,r), \mathsf{F}_B(y,0,r))$ by (A,B) and the random variable $(\mathsf{F}_A(x,1,r), \mathsf{F}_B(y,1,r))$ by (A',B'). The above protocol can be rephrased as follows:

1. Sample $(a, b) \xleftarrow{R} (A, B)$.
2. Sample a' from A' conditioned on $B' = b$.
3. Accept if and only if $a = a'$.

For any $b \in \mathcal{T}_B$, denote by A_b the variable A conditioned on B being b. Similarly, denote by A'_b the variable A' conditioned on B' being b. The probability that Alice rejects can now be written as follows:

$$\sum_{b \in \mathcal{T}_B} \Pr[B = b] \cdot \Pr_{\substack{a \xleftarrow{R} A_b \\ a' \xleftarrow{R} A'_b}} \left[a = a'\right] \tag{4}$$

In order to bound the second terms in the product above, we use the following bound on the collision probability of two distributions in terms of their statistical distance, which we prove later in this section.

Proposition 1. *For any two distributions p and q over a domain of size m such that $\Delta(p; q) = \delta$,*

$$\Pr_{\alpha \xleftarrow{R} p, \beta \xleftarrow{R} q}[\alpha = \beta] = \sum_i p_i q_i \geq (1 - \delta)^2 / m.$$

Applying the above proposition to the two distributions A_b and A'_b for each b, we see that the probability that Alice rejects – the quantity in (4) – is bounded as:

$$\Pr[\text{Alice rejects}] \geq \sum_{b \in \mathcal{T}_B} \Pr_r[B = b] \cdot \left(1 - \Delta\left(A_b; A'_b\right)\right)^2 / 2^{t_A} \tag{5}$$

Next we show that, with high probability over B, the quantity that the probability is multiplied with in the above expression is not too small. For this, we make use of the following simple proposition.

Proposition 2. *For any random variables (A, B) and (A', B') it holds that*

$$\mathop{\mathrm{E}}_{b \xleftarrow{R} B}\left[\Delta\left(A_b; A'_b\right)\right] \leq 2\,\Delta\left((A, B); (A', B')\right),$$

where A_b (resp., A'_b) denotes the random variable A conditioned on B being b (resp., random variable A' conditioned on B' being b).

Proof of Proposition 2. By the definition of statistical distance, the LHS equals to

$$\Delta\left((A_B, B); (A'_B, B)\right) = \Delta\left((A, B); (A'_B, B)\right).$$

By some elementary manipulations, we can bound the above as:

$$\begin{aligned}
\Delta\left((A, B); (A'_B, B)\right) &\leq \Delta\left((A, B); (A'_{B'}, B')\right) + \Delta\left(B; B'\right) \\
&= \Delta\left((A, B); (A', B')\right) + \Delta\left(B; B'\right) \\
&\leq 2\,\Delta\left((A, B); (A', B')\right),
\end{aligned}$$

where the last relation is by the data processing inequality. □

In our setting, the s-privacy of the CDS protocol implies that $\Delta\left((A,B);(A',B')\right) \leq 2s$, and so, by Proposition 2,

$$\mathop{\mathrm{E}}_{b \xleftarrow{R} B}\left[\Delta\left(A_b;A'_b\right)\right] \leq 4s.$$

Applying Markov's inequality, we get that

$$\Pr_{b \xleftarrow{R} B}\left[\left(1-\Delta\left(A_b;A'_b\right)\right)^2 < (1-8s)^2\right] = \Pr_{b \xleftarrow{R} B}\left[\Delta\left(A_b;A'_b\right) > 8s\right] < \frac{1}{2}. \tag{6}$$

Using the Eqs. (5) and (6), we can bound the rejection probability of Alice as:

$$\Pr\left[\text{Alice rejects}\right] \geq \sum_{b \in \mathcal{T}_B} \Pr_r\left[B=b\right] \cdot \left(1-\Delta\left(A_b;A'_b\right)\right)^2 / 2^{t_A} \geq \frac{1}{2} \cdot \frac{(1-8s)^2}{2^{t_A}}$$

Thus, by repeating the above protocol $\Theta(2^{t_A}/(1-8s)^2)$ times, we get a one-way communication protocol for f that is correct with a large constant probability. The communication complexity of the resulting protocol is $\Theta(2^{t_A}\rho/(1-8s)^2)$. □

Finally, we prove Proposition 1, which is a bound on collision probability of two distributions in terms of their statistical distance.

Proof of Proposition 1. Given distributions p and q over a domain of size m such that $\Delta\left(p;q\right) = \delta$, we wish to bound their collision probability $\sum_i p_i q_i$. For any i, let $s_i = \min(p_i, q_i)$. We immediately have the following relationship:

$$\sum_i p_i q_i \geq \sum_i s_i^2 \tag{7}$$

We next aim to relate the sum of s_i^2 to the self-collision probability of the distribution obtained by normalizing the s_i's. The self-collision probability of any distribution is always bounded below by $1/m$ due to the relationship between the L_2 and L_1 norms. This gives us the following:

$$\sum_i s_i^2 = \left(\sum_i s_i\right)^2 \cdot \sum_i \left(\frac{s_i}{\sum_i s_i}\right)^2 \geq \frac{\left(\sum_i s_i\right)^2}{m} \tag{8}$$

It is left to bound the above normalization factor. Let I_p denote the set of i's where $p_i \geq q_i$. We can bound the sum of the s_i's as follows:

$$\begin{aligned}
\sum_i s_i &= \sum_{i \in I_p} q_i + \sum_{i \notin I_p} p_i \\
&= \sum_{i \in I_p} (q_i - p_i + p_i) + \sum_{i \notin I_p} p_i \\
&= \sum_{i \in I_p} (q_i - p_i) + \sum_i p_i \\
&= -\delta + 1
\end{aligned} \tag{9}$$

where the last relation follows from the definition of statistical distance. Putting Eqs. (7) to (9) together gives us the proposition. □

8 Imperfect CDS vs. Interactive Proofs

Recall that Theorem 5 shows that (imperfect) CDS complexity is at least polynomial in the AM communication complexity of f. As mentioned in the introduction, the theorem is proved in two steps. In Section 8.2 we show that a CDS protocol can be transformed into a constant-round private-coins interactive-proof (that, in addition, achieves a zero-knowledge property). Then, in Section 8.3, we note that, just like in the computational setting, such interactive proofs can be converted to an AM protocol with polynomial overhead [Bab85, GS89]. The definitions of the relevant communication complexity measures are given in Section 8.1.

8.1 Defining IP and AM and zero-knowledge classes

Communication complexity analogues of interactive proof systems [Bab85, GMR85, BM88] have been considered in [Lok01, Kla11]. (This manner of defining communication classes based on complexity classes is first seen in [BFS86].) Below we define these measures in a slightly non-standard way in order to capture both the private-coins and public-coins setting in a unified way.

Both the models of IP and AM communication complexity involve three parties Alice, Bob, and a prover (sometimes referred to as Merlin). Alice and Bob hold inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, for a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Merlin knows both x and y, and aims to prove to Alice and Bob that $f(x,y) = 1$. We further assume that Alice and Bob have access to a shared random string that is not visible to Merlin, and that Alice and Bob have no other source of randomness.

Definition 4 (IP)**.** An IP protocol for a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with k rounds proceeds as follows. Each round begins with a two-party protocol between Alice and Bob after which Alice sends a message to Merlin, who sends back a message to Alice. At the end of all k rounds, Alice and Bob communicate and Alice decides whether to accept or reject. The protocol is said to compute f with completeness error of ϵ and soundness error of δ if it satisfies the following properties:

- **Completeness:** For all inputs (x,y) with $f(x,y) = 1$, the probability that there exists a proof strategy for Merlin such that (x,y) is accepted is at least $1 - \epsilon$.
- **Soundness:** For all inputs (x,y) with $f(x,y) = 0$, the probability that there exists a proof strategy such that (x,y) is accepted is at most δ.

The cost of an IP-protocol is the maximum over all inputs of the total communication complexity of the protocol. The IP_k complexity of f, denoted $\mathrm{IP}_k(f)$ is the smallest cost of a k-round IP-protocol computing f with soundness and completeness error of $\epsilon = \delta = 1/3$. We use similar notation to specify the number of rounds in the complexity measures defined below as well.

Definition 5 (AM)**.** An IP protocol for a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ is a *public-coin protocol* if all the messages from Alice to Merlin consist of public coins (that are shared with Bob), and, eventually, all the shared coins are being sent to Merlin. This means that the only communication between Alice and Bob takes place at the end, and the final acceptance decision is a *deterministic* function of the transcript and the inputs x and y. A single round (i.e., 2-move) public-coin protocol with soundness error of $\delta = 1/3$ and perfect completeness (i.e., $\epsilon = 0$) is referred to as an AM protocol. The cost of an AM-protocol is the maximum over all inputs of the

total communication complexity of the protocol. The AM complexity of f, denoted by $\mathrm{AM}(f)$, is the smallest cost of an AM-protocol computing f.[11]

We move on to define zero-knowledge protocols. Our definition borrows from a similar one in [BCH+17].

Definition 6 (SZK and SZK′)**.** Let Π be a (k-round) IP protocol for a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. We say that a randomized two-party public-coin protocol Π_S, involving parties S_A and S_B, *simulates* Π with an error of η if for every pair of inputs (x, y) for which $f(x, y) = 1$ the following holds. When Π_S is invoked with inputs x and y for the parties S_A and S_B, respectively, the parties produce outputs $\widehat{V}_A$ and $\widehat{V}_B$, respectively such that

$$\Delta\left((V_A, V_B); (\widehat{V}_A, \widehat{V}_B)\right) \leq \eta,$$

where V_A and V_B are (jointly distributed) random variables that describe the view of Alice and Bob when the protocol Π is invoked on x and y (and Merlin follows the honest strategy defined in Π). Here, the *view* of a party consists of all messages sent and received by it, along with any private or shared randomness that it sees.[12]

Let c_M be the number of bits of communication between Merlin and Alice, c_V that between Alice and Bob, and c_S that between S_A and S_B. The cost of such an SZK protocol is $(c_M + \max(c_V, c_S))$. The SZK complexity of f, denoted $\mathrm{SZK}(f)$, is the smallest cost of an SZK protocol computing f with simulation error at most 0.1.

Further, let ρ_V be the number of bits of randomness used by Alice and Bob, and ρ_S be that used by S_A and S_B. The SZK′ cost of such an SZK protocol is $(c_M + \max(c_V, c_S) + \max(\rho_V, \rho_S))$. The SZK′ complexity of f, denoted $\mathrm{SZK}'(f)$, is the smallest SZK′ cost of an SZK protocol for f with correctness, privacy and simulation errors all negligible in the input size.

8.2 CDS implies 1-round zero-knowledge interactive proof

Here, we show that a CDS protocol of a predicate f can be converted into a single-round (private-coin) interactive proof protocol. Moreover, the protocol is also zero-knowledge.

Lemma 6. *Suppose that the predicate f has a* CDS *with privacy and correctness error of ϵ that supports secrets of length ℓ bits with communication complexity of t. Then, f has an* IP_1 *protocol with communication of $\ell + t$ completeness error of ϵ and soundness error of $2^{-\ell} + \epsilon$. Moreover, the protocol is zero-knowledge with simulation error of ϵ where the communication complexity of the simulator is t.*

Proof. The idea is to imitate the standard interactive proof of Graph Nonisomorphism [GMW91] (or more generally for the Statistical Difference problem [SV03]). Given a CDS protocol $(\mathsf{F}_A, \mathsf{F}_B)$ for f, we construct an IP_1 protocol for f as follows.

1. Alice and Bob pick a random secret message $z \in \{0,1\}^\ell$ and a random $r \in \mathcal{R}$ using their common randomness.

[11] In some references the AM complexity is defined as the minimal length of Merlin's message. The difference between the two variants does not affect our results since the complexity of our AM protocols is dominated anyway by Merlin's message.

[12] Our definition only deals with honest-verifier SZK protocols; the case of malicious verifiers demands study on its own.

2. Bob sends to Alice the value $\mathsf{F}_B(y, z, r)$ and Alice delivers it to Merlin together with $\mathsf{F}_A(x, z, r)$.
3. Merlin sends z' to Alice.
4. Alice accepts if and only if $z' = z$.

The completeness of this protocol follows from the correctness of the CDS protocol – if $f(x, y) = 1$, then $(\mathsf{F}_A(x, b, r), \mathsf{F}_B(y, b, r))$ reveals the message z, except with probability ϵ, and Merlin can then guess it correctly.

Similarly, soundness follows from the privacy of the CDS protocol. If $f(x, y) = 0$ then the success probability p of of Merlin is at most ϵ larger than the success probability p' in a protocol in which Alice and Bob's messages are sampled by the CDS simulator. Since the simulator samples these messages independently of the secret z, we conclude that $p' = 2^{-\ell}$ and $p \leq \epsilon + 2^{-\ell}$.

Finally, the transcript of this protocol is simulated by running the Alice and Bob parts of the protocol and setting Merlin's response to always be z – if $f(x, y) = 1$, this is close to the actual transcript by the completeness of the protocol (as in this case Merlin's response, except with probability ϵ, is indeed z). □

In [AARV17, Theorem 5.1] it is shown that a standard CDS protocol (that supports a single bit secret with correctness error and privacy error of 0.1) can be upgraded into a CDS that supports $\Omega(k)$-bit secrets with correctness error and privacy error of 2^{-k} with only a multiplicative overhead of $O(k)$ in both communication and randomness complexity. By taking k to be a sufficiently large constant, we get the following theorem.

Theorem 8. *For any Boolean function f,* $\mathrm{CDS}(f) \geq \Omega(\mathrm{IP}_1(f))$. *Moreover,* $\mathrm{CDS}(f) \geq \Omega(\mathrm{SZK}_1(f))$.

More generally, one can achieve an SZK_1 protocol for f with correctness, soundness and simulation errors of ϵ with complexity of $O(\mathrm{CDS}(f) \log(1/\epsilon))$.

8.3 From IP_1 to AM

We observe that, as in the computational setting, 2-round private-coins interactive proofs can be converted to single round Arthur-Merlin protocols with polynomial communication overhead.

Theorem 9. *For every predicate f,* $\mathrm{IP}_1(f) \geq \mathsf{poly}(\mathrm{AM}(f), \log n)$ *where n is the length of the inputs to f.*

The proof follows by applying the following lemmas.

Lemma 7 (follows from [GS89])**.** *Suppose f has a single-round private-coin* IP *with communication and randomness complexity at most t and error of at most $1/3$. Then, f also has a two-round public-coin* IP *protocol with perfect completeness, soundness error of $1/3$, and communication and randomness complexity of at most* $\mathsf{poly}(t)$.

Proof Sketch. Goldwasser and Sipser (GS) [GS89] proved that, in the computational setting, k-round IP protocol Π can be converted into a $(k+1)$-round IP public-coin protocol Π' with perfect completeness. Let us focus on $k = 1$ and let us assume that Π has an error which is negligible in t. (This can be guaranteed via standard amplification at the expense of increasing the randomness and communication by a polynomial or even polylogarithmic factor). Since $k = 1$, the verifier in Π is specified by a query generation algorithm, and by an acceptance predicate $V_\Pi(w, r, a)$ that given an input to the problem $w = (x, y)$, private randomness r, and the prover's answer a decides whether to accept the proof. We note that, in this setting, the GS transformation has the following properties:

1. The communication complexity of Π' is polynomial in the communication and randomness of Π.

2. In Π' Arthur decides whether to accept the transcript by computing a conjunction of two predicates. The first predicate depends only on the randomness and on Merlin's messages (but is independent of the input $w = (x, y)$ to the language). The second predicate is simply the acceptance predicate $V_\Pi(w, r, a)$ of the verifier of Π, applied to the input $w = (x, y)$ and to a pair (r, a) that is being sent as part of Merlin's message in Π'.

We can therefore apply the transformation in the communication complexity setting. In particular, Alice's final decision in Π' is determined by locally computing C (based on Merlin's messages and the shared randomness) and by running with Bob the protocol for computing $V_\Pi((x, y), r, a)$. The communication complexity of this protocol is upper-bounded by the communication complexity of Π. □

It is not hard to verify that the round-reduction of Babai [Bab85] immediately applies to the communication complexity setting.

Lemma 8 ([Bab85])**.** *Suppose f has a two-round public-coin* IP *protocol with communication complexity at most t. Then,* $\mathrm{AM}(f) \leq O(t \log t)$.

Proof Sketch. Let t denote the communication complexity of the two-round public-coin IP protocol Π. First, modify the protocol as follows. Alice begins by sending her first message q_1 as in Π, Merlin responds with the corresponding response a_1 as in Π. In the second round, Alice sends $\ell = O(\log t)$ fresh random copies $\vec{q}_2 = (q_2^1, \ldots, q_2^\ell)$ of her second message in Π, and Merlin responds with $(a_2^1, \ldots, a_2^\ell)$. Finally, Alice and Bob accept if, for every $i \in [\ell]$ the transcript (q_1, a_1, q_2^i, a_2^i) is accepted by the original protocol. It is not hard to see that for no instances, conditioned on most prefixes q_1, the soundness error is 2^{-2t}. Therefore, by a union-bound over all a_1, with high probability over $q_1, \vec{q}_2$, every answer $(a_1, \vec{a}_2)$ will be rejected. Hence, the protocol can be further modified into a single round protcol in which Alice just sends $(q_1, \vec{q}_2)$ at the first move, gets back an answer $(a_1, \vec{a}_2)$, and accepts by running the acceptance protcol with Bob over all ℓ instances. The lemma follows. □

By combining Lemmas 7 and 8, we conclude a single-round interactive proof Π of f can be converted into an AM protocol Π' whose communication complexity is polynomial in the communication complexity and *randomness complexity* of Π. In order to prove Theorem 9, we prove that the randomness complexity of such Π is not much larger than the communication complexity of Π.

Lemma 9. *Suppose there is an* IP_1 *protocol for f with communication t and randomness ρ. Then there is an* IP_1 *protocol for f with communication $O(t)$ and randomness of $O(t + \log n)$, where $n = \log |\mathcal{X} \times \mathcal{Y}|$ is the total bit-length of the input.*

Proof. As in the proof of the CDS sparsification lemma (Lemma 1), we will use a properly chosen pseudorandom generator to partially derandomize the original protocol Π. Observe that the randomness (shared by Alice and Bob) affects two different computations in Π: (1) the generation of the query to the prover; and (2) the final reject/accept decision of Alice and Bob. Correspondingly, we will have to carefully define the class of adversaries that should be fooled by the PRG. Details follow.

Let $\mathcal{R}$ denote the randomness space of Π, let $\mathcal{Q}$ denote the space of all possible messages ("queries") sent by Alice to the prover, and let $\mathcal{Z}$ denote the set of all possible answers from the prover to Alice. For every fixing of input x, y, let $q_{x,y}(r)$ denote the mapping that takes a random string $r \in \mathcal{R}$ (shared by Alice and Bob) and maps it to the message $q \in \mathcal{T}$ that is being sent by Alice to the prover. For every fixed inputs x, y, the prover's strategy can be defined as a mapping $p_{x,y} : \mathcal{Q} \to \mathcal{Z}$ that maps a query q to an answer z. For every x, y and every possible prover strategy $p = p_{x,y}$, let $A_{x,y,p}$ be the mapping that takes randomness $r \in \mathcal{R}$ and outputs a bit b that determines whether the interaction that corresponds to inputs x, y randomness r and prover's message $z = p(q_{x,y}(r))$ is being accepted by Alice.

Let $\mathcal{D}$ be the set of adversaries $A_{x,y,p}$ that correspond to all possible choices of $x \in \mathcal{X}, y \in \mathcal{Y}$ and $p : \mathcal{Q} \to \mathcal{Z}$. Let $G : \mathcal{L} \to \mathcal{R}$ be a PRG that ϵ-fools the class $\mathcal{D}$. (Note that in this case $\mathcal{D}$ is a class of Boolean adversaries.) Consider the modified protocol Π' in which Alice and Bob share a random seed $s \overset{R}{\leftarrow} \mathcal{L}$ and invoke the protocol Π with shared randomness $r' = G(s)$.

We claim that the completeness error of Π' is at most ϵ larger than the completeness error of Π. Indeed, consider a yes instance x, y and assume that the prover in Π' uses the same optimal strategy $p^\star$ used in Π over the inputs x, y. Then, the rejection probability in Π' is

$$\Pr_{s \overset{R}{\leftarrow} \mathcal{L}} [A_{x,y,p^\star}(G(s)) = 0]$$

which is at most ϵ larger than the rejection probability in Π,

$$\Pr_{r \overset{R}{\leftarrow} \mathcal{R}} [A_{x,y,p^\star}(r) = 0] \leq 1/3.$$

We move on to analyze the soundness error. Fix a no instance (x, y) and consider an arbitrary strategy $p = p_{x,y}$ for the "cheating" prover. Then, the rejection probability in Π' is

$$\Pr_{s \overset{R}{\leftarrow} \mathcal{L}} [A_{x,y,p}(G(s)) = 0]$$

which is at most ϵ smaller than the rejection probability in Π,

$$\Pr_{r \overset{R}{\leftarrow} \mathcal{R}} [A_{x,y,p}(r) = 0] \geq 2/3.$$

Note that the number of functions in $\mathcal{D}$ is $2^{|\mathcal{X}|+|\mathcal{Y}|}|\mathcal{Z}|^{|\mathcal{Q}|}$ and therefore $\log\log|\mathcal{D}| \leq O(\log n + t))$ where n denotes the total input length and t denotes the communication complexity. By Claim 1, for $\epsilon = 0.01$, there exists a PRG that ϵ-fools $\mathcal{D}$ with seed length of $O(\log n + t)$. This yields a protocol Π' with communication complexity of t, randomness complexity of $O(\log n + t)$, and soundness and completeness error smaller than 0.4. (The latter can be reduced to $1/3$ by repeating the protocol in parallel a constant number of times. The total communication and randomness complexity are increased by a constant factor.) □

9 SZK and CDS

In this section, we further explore the relation between CDS complexity and SZK complexity. In particular, in Section 9.1 we introduce an interactive variant of CDS, referred to as CDS′, and show

(in Section 9.2) that the CDS′ complexity of a predicate is roughly the same, up to polynomial factor, as its SZK′ complexity. (Recall that the latter notion is the communication complexity analogue of statistical zero-knowledge protocols except that we measure both randomness and communication; See Section 8.1.) The proof is based on communication-complexity analogues of known completeness theorems for SZK protocols in the computational setting (as in [SV03]).

9.1 CDS with Interaction

Recall that standard CDS can be viewed as a non-interactive protocol for generating a CDS encoding (as per Remark 4.1). Below, we consider a variant in which the CDS encoding is generated using a two-party protocol. (Unlike the non-interactive setting, in this setting linear communication can be achieved trivially and so only sub-linear protocols are meaningful.)

Definition 7 (CDS′)**.** Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. A CDS′ protocol Π is a two-party protocol between Alice and Bob where Alice holds an input $x \in \mathcal{X}$, Bob holds an input $y \in \mathcal{Y}$, and both of them get a secret $z \in \mathcal{Z}$ and an access to common random string r. We assume that when the protocol terminates both parties get the same output, denoted by $\Pi(x, y, z; r)$. The random variable $\Pi(x, y, z)$, induced by a uniform choice of r, should satisfy the following requirements:

1. (c-Correctness) There exists a deterministic algorithm Dec, called a *decoder*, such that for every 1-input (x, y) of f and any secret $z \in \mathcal{Z}$ we have that:
$$\Pr[\mathsf{Dec}(x, y, \Pi(x, y, z)) \neq z] \leq c$$
2. (s-Privacy) There exists a simulator algorithm Sim such that for every 0-input (x, y) of f and any secret $z \in \mathcal{Z}$, the following holds:
$$\Delta\left(\mathsf{Sim}(x, y); \Pi(x, y, z)\right) \leq s$$

The *communication complexity* of the CDS′ protocol is the maximum, over all (x, y, z), of the sum of the length of the output $|\Pi(x, y, z)|$ and the number of bits of communication between Alice and Bob. Its *randomness complexity* is number of bits of shared randomness used. If c and s are zeros, such a CDS′ scheme is called *perfect*.

We denote by $\mathrm{CDS}'(f)$ the least communication complexity of any CDS′ protocol for f with correctness and privacy errors at most 0.1 (with single-bit secrets).

In a manner identical to that of Theorem 8, where it is shown that $\mathrm{CDS}(f)$ is lower-bounded by $\Omega(\mathrm{IP}_1(f))$ for any Boolean function f, it can be shown that $\mathrm{CDS}'(f) \geq \Omega(\mathrm{IP}_1(f))$. Together with Theorem 9, we now have the following.

Theorem 10. *For every Boolean function f on n-bit inputs,* $\mathrm{CDS}'(f) \geq \mathsf{poly}(\mathrm{AM}(f), \log n)$.

9.2 SZK′ and CDS′

Here, we show that the CDS′ complexity of any function f is roughly equivalent to its SZK′ complexity, upto being a small polynomial in it.

Theorem 11. *For any Boolean function f on n-bit inputs,*
$$\frac{\mathrm{SZK}'(f)}{\mathrm{polylog}(n)} \leq \mathrm{CDS}'(f) \leq O(\mathrm{SZK}'(f)^6)$$

Proof of Theorem 11. To show the lower bound, given a CDS′ protocol for f with communication t, we need to construct an SZK′ protocol for f, where the communication and randomness complexities are at most $t\,\mathrm{polylog}(n)$. To get such a bound on the randomness, we need the following analogue of the sparsification lemma for CDS (Lemma 1) that follows from an identical proof.

Lemma 10. *Suppose there is a* CDS′ *protocol for f with communication t and randomness ρ, and with correctness error c and privacy error s. Then, for any $\epsilon > 0$, there is a $(c+\epsilon, s+\epsilon)$-*CDS′ *protocol for f with communication t and randomness*

$$2t + \log t + \log n + 2\log(1/\epsilon),$$

where n is the total bit-length of the input and the secret. Further, if the former protocol is perfectly correct, then so is the latter.

We also note that the amplification procedure for CDS from [AARV17] extends immediately to CDS′, giving the following lemma.

Lemma 11. *Suppose there is a* CDS′ *protocol for f with communication t and randomness ρ, and correctness and privacy errors at most* 0.1*. Then, for every integer k there exists a* CDS′ *protocol for f with privacy and correctness errors of $2^{-\Omega(k)}$, communication complexity $O(kt)$, and randomness complexity $O(k\rho)$.*

Putting together the above two lemmas, the initial CDS′ protocol for f implies another protocol with both randomness and communication complexities $t \cdot \mathrm{polylog}(n)$, and privacy and correctness errors negligible in n. The required SZK′ protocol now follows exactly as an SZK protocol does from CDS in the proof of Lemma 6.

To show the upper bound, we show how to get a CDS′ protocol for a function f from any SZK′ protocol for it. We do this by showing (in Lemma 12) that an analogue of the Statistical Difference problem (which is complete for SZK in the computational setting) is, in an appropriate sense, "complete" for SZK′ communication complexity. Below, for a two-party protocol Π, we denote by $\Pi(x, y)$ the (distribution of the) joint output of the parties in Π when given inputs x and y respectively.

Lemma 12. *Suppose* $\mathrm{SZK}'(f) = t$*. Then, there are two-party randomized communication protocols Π_0 and Π_1, each with communication complexity, randomness complexity and output length all $O(t^6)$, with the following properties.*

- *If $f(x, y) = 1$, then $\Delta\left(\Pi_0(x, y); \Pi_1(x, y)\right) \geq 0.99$.*
- *If $f(x, y) = 0$, then $\Delta\left(\Pi_0(x, y); \Pi_1(x, y)\right) \leq 0.01$.*

Given f such that $\mathrm{SZK}'(f) = t$, we wish to construct a CDS′ protocol for it with complexity at most $O(t^6)$. For this, we use the pair of protocols (Π_0, Π_1) implies by Lemma 12. Our CDS′ protocol $\Pi(x, y, z)$ is simply as follows:

- Alice and Bob run the protocol $\Pi_0(x, y)$ if $z = 0$ and $\Pi_1(x, y)$ if $z = 1$, and output whatever these protocols output.

For correctness, note that whenever $f(x, y) = 1$, by the guarantees of Lemma 12, we have:

$$\Delta\left(\Pi(x, y, 0); \Pi(x, y, 1)\right) = \Delta\left(\Pi_0(x, y); \Pi_1(x, y)\right) \geq 0.99$$

Thus the ideal distinguisher between $\Pi_0(x, y)$ and $\Pi_1(x, y)$ can be used as the decoder. When the secret z is picked uniformly at random, by the properties of statistical distance, it will be correct with probability at least $1/2 + 0.99/2 = 0.995$. So, for any z, it will be correct with probability at least 0.99.

For privacy, the simulator $\mathsf{Sim}(x, y)$ runs $\Pi(x, y, 0)$ and outputs whatever it outputs. Again, from Lemma 12, when $f(x, y) = 0$, for any z,

$$\Delta\left(\mathsf{Sim}(x, y); \Pi(x, y, z)\right) = \Delta\left(\Pi(x, y, 0); \Pi(x, y, z)\right) \leq \Delta\left(\Pi_0(x, y); \Pi_1(x, y)\right) \leq 0.01$$

The communication complexity of this CDS′ protocol is the length of the outputs of Π_0 and Π_1, which is $O(t^6)$. This proves the theorem. □

Lemma 12 is proven along the lines of the reduction to the "Statistical Difference" problem from any problem with honest-verifier SZK proofs in the computational setting. The following lemma is a quantitative statement of this completeness that is implicit in [SV03] and [Vad99, Chap. 3].

Lemma 13. *Suppose that a language L has a statistical zero-knowledge protocol with simulator S. Then, there exist a pair of oracle-aided circuits C_0 and C_1 that take as input an instance x and a random string r such that, when given oracle access to S, the distributions of $C_0^S(x, r)$ and $C_1^S(x, r)$, induced by a uniform random choice of r, have the following properties:*

- *If $x \in L$, then $\Delta\left(C_0^S(x, r); C_1^S(x, r)\right) \geq 0.99$.*
- *If $x \notin L$, then $\Delta\left(C_0^S(x, r); C_1^S(x, r)\right) \leq 0.01$.*

Moreover, each of these circuits is of size $O(t^6)$ and makes $O(t^5)$ oracle calls, where t is the sum of the length of the transcript output by S and the number of bits of randomness it uses.

Lemma 12 follows immediately from the above by instantiating the simulator oracle with the simulator of the SZK′ protocol. For the sake of completeness, we unravel Lemma 13 and walk through the entire proof of Lemma 12 in Appendix B.

Acknowledgements

We thank Pritish Kamath and Hoeteck Wee for helpful discussions, and Mika Göös for helpful pointers. The first author was supported by the European Union's Horizon 2020 Programme (ERC-StG-2014-2020) under grant agreement no. 639813 ERC-CLC, by an ICRC grant and by the Check Point Institute for Information Security. The second author was supported in part by NSF Grant CNS-1350619, and by the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Research Office under contracts W911NF-15-C-0226 and W911NF-15-C-0236.

References

[AA18] Benny Applebaum and Barak Arkis. On the power of amortization in secret sharing: d-uniform secret sharing and CDS with constant information rate. In Amos Beimel and Stefan Dziembowski, editors, *Theory of Cryptography - 16th International Conference, TCC 2018, Panaji, India, November 11-14, 2018, Proceedings, Part I*, volume 11239 of *Lecture Notes in Computer Science*, pages 317–344. Springer, 2018.

[AARV17] Benny Applebaum, Barak Arkis, Pavel Raykov, and Prashant Nalini Vasudevan. Conditional disclosure of secrets: Amplification, closure, amortization, lower-bounds, and separations. In Katz and Shacham [KS17], pages 727–757.

[AHMS18] Benny Applebaum, Thomas Holenstein, Manoj Mishra, and Ofer Shayevitz. The communication complexity of private simultaneous messages, revisited. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part II*, volume 10821 of *Lecture Notes in Computer Science*, pages 261–286. Springer, 2018.

[AIR01] William Aiello, Yuval Ishai, and Omer Reingold. Priced oblivious transfer: How to sell digital goods. In Birgit Pfitzmann, editor, *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, volume 2045 of *Lecture Notes in Computer Science*, pages 119–135. Springer, 2001.

[AR16] Benny Applebaum and Pavel Raykov. From private simultaneous messages to zero-information arthur-merlin protocols and back. In *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part II*, pages 65–82, 2016.

[Att14] Nuttapong Attrapadung. Dual system encryption via doubly selective security: Framework, fully secure functional encryption for regular languages, and more. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 557–577. Springer, 2014.

[Bab85] László Babai. Trading group theory for randomness. In Sedgewick [Sed85], pages 421–429.

[BCH+17] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 708–719. IEEE Computer Society, 2017.

[BD91] Ernest F. Brickell and Daniel M. Davenport. On the classification of ideal secret sharing schemes. *J. Cryptology*, 4(2):123–134, 1991.

[BDRV18] Itay Berman, Akshay Degwekar, Ron D. Rothblum, and Prashant Nalini Vasudevan. From laconic zero-knowledge to public-key cryptography. *IACR Cryptology ePrint Archive*, 2018:548, 2018.

[BFMP17] Amos Beimel, Oriol Farrs, Yuval Mintz, and Naty Peter. Linear secret-sharing schemes for forbidden graph access structures. To appear in TCC 2017, 2017.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347. IEEE Computer Society, 1986.

[BIKK14] Amos Beimel, Yuval Ishai, Ranjit Kumaresan, and Eyal Kushilevitz. On the cryptographic complexity of the worst functions. In Lindell [Lin14], pages 317–342.

[BKN18] Amos Beimel, Eyal Kushilevitz, and Pnina Nissim. The complexity of multiparty psm protocols and related models. To appear in Eurocrypt 2018, 2018. https://eprint.iacr.org/2018/148.

[BM88] László Babai and Shlomo Moran. Arthur-merlin games: A randomized proof system, and a hierarchy of complexity classes. *J. Comput. Syst. Sci.*, 36(2):254–276, 1988.

[BP18] Amos Beimel and Naty Peter. Optimal linear multiparty conditional disclosure of secrets protocols. Cryptology ePrint Archive, Report 2018/441, 2018. https://eprint.iacr.org/2018/441.

[CSGV93] Renato M. Capocelli, Alfredo De Santis, Luisa Gargano, and Ugo Vaccaro. On the size of shares for secret sharing schemes. *J. Cryptology*, 6(3):157–167, 1993.

[DI06] Bella Dubrov and Yuval Ishai. On the randomness complexity of efficient sampling. In Jon M. Kleinberg, editor, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 711–720. ACM, 2006.

[Dod12] Yevgeniy Dodis. Shannon impossibility, revisited. In Adam D. Smith, editor, *Information Theoretic Security - 6th International Conference, ICITS 2012, Montreal, QC, Canada, August 15-17, 2012. Proceedings*, volume 7412 of *Lecture Notes in Computer Science*, pages 100–110. Springer, 2012.

[FKN94] Uriel Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation (extended abstract). In Frank Thomson Leighton and Michael T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 554–563. ACM, 1994.

[FN93] Amos Fiat and Moni Naor. Broadcast encryption. In Douglas R. Stinson, editor, *Advances in Cryptology - CRYPTO '93, 13th Annual International Cryptology Conference, Santa Barbara, California, USA, August 22-26, 1993, Proceedings*, volume 773 of *Lecture Notes in Computer Science*, pages 480–491. Springer, 1993.

[GIKM00] Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. Protecting data privacy in private information retrieval schemes. *J. Comput. Syst. Sci.*, 60(3):592–629, 2000.

[GKW15] Romain Gay, Iordanis Kerenidis, and Hoeteck Wee. Communication complexity of conditional disclosure of secrets and attribute-based encryption. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part II*, volume 9216 of *Lecture Notes in Computer Science*, pages 485–502. Springer, 2015.

[GMR85] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof-systems (extended abstract). In Sedgewick [Sed85], pages 291–304.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity for all languages in NP have zero-knowledge proof systems. *J. ACM*, 38(3):691–729, 1991.

[GPW15] 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.

[GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Computational Complexity*, 27(2):245–304, 2018.

[GS89] Shafi Goldwasser and Michael Sipser. Private coins versus public coins in interactive proof systems. *Advances in Computing Research*, 5:73–90, 1989.

[HR11] Thomas Holenstein and Renato Renner. On the randomness of independent experiments. *IEEE Transactions on Information Theory*, 57(4):1865–1871, 2011.

[IK04] Yuval Ishai and Eyal Kushilevitz. On the hardness of information-theoretic multiparty computation. In Christian Cachin and Jan Camenisch, editors, *Advances in Cryptology - EUROCRYPT 2004, International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004, Proceedings*, volume 3027 of *Lecture Notes in Computer Science*, pages 439–455. Springer, 2004.

[IKP10] Yuval Ishai, Eyal Kushilevitz, and Anat Paskin. Secure multiparty computation with minimal interaction. In Tal Rabin, editor, *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, volume 6223 of *Lecture Notes in Computer Science*, pages 577–594. Springer, 2010.

[IW14] Yuval Ishai and Hoeteck Wee. Partial garbling schemes and their applications. In 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*, pages 650–662. Springer, 2014.

[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.

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

[KNR99] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[KS17] Jonathan Katz and Hovav Shacham, editors. *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part I*, volume 10401 of *Lecture Notes in Computer Science*. Springer, 2017.

[Lin14] Yehuda Lindell, editor. *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, volume 8349 of *Lecture Notes in Computer Science*. Springer, 2014.

[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.

[LV18] Tianren Liu and Vinod Vaikuntanathan. Breaking the circuit-size barrier in secret sharing. To appear in STOC2018, 2018. https://eprint.iacr.org/2018/333.

[LVW17a] Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Conditional disclosure of secrets via non-linear reconstruction. In Katz and Shacham [KS17], pages 758–790.

[LVW17b] Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Towards breaking the exponential barrier for general secret sharing. To appear in Eurocrypt 2018, 2017. https://eprint.iacr.org/2017/1062.

[MNSW98] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, 1991.

[RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In Bimal K. Roy, editor, *Advances in Cryptology - ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4-8, 2005, Proceedings*, volume 3788 of *Lecture Notes in Computer Science*, pages 199–216. Springer, 2005.

[Sed85] Robert Sedgewick, editor. *Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA*. ACM, 1985.

[Sha49] Claude E. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28:656–715, 1949.

[SS97] Hung-Min Sun and Shiuh-Pyng Shieh. Secret sharing in graph-based prohibited structures. In *Proceedings IEEE INFOCOM '97, The Conference on Computer Communications, Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Driving the Information Revolution, Kobe, Japan, April 7-12, 1997*, pages 718–724. IEEE, 1997.

[SV03] Amit Sahai and Salil P. Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003.

[Vad99] Salil Pravin Vadhan. *A study of statistical zero-knowledge proofs*. PhD thesis, Massachusetts Institute of Technology, 1999.

[Wee14] Hoeteck Wee. Dual system encryption via predicate encodings. In Lindell [Lin14], pages 616–637.

[Wee18] Hoteck Wee. Personal Communication, 2018.

A A ppCDS protocol for NEQ_n

Lemma 14. *For every constant c, there is a* CDS *protocol for NEQ_n with perfect privacy and correctness error of 2^{-c} that supports secrets of length c bits with communication complexity of $4c$ bits. The randomness complexity is $O(n+c)$ and it can be reduced to $O(\log n + c + \log(1/\epsilon))$, for an arbitrary $\epsilon > 0$, at the expense of increasing the correctness error to $\delta = 2^{-c} + \epsilon$. Moreover, the protocol has a ZPP-type decoder, i.e., (on 1-inputs) the decoder never errs but may output a special "I don't know" symbol with probability at most δ.*

Proof. The protocol is defined as follows.

- Shared Randomness: The parties share a pairwise independent hash function h from n-bits to c-bits, and a random c-bit string r.
- Alice sends $(h(x), z \cdot h(x) + r)$ and Bob sends $(h(y), z \cdot h(y) + r)$, where multiplication is over the field $\mathbb{F}_{2^c}$.
- Decoding: Given (a, b) from Alice and (c, d) from Bob, if $(a - c) = 0$ output "I don't know", else, output $b - d/(a - c)$.

By pairwise independent, when $x \neq y$ the probability of error (i..e, $a - c = 0$) is at most 2^{-c}. Privacy follows by noting that when $x = y$, the pair $(a, b) = (c, d)$ is distributed uniformly over $\mathbb{F}_{2^c} \times \mathbb{F}_{2^c}$.

To reduce the randomness complexity replace h with ϵ-almost 2-wise independent hash function h. The error probability grows by ϵ. To keep perfect privacy, make sure that h is 1-wise independent (i.e., for every x, the value $h(x)$ is uniform). It is possible to sample such a hash function using $O(\log n + c + \log(1/\epsilon))$ random bits. □

B Completeness for SZK′ – Proof of Lemma 12

Lemma 12 is proven along the lines of the construction showing a reduction to the "Statistical Difference" problem from any problem with honest-verifier SZK proofs in the computational setting. We follow rather closely such a reduction presented in [Vad99, Chap. 3].

Proof of Lemma 12. We are given an SZK′ protocol Π for the function f with complexity t, completeness error c, soundness error s, and simulation error η (which is negligible in n). Let the various quantities c_V, c_M, c_S, ρ_V, and ρ_S be as in Definition 6 – note that all of these are at most t which, in turn, is at most n. Let the number of rounds in the protocol be v – this means the number of messages (sent between Alice and Merlin) is $2v$. For notational convenience, we assume, without loss of generality, that the final message is sent by Alice and consists of all the random bits she shared with Bob (this is the $(2v-1)^{\text{th}}$ message, and the $2v^{\text{th}}$ message is empty).

Denote by S the simulator for this SZK′ protocol Π. Recall that S (written as Π_S in Definition 6) is a public-coin two-party protocol that simulates the views of Alice and Bob – that is, it has two parties that output $\widehat{V}_A$ and $\widehat{V}_B$, the joint distribution of which is statistically close to that of the actual views of Alice and Bob if $f(x,y)=1$. Throughout this proof, we will only use the part of the simulated transcript corresponding to the communication between Alice and Merlin, and the shared randomness between Alice and Bob, and we simply ignore the rest of the simulator's output.

Fix some input (x,y). Denote by S_i (for $i \in [2v]$) the protocol that runs the simulator protocol S on (x,y) and outputs just the first i simulated messages between Alice and Merlin. We define the following two intermediate protocols:

- The protocol Π'_0 runs one instance each of $S_2, S_4, \ldots, S_{2v}$, and outputs the concatenation of their outputs.
- The protocol Π'_1 works as follows:
  1. Run one instance each of $S_1, S_3, \ldots, S_{2v-1}$, and output their outputs.
  2. Output $(\rho_V - 4)$ uniformly random bits.
  3. Run S independently $100 \log c_M$ times, and if a majority of the resulting transcripts are rejecting, additionally output $(vc_M + 2)$ uniformly random bits.

So far, the outputs of either of Π'_0 and Π'_1 are $O(t^2)$ bits long (as the output of any S_i is at most $O(t)$ bits), and the same is true of their communication and randomness complexities. We first show the following differences in the behaviour of the entropies of the output distributions of the above protocols depending on the value of $f(x,y)$.

Proposition 3. *If $f(x,y)=1$, then $H(\Pi'_0) \geq H(\Pi'_1)+1$. And if $f(x,y)=0$, then $H(\Pi'_1) \geq H(\Pi'_0)+1$.*

We prove Proposition 3 in Section B.1. The next step is to extract this purported difference in the entropies of the protocols. Let the output lengths of Π'_0 and Π'_1 be m, and say they each use ρ bits of shared randomness. (Recall that both $m = O(t^2)$ and $\rho = O(t^2)$.) Let Π''_0 (resply. Π''_1) denote the protocol obtained by repeating Π'_0 (respectively, Π'_1) $q = \Theta(\rho^2)$ times, where the exact value of q will be determined later. Pick a family $G = \left\{ g : \{0,1\}^{q(m+\rho)} \to \{0,1\}^{qm} \right\}$ of 2-universal hash functions that can be sampled with $O(q(m+\rho))$ bits of randomness (e.g., by using Toeplitz

matrices). For $r \in \{0,1\}^{q\rho}$, denote by $\Pi''_0(r)$ the protocol Π_0 run with r as the shared randomness.

Our final protocols are defined as follow:

- Π_0:
  1. Pick an $r \xleftarrow{R} \{0,1\}^{q\rho}$ and run $\Pi''_0(r)$ to get its output π_0.
  2. Run Π''_1 to get its output π_1.
  3. Pick a hash function $g \xleftarrow{R} G$ and output $(\pi_0, g, g(r, \pi_1))$.
- Π_1:
  1. Run Π''_0 to get its output π_0.
  2. Pick a hash function $g \xleftarrow{R} G$ and an $r' \xleftarrow{R} \{0,1\}^{qm}$ and output (π_0, g, r').

It can be verified that the communication and randomness complexities and the output lengths of the above protocols are all $O(q(m+\rho)) = O(t^6)$. The following proposition (whose proof is postponed to Section B.2 together with Proposition 3, now completes the proof of the lemma.

Proposition 4. *If* $H(\Pi'_0) > H(\Pi'_1) + 1$, *then* $\Delta(\Pi_0(x,y); \Pi_1(x,y)) \geq 0.99$. *And if* $H(\Pi'_1) > H(\Pi'_0) + 1$, *then* $\Delta(\Pi_0(x,y); \Pi_1(x,y)) \leq 0.01$.

□

B.1 Proof of Proposition 3

Recall that we have fixed an input (x, y). First, we write the entropies of the outputs of the two protocols as follows:

$$H(\Pi'_0) = \sum_{i=1}^{v} H(S_{2i})$$
$$H(\Pi'_1) = \sum_{i=1}^{v} H(S_{2i-1}) + (\rho_V - 4) + H(P)$$

where by P we denote the random variable that is (vc_M+2) random bits if the majority of $100 \log c_M$ runs of S on (x, y) produce rejecting transcripts, and is empty otherwise.

The difference between these entropies is:

$$H(\Pi'_1) - H(\Pi'_0) = \rho_V - \sum_{i=1}^{v} (H(S_{2i}) - H(S_{2i-1})) + H(P) - 4 \tag{10}$$

To analyze this difference, we make use of the following corollaries of Lemmas 3.3.8, 3.3.11, and 3.3.12 from [Vad99]. While these were originally stated in the computational setting, it may be seen that these lemmas only concern the properties of the simulator's output in an interactive protocol, and are oblivious to the model of computation of the verifier as long as all of its randomness is accounted for in the following statements.

Proposition 5. *Let $\eta_{x,y}$ be the statistical distance between the distribution of the output of S and that of the actual transcript of Π on the input (x, y). Then,*

$$\rho_V - \sum_{i=1}^{v} \left(H(S_{2i}) - H(S_{2i-1})\right) \leq 2v \left(c_M \eta_{x,y} + h(\eta_{x,y})\right)$$

where h is the binary Shannon entropy.

Proposition 6. *On input (x, y), let p be the probability that S outputs an accepting transcript, and q be the maximum, over all strategies of Merlin, that Alice and Bob accept. Then, if $p \geq q$,*

$$d(p||q) \leq \rho_V - \sum_{i=1}^{v} \left(H(S_{2i}) - H(S_{2i-1})\right)$$

where d is the binary KL-divergence.

Suppose $f(x, y) = 1$. We first bound the entropy of the variable P. Let R denote the event that a majority of the $100 \log c_M$ runs of S (used in sampling P) produce rejecting transcripts. As long as the correctness and simulation errors are small enough constants, by the Chernoff bound, $\Pr[R]$ is less than, say, $1/5c_M^3$. And unless R happens, P is empty and has no entropy. As R is also completely determined by P, we may bound $H(P)$ as:

$$H(P) = H(P, R) = H(R) + H(P|R) \leq h\left(\frac{1}{5c_M^3}\right) + \frac{1}{5c_M^3} \cdot (c_M v + 2) + \left(1 - \frac{1}{5c_M^3}\right) \cdot 0 < 2 \quad (11)$$

Putting together Eqs. (10) and (11) and Proposition 5, we have:

$$H(\Pi_1') - H(\Pi_0') < 2v \left(c_M \eta_{x,y} + h(\eta_{x,y})\right) - 2$$

which, if the simulation error is smaller than, say, $1/5c_M v$, is smaller than (-1). This proves the first part of the proposition.

Next, suppose $f(x, y) = 0$. If p – the probability that S outputs accepting transcripts – is at most $1/4$ then, again by the Chernoff bound, the probability that P is empty is at most $1/5c_M^3$. Thus, $H(P)$ is at least $(1 - 1/5c_M^3) \cdot (c_M v + 2) \geq (c_M v + 1) \geq H(\Pi_0') + 1$.

Suppose p is larger than $1/4$. Note that, by the soundness of the SZK$'$ protocol, q – the maximum probability that Alice accepts for any Merlin strategy – is at most s, the soundness error. Thus, $d(p||q)$ is at least $d(0.25||s)$, which in turn is at least 5 if s is a small enough constant. Now, using Eq. (10) and Proposition 6 and the fact that $H(P) \geq 0$, we have:

$$H(\Pi_1') - H(\Pi_0') \geq d(p||q) - 4 \geq 1$$

Thus, if $f(x, y) = 0$, in either case, $H(\Pi_1') \geq H(\Pi_0') + 1$. This completes the proof of Proposition 3. □

B.2 Proof of Proposition 4

Our arguments in this proof will be based on the entropy that the random string $r \in \{0, 1\}^\rho$ and the output of Π_1' have given the output of $\Pi_0'(r)$. Denoting the random variable corresponding to

r by R, this quantity is written as follows:

$$\begin{aligned} H(R, \Pi_1' | \Pi_0'(R)) &= H(\Pi_1') + H(R | \Pi_0'(R)) \\ &= H(\Pi_1') + H(R, \Pi_0'(R)) - H(\Pi_0'(R)) \\ &= H(\Pi_1') + H(R) - H(\Pi_0') \\ &= \rho + H(\Pi_1') - H(\Pi_0') \end{aligned}$$

where the second-to-last equality follows from the fact that r is the only source of randomness for the protocol Π_0'.

The statistical distance between the outputs of Π_0 and Π_1 is that between $(\Pi_0'^q(R^q), G, G(R^q, \Pi_1^q))$ and $(\Pi_0'^q(R^q), G, U)$, where G is the hash function chosen by the protocol and U is the uniform distribution over strings of the appropriate length. $\Pi_0'^q$, $\Pi_1'^q$ and R^q denote independent q-fold repetitions of Π_0', Π_1' and R, respectively.

Suppose that $H(\Pi_1') \geq H(\Pi_0') + 1$. We wish to show that the hash of $(R^q, \Pi_1'^q)$ is close to uniform. We do this using the following appropriate leftover hash lemma, which follows from results in [HR11, RW05] (see Theorem 4.8 and Lemma 4.9 in [BDRV18] for the necessary summary).

Lemma 15 (Leftover Hash Lemma with Flattening)**.** *For some natural numbers q, n and m, let $G = \{g : \{0,1\}^{qn} \to \{0,1\}^m\}$ be a family of universal hash functions. For any random variables (X, Y) where X is distributed over $\{0,1\}^n$, let (X^q, Y^q) denote the q-fold repetition of (X, Y). Then, for any $\delta \geq 0$,*

$$\Delta\left((Y^q, G, G(X^q, Y^q)); (Y^q, G, U)\right) \leq 2^{-\frac{q\delta^2}{3n^2}} + \frac{1}{2} \cdot \sqrt{2^{-q(H(X|Y)-\delta)} \cdot 2^m}$$

where U is uniform over $\{0,1\}^{qm}$.

In our case, the variable X corresponds to (R, Π_1'), Y corresponds to $\Pi_0'(R)$, and the conditional entropy $H(X|Y)$ is at least $(\rho+1)$. We pick δ to be a small constant (say 0.01), and set q to be dt^4 for a constant d that is large enough (say 10^5) for the following bounds. The length of the output of the hash function here is $q\rho$, and applying Lemma 15, we have the following:

$$\begin{aligned} \Delta\left(\Pi_0; \Pi_1\right) &= \Delta\left((\Pi_0'^q(R^q), G, G(R^q, \Pi_1^q)); (\Pi_0'^q(R^q), G, U)\right) \\ &\leq 2^{-\frac{10^5 t^4 \cdot \delta^2}{3 \cdot O(t^2)^2}} + \frac{1}{2} \cdot \sqrt{2^{-q(\rho+1-\delta-\rho))}} \\ &\leq 0.01 \end{aligned}$$

This proves the first part of the proposition.

Next, suppose that $H(\Pi_0') \geq H(\Pi_1') + 1$. In this case, $H(R, \Pi_1' | \Pi_0'(R))$ is at most $(\rho - 1)$. In this case, we wish to show that the above distributions are far. We make use of the following lemma that is again to be found in [HR11] (as Theorem 2).

Lemma 16. *Let (X, Y) be random variables where X is distributed over $\{0,1\}^n$, and let (X^q, Y^q) denote their q-fold repetition. For any $\delta \in [0, n]$, and any y in the support of Y^q, there exists a set $T_y^\delta \subseteq \{0,1\}^{qn}$ that is of size at most $2^{q(H(X|Y)+\delta)}$ such that:*

$$\Pr_{(x,y) \leftarrow (X^q, Y^q)}\left[x \in T_y^\delta\right] \geq 1 - 2 \cdot 2^{-\frac{q\delta^2}{3n^2}}$$

Taking X, Y and δ to be as earlier, Lemma 16 implies that for any π_0 in the support of $\Pi_0'^q(R)^q$, there is a set $T^\delta_{\pi_0}$ of size at most $2^{q(\rho-1+\delta)}$ such that:

$$\Pr_{r\leftarrow\{0,1\}^{q\rho},\pi_0\leftarrow\Pi_0'^q(r),\pi_1\leftarrow\Pi_1'^q}\left[(r,\pi_1)\in T^\delta_{\pi_0}\right]\geq 1-2\cdot 2^{-\frac{10^5t^4\delta^2}{3\cdot O(t^2)^2}}$$

For a hash function g, let $g(T^\delta_{\pi_0})$ denote the set of images of elements of $T^\delta_{\pi_0}$; note that $\left|g(T^\delta_{\pi_0})\right|\leq \left|T^\delta_{\pi_0}\right|\leq 2^{q(\rho-1+\delta)}$. The statistical distance between Π_0 and Π_1 can be written as follows:

$$\begin{aligned}
\Delta\left(\Pi_0;\Pi_1\right) &= \Delta\left((\Pi_0'^q(R^q),G,G(R^q,\Pi_1^q));(\Pi_0'^q(R^q),G,U)\right)\\
&\geq \Pr_{r\leftarrow\{0,1\}^{q\rho},\pi_0\leftarrow\Pi_0'^q(r),\pi_1\leftarrow\Pi_1'^q,g\leftarrow G}\left[g(r,\pi_1)\in g(T^\delta_{\pi_0})\right]-\Pr_{\pi_0\leftarrow\Pi_0'^q,g\leftarrow G,r'\leftarrow\{0,1\}^{q\rho}}\left[r'\in g(T^\delta_{\pi_0})\right]\\
&\geq 1-2\cdot 2^{-\frac{10^5t^4\delta^2}{3\cdot O(t^2)^2}}-\frac{2^{q(\rho-1-\delta)}}{2^{q\rho}}\\
&\geq 0.99
\end{aligned}$$

This proves Proposition 4. □