Communication with Contextual Uncertainty

Citation: Ghazi, Badih, et al. "Communication with Contextual Uncertainty." Computational Complexity, vol. 27, no. 3, Sept. 2018, pp. 463–509.

As Published: https://doi.org/10.1007/s00037-017-0161-3

Publisher: Springer International Publishing

Persistent URL: http://hdl.handle.net/1721.1/118914

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

COMMUNICATION WITH CONTEXTUAL UNCERTAINTY

Badih Ghazi, Ilan Komargodski,
Pravesh K. Kothari, and Madhu Sudan

July 31, 2017

Abstract. We introduce a simple model illustrating the utility of context in compressing communication and the challenge posed by uncertainty of knowledge of context. We consider a variant of *distributional* communication complexity where Alice gets some information $X \in \{0,1\}^n$ and Bob gets $Y \in \{0,1\}^n$, where (X,Y) is drawn from a known distribution, and Bob wishes to compute some function $g(X,Y)$ or some close approximation to it (i.e., the output is $g(X,Y)$ with high probability over (X,Y)). In our variant, Alice does not know g, but only knows some function f which is a very close approximation to g. Thus, the function being computed forms the context for the communication. It is an enormous implicit input, potentially described by a truth table of size 2^n. Imprecise knowledge of this function models the (mild) uncertainty in this context.

We show that uncertainty can lead to a huge cost in communication. Specifically, we construct a distribution μ over $(X,Y) \in \{0,1\}^n \times \{0,1\}^n$ and a class of function pairs (f,g) which are very close (i.e., disagree with $o(1)$ probability when (X,Y) are sampled according to μ), for which the communication complexity of f or g in the standard setting is *one bit* whereas the (two-way) communication complexity in the uncertain setting is at least $\Omega(\sqrt{n})$ bits even when allowing a constant probability of error.

It turns out that this blow-up in communication complexity can be attributed in part to the mutual information between X and Y. In particular, we give an efficient protocol for communication under contextual uncertainty that incurs only a small blow-up in communication if this mutual information is small. Namely, we show that if g has a communication protocol with complexity k in the standard setting and the mutual information between X and Y is I, then g has a one-way communication protocol with complexity $O((1+I)\cdot 2^k)$ in the uncertain setting. This result is an immediate corollary of an even stronger result which shows that if g has one-way communication complexity k, then it has one-way uncertain communication complexity at most $O((1+I)\cdot k)$. In the particular case where the input distribution is a product distribution (and so $I = 0$), the protocol in the uncertain setting only incurs a *constant factor* blow-up in one-way communication and error.

Keywords. Reliable Communication, Context, Uncertainty, Communication Complexity.

Subject classification. 68Q01

1. Introduction

Most forms of communication involve communicating players that share a large common *context* which they use to compress communication. In natural settings, the context may include understanding of language, and knowledge of the environment and laws. In designed (computer-to-computer) settings, the context includes "commonsense knowledge" as well as the knowledge of the operating system, communication protocols, and encoding/decoding mechanisms. This notion of "context" held by intelligent systems plays a fundamental role both in the classical study of artificial intelligence and in the emerging area of "conversational artificial intelligence" which underlies intelligent virtual assistants such as Siri, Google Assistant and Amazon Alexa. Remarkably, especially in the natural setting, context can seemingly be used to *compress communication*, even when it is *enormous* and *not shared perfectly.* This ability to communicate despite a major source of uncertainty has led to a series of works attempting to model various forms of communication amid uncertainty, starting with Juba & Sudan (2008), Goldreich *et al.* (2012) followed by Juba *et al.* (2011), Juba & Sudan (2011), Juba & Williams (2013), Haramaty & Sudan (2014) and Canonne *et al.* (2015). The latter works implicitly give examples of context which share the three features mentioned above — the context helps compress communication, even though it is large and imperfectly shared. This current work is the first in this series to explicitly highlight this notion and features of context. It does so while studying a theme that is new to this series of works, namely a *functional* notion of uncertainty. We start by describing the setup for our model and then present our model and results below, before contrasting them with some of the previous works.

Our model builds on the classical setup of communication complexity due to Yao (1979). The classical model considers two interacting players Alice and Bob each possessing some private information X and Y, with X known only to Alice and Y to Bob. In the general setting, both players can send messages to each other, while in the one-way setting only Alice sends a message to Bob. They (specifically Bob, in the one-way setting) wish to compute some joint (Boolean-valued) function $g(X, Y)$ and would like to do so while communicating the minimum possible number of bits. In this work we use the function g to model (part of) the *context* of the communication. Indeed it satisfies some of the essential characteristics of context: It is potentially "enormous". For example, if g were represented as a truth table of values and if X and Y are n-bit strings, then the representation of g would be 2^{2n} bits long. And indeed knowledge of this context can compress communication significantly: Consider the trivial collection of examples where $g(X, Y) = g'(X)$, i.e., g is simply a function of X. In this case, knowledge of the context (i.e., the function g') compresses communication to just one bit. In contrast, if Alice does not know the context, her other option is to send X to Bob which requires n bits of communication. This intuitive explanation can be formalized using the well known Indexing problem (Kushilevitz & Nisan 1997, Example 4.19) which essentially considers the setting where Alice has an "index" (corresponding to X) and Bob has a vector (corresponding to the truth table of g') and their goal is to compute the indexed value of the vector (i.e., computing $g'(X)$ in our correspondence). Standard lower bounds for Indexing (see Theorem 6.2) imply that $\Omega(n)$ bits of communication are needed to compute $g'(X)$.

In this work, we focus on the case where the context is imperfectly shared. Specifically, we consider the setting where Bob knows the function g and Alice only knows some (close) approximation f to g (with f not being known to Bob).[1] This leads to the questions: How should Alice and Bob interact while accounting for this uncertainty about their shared context? What quantitative effect does this uncertainty have on the communication complexity of computing $g(X, Y)$?

It is clear that if $X \in \{0,1\}^n$, then n bits of communication suffice — Alice can simply ignore f and send X to Bob. We wish to consider settings that improve on this. To do so, a necessary condition is that g must have low communication complexity in the standard model. However, this necessary condition does not seem to be sufficient to compute g correctly on every input — since Alice only has an approximation f to g. (In Theorem 6.1 in Section 6, we formally prove this assertion by giving a function g with low communication complexity, but where computing $g(X, Y)$ takes $\Omega(n)$ bits in the worst-case if Alice is only given an approximation f to g.) Thus, we settle for a weaker goal, namely, that of computing g correctly only on most inputs. This puts us in a distributional communication complexity setting. A necessary condition now is that g must have a low-error low-communication protocol in the standard (distributional complexity) setting. The question is then: can g be computed with low error and low communication when Alice only knows an approximation f to g (with f being unknown to Bob)? Formalizing this model still requires some work and we do so next.

1.1. Uncertain-Communication Complexity. We first recall the standard model of communication complexity, in the distributional setting. For contrast with our model, we sometimes refer to

[1] We note that the assumption that Bob knows the precise function g to be computed is not a restrictive assumption but merely a convention that is consistent with our earlier suggestion that Bob wishes to compute the function g. We could have equally well asserted that the function to be computed is f (and so only Alice knows the function to be computed), or picked a neutral setting saying the function to be computed is h which is very close to both f and g. The definitions don't really make a significant difference to the communication problem since any protocol Π that computes a function close to g is also close to f or h, and hence all versions have the same communication complexity with small changes in error.

this as the model of "certain-communication".

Let Π denote a communication protocol that specifies how Alice with input X and Bob with input Y interact. I.e., Π includes functions that specify: (1) given a history of transmissions, if the communication should continue and if so which one of Alice or Bob should speak next, (2) given a history of transmissions and the speaker's private input (one of X or Y), what the speaker's next message should be; and (3) what the output of the protocol is when the communication stops. We let $\Pi(X, Y)$ denote the output of the protocol. Note that protocols may involve private or public (shared) randomness and if so $\Pi(X, Y)$ is a random variable. We let the communication complexity of Π, denoted $\mathsf{CC}(\Pi)$, be the maximum number of bits exchanged by a protocol, maximized over all inputs and all (private or public) random coins. We say that a protocol is one-way if all communication comes from one speaker, typically from Alice to Bob.

In order to describe what it means for a protocol to compute a close approximation to a given function, we describe our distance measure on functions. For a distribution μ supported on $\{0,1\}^n \times \{0,1\}^n$, we let $\delta_\mu(f, g)$ denote the probability that f and g differ on a random input drawn from μ, i.e., $\delta_\mu(f, g) := \Pr_{(X,Y)\sim\mu}[f(X, Y) \neq g(X, Y)]$. If exactly one of f or g is probabilistic then we include the randomness in the probability space.[2] We say f and g are *δ-close* (with respect to μ) if $\delta_\mu(f, g) \leq \delta$.

For parameter $\epsilon > 0$, the *distributional communication complexity* (in the setting of certain-communication) of a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ over a distribution μ, denoted $\mathsf{CC}^\mu_\epsilon(f)$, is the minimum communication complexity of a protocol, minimized over all protocols that compute a function that is ϵ-close to f, i.e.,

$$\mathsf{CC}^\mu_\epsilon(f) \triangleq \min_{\Pi:\delta_\mu(f,\Pi)\leq\epsilon} \{\mathsf{CC}(\Pi)\}.$$

Similarly, $\mathsf{owCC}^\mu_\epsilon(f)$ denotes the corresponding *one-way* communication complexity of f.

[2] The correct generalization to the case when both f and g are probabilistic is to take the expectation of the statistical distance (also known as total variation distance) between $f(X, Y)$ and $g(X, Y)$, but we won't need to consider this setting in this paper.

We now turn to defining the measure of complexity in the uncertain setting. Ideally, we would like to define the uncertain-communication complexity of computing some function g, given that Alice has some nearby function f. But this definition will not make sense as such! Even if Alice doesn't know g the protocol might itself "know" g. (Formally, the protocol Π that minimizes the communication complexity should not depend on g, but how does one forbid this?) So the right formulation is to define the communication complexity of an entire family $\mathcal{F}$ of pairs of functions, $\mathcal{F} \subseteq \{(f, g) \mid f, g\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}\}$. We define such a measure shortly, but before doing so, we discuss one more aspect of uncertain-communication.

One view of communication, applicable in the uncertain-communication setting as well as the certain-communication setting, is to make the function being computed an explicit input to the communicating players, say by presenting it as a truth table. Thus, in the setting of uncertain-communication, we may view the goal as computing the universal function $U\colon ((f, X), (g, Y)) \mapsto g(X, Y)$, where Alice's input is (f, X) and Bob's input is (g, Y). In the certain-communication setting we would further require $f = g$, but in our "uncertain" setting we don't. Instead, the functions f, g are adversarially chosen subject to the restrictions that they are close to each other (under some distribution μ on the inputs) and that g (and hence f) has a low-error low-communication protocol. The pair (X, Y) is drawn from the distribution μ (independent of the choice of f and g). The players both know μ in addition to their respective inputs.

Under this view, a protocol Π solving an uncertain communication problem is simply a protocol for the universal communication problem with its communication complexity being the maximum communication over all inputs and all possible randomness.[3] The ability to solve communication problems from $\mathcal{F}$ under distribution

[3] It might seem more appropriate to define the communication as the maximum only over pairs $(f, g) \in \mathcal{F}$ and (X, Y) in the support of μ, but this does not make a difference for optimal protocols. A protocol can be modified to stop after a given number of bits of communication, and the result would only affect the accuracy of the output, which thereby becomes the only parameter tied to the problem being solved.

μ is taken into account in defining the error of this protocol. For a protocol Π computing a (probabilistic) function $\Pi((f,X),(g,Y))$ we let $\Pi_{(f,g)}$ denote the function $\Pi_{(f,g)}(X,Y) = \Pi((f,X),g(Y))$.

Definition 1.1 (Uncertain-Communication Complexity). *The* uncertain-communication complexity *of a family $\mathcal{F}$ of pairs of functions (f,g) with respect to a distribution μ supported on $\{0,1\}^n \times \{0,1\}^n$, denoted $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F})$, is the minimum communication complexity of a protocol Π, minimized over all protocols Π such that for every $(f,g) \in \mathcal{F}$, the function $\Pi_{(f,g)}$ is ϵ-close to g. That is,*

$$\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \triangleq \min_{\{\Pi \mid \forall (f,g)\in\mathcal{F}\colon\, \delta_\mu(\Pi_{(f,g)},g)\le\epsilon\}} \{\mathsf{CC}(\Pi)\}.$$

Similarly, let $\mathsf{owCCU}^{\mu}_{\epsilon}(\mathcal{F})$ denote the one-way uncertain-communication complexity *of $\mathcal{F}$.*

Our goal is to study $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F})$ for a family $\mathcal{F}$, but this can be small only if the certain-communication complexity of functions in $\mathcal{F}$, specifically $\mathsf{CC}^{\mu}_{\epsilon}(\mathcal{F}) \triangleq \max_{(f,g)\in\mathcal{F}}\{\mathsf{CC}^{\mu}_{\epsilon}(g)\}$, is small. Furthermore, we want to model "mild" uncertainty (and not total uncertainty) between Alice and Bob. To this end, we define the distance of a family $\mathcal{F}$, denoted by $\delta_\mu(\mathcal{F})$, to be the maximum over all $(f,g) \in \mathcal{F}$ of $\delta_\mu(f,g)$.

In what follows we will study the behavior of $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F})$ as a function of $\mathsf{CC}^{\mu}_{\epsilon}(\mathcal{F})$ and $\delta_\mu(\mathcal{F})$ and especially focus on the case where $\delta_\mu(\mathcal{F}) \ll \epsilon$ (so the uncertainty between Alice and Bob is very small compared to the error they are willing to tolerate).

1.2. Results. For general distributions, it turns out we can prove a large gap between the uncertain-communication complexity of functions and their certain-communication complexity.

Theorem 1.2. *For every constant $\delta \in (0,1)$ and $\epsilon \in (0,0.5)$, there exist constants $\tau > 0$ and $c < \infty$ such that for all n, there is a distribution μ supported on $\{0,1\}^n \times \{0,1\}^n$ and a function class $\mathcal{F}$ satisfying $\delta_\mu(\mathcal{F}) \le \delta$ and $\mathsf{owCC}^{\mu}_{0}(\mathcal{F}) \le 1$ such that $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \ge \tau \cdot \sqrt{n} - c$.*

In particular, if δ is any positive constant (e.g., 0.001), then Theorem 1.2 asserts the existence of a distribution and a class of distance-δ functions for which the zero-error (one-way) communication complexity in the standard model is a single bit, but under contextual uncertainty, any two-way protocol (with an arbitrary number of rounds of interaction) having a noticeable advantage over random guessing requires $\Omega(\sqrt{n})$ bits of communication!

Given the strong negative result in Theorem 1.2, a natural question is to understand if there are any non-trivial settings where the uncertain-communication complexity is close to the certain-communication complexity. Surprisingly, it turns out that uncertain-communication complexity can *always* be upper-bounded in terms of the certain-communication complexity and the *mutual information* of the input distribution. Recall that for random variables (X, Y) drawn from some joint distribution, the mutual information between X and Y, denoted $I(X;Y)$, measures the amount of information that X has about Y (or vice versa)[4]. Theorem 1.3 shows that if μ is a distribution on which f and g are close and each has a *one-way* certain-communication complexity of at most k bits (for all $(f, g) \in \mathcal{F}$), then the family $\mathcal{F}$ has one-way uncertain communication complexity of at most $O(k \cdot (1 + I))$ bits with I being the mutual information of $(X, Y) \sim \mu$. We denote by $\mathsf{CC}^{\mu}_{\epsilon}(\mathcal{F})$ (resp. $\mathsf{owCC}^{\mu}_{\epsilon}(\mathcal{F})$) the maximum over all $(f, g) \in \mathcal{F}$ of $\mathsf{CC}^{\mu}_{\epsilon}(g)$ (resp. $\mathsf{owCC}^{\mu}_{\epsilon}(g)$).[5] We prove the following theorem.

Theorem 1.3. *There exists a constant c such that for all positive integers k and n and positive reals ϵ, δ, θ, for every distribution μ over $\{0,1\}^n \times \{0,1\}^n$, and every family $\mathcal{F}$ of pairs of Boolean functions satisfying $\delta_\mu(\mathcal{F}) \le \delta$ and $\mathsf{owCC}^{\mu}_{\epsilon}(\mathcal{F}) \le k$, it holds that*

$$\mathsf{owCCU}^{\mu}_{\epsilon+2\delta+\theta}(\mathcal{F}) \le c \cdot \frac{\left(k + \log\left(\frac{1}{\theta}\right)\right)}{\theta^2} \cdot \left(1 + \frac{I(X;Y)}{\theta^2}\right).$$

Using the well-known fact that the one-way certain-communication of any function is at most exponential in its two-way communi-

[4] Formally, given a distribution μ over a pair (X, Y) of random variables with marginals μ_X and μ_Y over X and Y respectively, the *mutual information* of X and Y is defined as $I(X;Y) \triangleq \mathbb{E}_{(a,b)\sim\mu}[\log(\frac{\mu(a,b)}{\mu_X(a)\mu_Y(b)})]$.

[5] Note that if $\delta_\mu(f, g) \le \delta$ and $\mathsf{CC}^{\mu}_{\epsilon}(g) \le k$, then $\mathsf{CC}^{\mu}_{\epsilon+\delta}(f) \le k$.

cation complexity (e.g., (Kushilevitz & Nisan 1997, Exercise 4.21)), Theorem 1.3 also immediately implies the next corollary.

Corollary 1.4. *There exists a constant c such that for all positive integers k and n and positive reals ϵ, δ, θ, for every distribution μ over $\{0,1\}^n \times \{0,1\}^n$, and every family $\mathcal{F}$ of pairs of Boolean functions satisfying $\delta_\mu(\mathcal{F}) \leq \delta$ and $\mathsf{CC}^\mu_\epsilon(\mathcal{F}) \leq k$, it holds that*

$$\mathsf{owCCU}^\mu_{\epsilon+2\delta+\theta}(\mathcal{F}) \leq c \cdot \frac{\left(2^k + \log\left(\frac{1}{\theta}\right)\right)}{\theta^2} \cdot \left(1 + \frac{I(X;Y)}{\theta^2}\right). \tag{1.5}$$

We stress that the exponential blow-up in (1.5) can be significantly smaller than the length n of the inputs (which is the trivial upper bound on the communication in the uncertain case). In the special case where μ is a product distribution, we have $I(X;Y) = 0$ and so we obtain the following particularly interesting corollary of Theorem 1.3.

Corollary 1.6. *There exists a constant c such that for all positive integers k and n and positive reals ϵ, δ, θ, for every* product *distribution μ over $\{0,1\}^n \times \{0,1\}^n$, and every family $\mathcal{F}$ of pairs of Boolean functions satisfying $\delta_\mu(\mathcal{F}) \leq \delta$ and $\mathsf{owCC}^\mu_\epsilon(\mathcal{F}) \leq k$, it holds that*

$$\mathsf{owCCU}^\mu_{\epsilon+2\delta+\theta}(\mathcal{F}) \leq c \cdot \frac{\left(k + \log\left(\frac{1}{\theta}\right)\right)}{\theta^2}.$$

In words, Corollary 1.6 says that for product distributions and for constant error probabilities, one-way uncertain-communication complexity is only a constant factor larger than the one-way certain-communication complexity.

One intuitive interpretation of the dependence on the mutual information $I(X;Y)$ in Theorem 1.3 is that the parties can make strong use of correlations among their inputs (i.e., between X and Y) in the standard setup. In contrast, they are unable to make such strong use in the uncertain case. Since the distribution μ in Theorem 1.2 has mutual information $\approx n$, Theorem 1.2 rules out improving the dependence on the mutual information in Theorem 1.3 to anything smaller than $\sqrt{I(X;Y)}$. It is a very interesting open question to determine the correct exponent of $I(X;Y)$ in Theorem 1.3.[6]

Finally, we point out that our results in Theorem 1.3, Corollary 1.4 and Corollary 1.6 achieve reliable communication despite uncertainty about the context *even when the uncertainty itself is hard to resolve.* To elaborate on this statement, note that one hope for achieving a low-communication protocol for g would be for Alice and Bob to first agree on some function h that is close to f and g, and then apply some low-communication protocol for this common function h. Such a protocol obviously exists if we assume g has a low-communication protocol, albeit with slightly higher error. (In particular, an ϵ-error protocol for g computes h with error $\epsilon + \delta_\mu(g,h)$.) This would be the "resolve the uncertainty first" approach.

We prove (in Theorem 1.7 below) that resolving the uncertainty is definitely an overkill and can lead to communication exponential in n (and much more so than the trivial protocol of sending x) and hence, this cannot be a way to prove Theorem 1.3. Namely, denote by $\textsc{Agree}_{\delta,\gamma}(\mathcal{F})$ the communication problem where Alice gets f and Bob gets g such that $(f,g) \in \mathcal{F}$ and their goal is for Alice to output h_A and Bob to output h_B such that $\delta(h_A, f), \delta(h_B, g) \leq \delta$ and $\Pr[h_A = h_B] \geq \gamma$, where the probability is over the internal randomness of the protocol. Even getting a positive agreement probability γ, leave alone getting agreement with high probability, turns out to require high communication as shown by the following theorem.

Theorem 1.7. *Let μ denote the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. For every $\delta, \delta' \in (0, 1/2)$ and $\gamma \in (0,1)$, there exist $\alpha > 0$ and $\beta < \infty$ and a family $\mathcal{F}$ of pairs of Boolean functions satisfying*

[6] We note that the upper bound of (roughly) $1 + I(X;Y)$ on the communication blow-up due to uncertainty in Theorem 1.3 holds for *every* function class and input distribution whereas the lower bound of $\sqrt{I}$ on this blow-up implied by Theorem 1.2 holds for *some* function class and input distribution. In particular, if the distribution μ only puts mass on points (X,Y) for which $X = Y$, then the mutual information can be very large while there would be no blow-up in communication due to uncertainty (since on such distributions no communication is needed to compute any function).

$\delta_\mu(\mathcal{F}) \leq \delta$ *and* $\mathsf{CC}_0^\mu(\mathcal{F}) = 0$, *such that*

$$\mathsf{CC}(\textsc{Agree}_{\delta',\gamma}(\mathcal{F})) \geq \alpha \cdot 2^n - \beta.$$

In particular, the theorem shows that there's a class of function pairs (f, g) where f and g are very close (say $\delta(f, g) \leq .01$) but agreeing on a function h with even a slight correlation with f and g (say $\delta(f, h), \delta(g, h) \leq .499$) incurs an exponentially high communication cost in n.

1.3. Prior Work. The first works to consider communication with uncertainty in a manner similar to this work were those of Juba & Sudan (2008) and Goldreich *et al.* (2012). Their goal was to model an extreme form of uncertainty, where Alice and Bob do not have any prior (known) commonality in context and indeed both come with their own "protocol" which tells them how to communicate. So communication is needed even to resolve this uncertainty. While their setting is thus very broad, the solutions they propose are less communication-efficient and typically involve resolving the uncertainty as a first step.

The later works Juba *et al.* (2011), Haramaty & Sudan (2014) and Canonne *et al.* (2015) tried to restrict the forms of uncertainty to see when it could lead to more efficient communication solutions. For instance, Juba *et al.* (2011) consider the compression problem when Alice and Bob do not completely agree on the prior. This introduces some uncertainty in the beliefs, and they provide fairly efficient solutions by restricting the uncertainty to a manageable form. Canonne *et al.* (2015) were the first to connect this stream of work to communication complexity, which seems to be a good umbrella to study the broader communication problems. The imperfectness they study is however restricted to the randomness shared by the communicating parties, and does not incorporate any other elements. (We point out that the setup of communication with imperfectly shared randomness had independently been studied by Bavarian *et al.* (2014) in the *simultaneous message passing model.* It was also further studied by Ghazi *et al.* (2016)). Canonne *et al.* (2015) suggest studying imperfect understanding of the function being computed as a general direction, though they do not suggest specific definitions, which we in particular do in this work.

1.4. Future Directions and Open Questions. In the current work, we introduce and study a simple model illustrating the role of context in communication and the challenge posed by uncertainty of knowledge of context. Several interesting questions are raised by this work.

On the technical side, it would be very interesting to determine the correct exponent of $I(X;Y)$ in Theorem 1.3. Theorem 1.3 and Theorem 1.2 imply that this exponent is between 1/2 and 1. Moreover, it would be nice to understand the needed dependence on k in the product $k \cdot I(X;Y)$ in Theorem 1.3. A very recent follow-up work Ghazi & Sudan (2017) obtained an improved lower bound of $\Omega(\sqrt{k} \cdot \sqrt{I(X;Y)})$, but the tight bound is still elusive. A related (but perhaps more challenging) question is whether the dependence on n can be improved from $\Omega(\sqrt{n})$ to $\Omega(n)$ in Theorem 1.7 (while keeping the communication in the standard case equal to $O(1)$). As discussed in Section 3, such an improvement would require a new construction of a family of pairs of Boolean functions and an input distribution since the $\Omega(\sqrt{n})$ lower bound is tight (up to a logarithmic factor) for the considered construction.

We point out that our protocol in Theorem 1.3 uses *shared randomness*. A very interesting question is whether shared randomness is actually *needed* for communication amid uncertainty. In fact, an ideal protocol for communication amid uncertainty would only use private randomness (or even no randomness at all). The follow-up work Ghazi & Sudan (2017) studied this question of the power of shared randomness in communication with uncertainty. In the case of *product distributions*, it was shown that *imperfectly shared randomness* (as studied in Bavarian *et al.* (2014), Canonne *et al.* (2015) and Ghazi *et al.* (2016)) is enough to incur not more than a constant factor blow-up in communication for constant error probabilities. Moreover, Ghazi & Sudan (2017) showed that the private-coin communication complexity with uncertainty is larger than the public-coin communication by a growing function of n. Nevertheless, the questions of determining the tight bounds for communication amid uncertainty in the deterministic, private-coin and imperfectly shared randomness setups remain open, and are likely to require fundamentally new ideas and constructions. For instance, can one prove a non-trivial upper bound – such as Theorem 1.3 – on the communication complexity of *deterministic* protocols?

It would also be extremely interesting to prove an analogue of Theorem 1.3 for two-way protocols. Our proof of Theorem 1.3 uses in particular the fact that any low-communication one-way protocol in the standard distributional communication model should have a canonical form: to compute $g(x, y)$, Alice tries to describe the entire function $g(x, \cdot)$ to Bob, and this does not create a huge overhead in communication. Coming up with a canonical form of two-way protocols that somehow changes gradually as we morph from g to f seems to be the essence of the challenge in extending Theorem 1.3 to the two-way setting. A concrete question here is whether the dependence on k in the special case of product distributions ((1.5) of Corollary 1.4 with $I(X; Y) = 0$) can be improved from 2^k to $\mathsf{poly}(k)$.

On the more conceptual side, arguably, the model considered in this work is realistic: communication has some goals in mind which we model by letting Bob be interested in a specific function of the joint information that Alice and Bob possess. Moreover, it is an arguably natural model to posit that the two are not in perfect synchronization about the function that Bob is interested in, but Alice can estimate the function in some sense. One aspect of our model that can be further refined is the specific notion of distance that quantifies the gap between Bob's function and Alice's estimate. In this work, we chose the Hamming distance which forms a good first starting point. We believe that it is interesting to propose and study other models of distance between functions that more accurately capture natural forms of uncertainty.

Finally, we wish to emphasize the mix of adversarial and probabilistic elements in our uncertainty model — the adversary picks (f, g) whereas the inputs (X, Y) are sampled from a distribution. We believe that richer mixtures of adversarial and probabilistic elements could lead to broader settings of modeling and coping with uncertainty — the probabilistic elements offer efficient possibilities that are often immediately ruled out by adversarial choices, whereas the adversarial elements prevent the probabilistic assumptions from being too precise.

Organization In Section 2, we carefully develop the uncertain communication complexity model after recalling the standard distributional communication complexity model. In Section 3, we prove the hardness of contextual agreement. In Section 4, we prove our main upper bound (Theorem 1.3). In Section 5, we prove our main lower bound (Theorem 1.2).

2. The Uncertain Communication Complexity Model

We start by recalling the classical communication complexity model of Yao (1979) and then present our definition and measures.

2.1. Communication Complexity. We start with some basic notation. For an integer $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1, \ldots, n\}$. We use $\log x$ to denote a logarithm in base 2. For two sets A and B, we denote by $A \triangle B$ their symmetric difference. For a distribution μ, we denote by $X \sim \mu$ the process of sampling a random variable from the distribution μ. Similarly, for a set $\mathcal{X}$ we denote by $X \sim \mathcal{X}$ the process of sampling a value X from the uniform distribution over $\mathcal{X}$. For any event E, let $\mathbb{1}(E)$ be the 0-1 indicator of E. For a pair (X, Y) of random variables sampled from a probability distribution μ, we denote by μ_X (respectively μ_Y) the marginal of μ over X (respectively Y). By $\mu_{Y|x}$, we denote the conditional distribution of μ over Y conditioned on $X = x$.

Given a distribution μ supported on a set $\mathcal{X}$ and functions $f, g \colon \mathcal{X} \to \Sigma$, we let $\delta_\mu(f, g)$ denote the (weighted and normalized) Hamming distance between f and g, i.e., $\delta_\mu(f, g) \triangleq \Pr_{X \sim \mu}[f(X) \neq g(X)]$. (Note that this definition extends naturally to probabilistic functions f and g, i.e., by letting $f(X)$ and $g(X)$ be sampled independently for every fixed value of X.) We say that f is δ-close to g (with respect to μ if μ is not clear from context) if $\delta_\mu(f, g) \leq \delta$.

We now turn to the definition of *communication complexity.* A more extensive introduction can be found in Kushilevitz & Nisan (1997). Let $f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a function and Alice and Bob be two parties. A protocol Π between Alice and Bob specifies how and what Alice and Bob communicate given their respective inputs and communication thus far. It also specifies when they stop and produce an output (that we require to be produced by Bob). A protocol is said to be *one-way* if it involves a single message from Alice to Bob, followed by Bob producing the output. The protocol Π is said to compute f if for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ it holds that $\Pi(x, y) = f(x, y)$. The communication cost of Π is the number of bits transmitted during the execution of the protocol between Alice and Bob, maximized over all possible inputs. The communication complexity of f is the minimal communication cost of a protocol computing f.

It is usual to relax the above setting by introducing a distribution μ over the input space $\mathcal{X} \times \mathcal{Y}$ and requiring the protocol to succeed with high probability (rather than with probability 1). We say that a protocol Π ϵ-*computes* a function f under distribution μ if $\delta_\mu(\Pi, f) \leq \epsilon$. We next define the *distributional communication complexity* both for *functions* (as usual in the field of communication complexity) and for *families of pairs of functions* (which, as discussed in Section 1, are central to our work).

Definition 2.1 (Distributional Communication Complexity). *Let $f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a Boolean function and μ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$. The* distributional communication complexity *of f under μ with error ϵ, denoted by $\mathsf{CC}^\mu_\epsilon(f)$, is defined as the minimum over all protocols Π that ϵ-compute f over μ, of the communication cost of Π. The* one-way distributional communication complexity *$\mathsf{owCC}^\mu_\epsilon(f)$ is defined similarly by minimizing over one-way protocols Π.*

Let $\mathcal{F} \subseteq \{f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}\}^2$ be a family of pairs of Boolean functions with domain $\mathcal{X} \times \mathcal{Y}$. We define the distributional communication complexity $\mathsf{CC}^\mu_\epsilon(\mathcal{F})$ of $\mathcal{F}$ as the maximum value of $\mathsf{CC}^\mu_\epsilon(g)$ over all pairs $(f, g) \in \mathcal{F}$. Similarly, we define the one-way distributional communication complexity $\mathsf{owCC}^\mu_\epsilon(\mathcal{F})$ of $\mathcal{F}$ as the maximum value of $\mathsf{owCC}^\mu_\epsilon(g)$ over all functions $(f, g) \in \mathcal{F}$.

We note that it is also common to provide Alice and Bob with a shared random string which is independent of x, y and f. In the distributional communication complexity model, it is a known fact that any protocol with shared randomness can be used to get a protocol that does *not* use shared randomness without increasing its distributed communication complexity Yao (1977).

In this paper, unless stated otherwise, whenever we refer to a protocol, we think of the input pair (x, y) as coming from a distribution.

2.2. Uncertain-Communication Complexity. We now turn to the central definition of this paper: *uncertain-communication complexity.* Our goal is to understand how Alice and Bob can communicate when the function that Bob wishes to determine is not known to Alice. In this setting, we make the functions g (that Bob wants to compute) and f (Alice's estimate of g) explicitly part of the input to the protocol Π. Thus, in this setting a protocol Π specifies how Alice with input (f, x) and Bob with input (g, y) communicate, and how they stop and produce an output. We denote the output by $\Pi((f, x), (g, y))$. We say that Π computes (f, g) if for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the protocol outputs $g(x, y)$. We say that a (public-coin) protocol Π ϵ-computes (f, g) over μ if $\Pr_{(X,Y)\sim\mu}[g(X, Y) \neq \Pi((f, X), (g, Y))] \leq \epsilon$.

Next, one may be tempted to define the communication complexity of a pair of functions (f, g) as the minimum over all protocols that compute (f, g) of their maximum communication. But this does not capture the uncertainty! (Rather, a protocol that works for the pair corresponds to both Alice and Bob knowing both f and g.) To model the uncertainty, we have to consider the communication complexity of a whole class of pairs of functions, from which the pair (f, g) is chosen (in our case by an adversary).

Let $\mathcal{F} \subseteq \{f \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}\}^2$ be a family of pairs of Boolean functions with domain $\mathcal{X} \times \mathcal{Y}$. We say that a public-coin protocol Π ϵ-computes $\mathcal{F}$ over μ if for every $(f, g) \in \mathcal{F}$, we have that Π ϵ-computes (f, g) over μ.

We now define the uncertain-communication complexity of a family of functions $\mathcal{F}$. (Note that this is exactly the same as in Definition 1.1.)

Definition 2.2 (Uncertain-Communication Complexity). *Let*

μ be a distribution on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{F} \subseteq \{f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}\}^2$. The uncertain-communication complexity *of $\mathcal{F}$, denoted $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F})$, is the minimum over all public-coin protocols* Π *that ϵ-compute $\mathcal{F}$ over μ, of the maximum communication complexity of* Π *over all $(f,g) \in \{h \colon \mathcal{X} \times \{\mathcal{Y} \to \{0,1\}\}^2$, all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and all settings of the public coins.*

The one-way *uncertain-communication complexity of $\mathcal{F}$, denoted by $\mathsf{owCCU}^{\mu}_{\epsilon}(\mathcal{F})$, is defined similarly by restricting to one-way protocols.*

We remark that while in the standard distributional model of Section 2.1, the "easy direction" of Yao's minimax principle (Yao 1977) implies that shared randomness can be assumed *without loss of generality*, this is not necessarily the case in Definition 2.2. This is because the function pair (f,g) is selected *adversarially* from the class $\mathcal{F}$ and hence shared randomness can help the protocol "fool" this adversary.[7]

Also, observe that in the special case where $\mathcal{F} = \{(f,g)\}$, Definition 2.2 reduces to the standard definition of distributional communication complexity (i.e., Definition 2.1) for the function-class $\mathcal{F} = \{(f,g)\}$, and we thus have $\mathsf{CCU}^{\mu}_{\epsilon}(\{(f,g)\}) = \mathsf{CC}^{\mu}_{\epsilon}(\{(f,g)\})$. Furthermore, the uncertain communication complexity is monotone, i.e., if $\mathcal{F} \subseteq \mathcal{F}'$ then $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \leq \mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}')$. Hence, we conclude that $\mathsf{CCU}^{\mu}_{\epsilon}(\mathcal{F}) \geq \mathsf{CC}^{\mu}_{\epsilon}(\mathcal{F})$.

In this work, we attempt to identify a setting under which the last lower bound above can be matched. If the set of functions $\Gamma(g) := \{f \mid (f,g) \in \mathcal{F}\}$ is not sufficiently informative about g, then it seems hard to conceive of settings where Alice and Bob can do non-trivially well. We thus impose a simple and natural restriction on $\Gamma(g)$, namely, that it consists of functions that are close to g (in δ_{μ}-distance). This leads us to the definition of the

[7] If the pair (f,g) was sampled from some fixed probability distribution, then shared randomness would no longer be needed and deterministic protocols would be optimal. However, an adversarial assumption on (f,g) is more desirable since it is more likely to model natural scenarios. The reason why we choose the input pair (x,y) from a fixed distribution is to be able to define a notion of *distance* between two functions. Henceforth, we assume that the pair (f,g) is chosen adversarially.

distance of a family of pairs of functions.

DEFINITION 2.3 (Distance of a family, $\delta_\mu(\mathcal{F})$). *Let $\mathcal{F} \subseteq \{f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}\}^2$ be a family of pairs of Boolean functions with domain $\mathcal{X} \times \mathcal{Y}$, and let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$. The μ-*distance of *$\mathcal{F}$, denoted $\delta_\mu(\mathcal{F})$, is defined as the maximum over all $(f,g) \in \mathcal{F}$ of the distance $\delta_\mu(f,g)$.*

An optimistic hope might be that given (f,g) the players can exchange a few bits and agree on a function h which is close to both f and g, and thus reduce the task to that of computing h in the standard (certain-communication) setting. Our Theorem 1.7 shows that this naive strategy cannot work, in that there exists a family of nearby functions where agreement takes exponentially more communication than the simple strategy of simply exchanging x and y. We then prove Theorem 1.3 which gives an upper bound on the one-way uncertain-communication complexity, $\mathsf{owCCU}^\mu_\epsilon(\mathcal{F})$, which is comparable to the one-way certain-communication complexity $\mathsf{owCC}^\mu_\epsilon(\mathcal{F})$, when $\delta_\mu(\mathcal{F})$ is small, and μ is a product distribution. More generally, the theorem shows that the bound grows slowly as long as the mutual information between X and Y is small. Finally we prove Theorem 1.2, showing that for general non-product distributions, $\mathsf{owCCU}^\mu_\epsilon(\mathcal{F})$ can be much larger than $\mathsf{owCC}^\mu_\epsilon(\mathcal{F})$ even when the distance $\delta_\mu(\mathcal{F})$ is a small. More precisely, we construct a family of close-by functions along with a distribution μ for which the one-way certain-communication complexity is a single bit whereas the two-way uncertain-communication complexity is at least $\Omega(\sqrt{n})$.

3. Hardness of Contextual Agreement

In this section, we show that even if f and g are very close and have small one-way distributional communication complexity over a distribution μ (for every $(f,g) \in \mathcal{F}$), agreeing on an h such that $\delta_\mu(h,f)$ and $\delta_\mu(g,h)$ are non-trivially small takes communication that is roughly the size of the binary representation of f (which is exponential in the size of the input). Thus, agreeing on h before simulating a protocol for h is exponentially costlier than even the trivial protocol where Alice sends her input x to Bob. Formally, we consider the following communication problem:

Definition 3.1 ($\textsc{Agree}_{\delta,\gamma}(\mathcal{F})$). *For any given family of pairs of functions $\mathcal{F} \subseteq \{f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}\}^2$, the $\mathcal{F}$-agreement problem with parameters $\delta, \gamma \geq 0$ (denoted by $\textsc{Agree}_{\delta,\gamma}(\mathcal{F})$) is the communication problem where Alice gets f and Bob gets g such that $(f,g) \in \mathcal{F}$ and their goal is for Alice to output h_A and Bob to output h_B such that $\delta(h_A, f), \delta(h_B, g) \leq \delta$ and $\Pr[h_A = h_B] \geq \gamma$, where the probability is over the internal randomness of the protocol.*

Somewhat abusing notation, we will use $\textsc{Agree}_{\delta,\gamma}(\mathcal{D})$ to denote the distributional problem where $\mathcal{D}$ is a distribution on $\{f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}\}^2$ and the goal now is to get agreement with probability γ over the randomness of the protocol and that of the inputs.

If the agreement problem could be solved with low communication for a family $\mathcal{F}$ of pairs of Boolean functions, then it would imply a natural protocol for $\mathcal{F}$ in the uncertain-communication case. The following theorem, which is a refinement of Theorem 1.7 proves that agreement is extremely expensive even when all the functions that appear in the class $\mathcal{F}$ have *zero* communication complexity.

Theorem 3.2. *Let μ denote the uniform distribution over $\mathcal{X} \times \mathcal{Y}$. For every $\delta, \delta' \in (0, 1/2)$, there exist $\alpha > 0$ and $\beta < \infty$ such that for every $\gamma > 0$ and finite sets $\mathcal{X}$ and $\mathcal{Y}$ the following hold: There is a family $\mathcal{F}$ of pairs of Boolean functions over $\mathcal{X} \times \mathcal{Y}$ satisfying $\delta_\mu(\mathcal{F}) \leq \delta$ and $\mathsf{CC}_0^\mu(\mathcal{F}) = 0$, such that $\mathsf{CC}(\textsc{Agree}_{\delta',\gamma}(\mathcal{F})) \geq \alpha|\mathcal{Y}| - \beta \log(1/\gamma)$ where μ is the uniform distribution over $\mathcal{X} \times \mathcal{Y}$.*

Note that Theorem 1.7 corresponds to the special case of $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$ and γ being an absolute constant.

Theorem 3.2 says that there is a family of pairs of functions supported on functions of *zero* communication complexity (with zero error) for which agreement takes communication polynomial in the size of the domain of the functions. Note that this is exponentially larger than the trivial communication complexity for any function g, which is at most $\min\{1 + \log|\mathcal{Y}|, \log|\mathcal{X}|\}$ (which would result from either Alice sending the binary representation of her input to Bob, or Bob sending the binary representation of his input to Alice). Furthermore, this lower bound holds even if the goal is to get agreement with probability only exponentially small in $|\mathcal{Y}|$, which is really tiny!

Our proof of Theorem 3.2 uses a lower bound on the communication complexity of the *agreement distillation (with imperfectly shared randomness)* problem defined in Canonne *et al.* (2015), who in turn rely on a lower bound for randomness extraction from correlated sources due to Bogdanov & Mossel (2011). We describe their problem below and the result that we use. We note that their context is slightly different and our description below is a reformulation. First, we define the notion of ρ-noisy sequences of bits. A pair of bits (a, b) is said to be a pair of ρ-noisy uniform bits if a is uniform over $\{0, 1\}$, and $b = a$ with probability $1 - \rho$ and $b \neq a$ with probability ρ. A pair of sequences of bits (r, s) is said to be ρ-noisy if $r = (r_1, \ldots, r_n)$ and $s = (s_1, \ldots, s_n)$ and each coordinate pair (r_i, s_i) is a ρ-noisy uniform pair drawn independently of all other pairs. For a random variable W, we define its min-entropy as $H_\infty(w) \triangleq \min_{w \in \mathsf{supp}(W)}\{-\log(\Pr[W = w])\}$.

Definition 3.3 (Agreement-Distillation$^k_{\gamma,\rho}$). *In this problem, Alice and Bob get as inputs r and s respectively, where (r, s) form a ρ-noisy sequence of bits. Their goal is to communicate deterministically and produce as outputs w_A (Alice's output) and w_B (Bob's output) with the following properties: (i) $H_\infty(w_A), H_\infty(w_B) \geq k$ and (ii) $\Pr_{(r,s)}[w_A = w_B] \geq \gamma$.*

Lemma 3.4 (Canonne *et al.* 2015, Theorem 2). *For every $\rho \in (0, 1/2)$, there exists $\alpha > 0$ and $\beta > 0$ such that for every k and γ, it holds that every deterministic protocol Π that solves* Agreement-Distillation$^k_{\gamma,\rho}$ *has communication complexity at least $\alpha k - \beta \log(1/\gamma)$.*

We note that while the agreement distillation problem is very similar to our agreement problem, there are some syntactic differences. We are considering pairs of functions with low communication complexity, whereas the agreement distillation problem considers arbitrary random sequences. Also, our output criterion is proximity to the input functions, whereas in the agreement distillation problem we need to produce high-entropy outputs. Finally, we want a lower bound for our agreement problem when Alice and Bob are allowed to share perfect randomness while the agreement distillation bound only holds for deterministic protocols. Nevertheless, we are able to reduce to their setting as we will see shortly.

Our proof of Theorem 3.2 uses the standard Chernoff-Hoeffding tail inequality for random variables that we include below. Denote $\exp(x) \triangleq e^x$, where e is the base of the natural logarithm.

Proposition 3.5 (Chernoff bound; see e.g., Mitzenmacher & Upfal 2005). *Let $X = \sum_{i=1}^n X_i$ be a sum of independent identically distributed random variables $X_1, \ldots, X_n \in \{0,1\}$. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i]$. It holds that for every $\delta \in (0,1)$,*

$$\Pr[X < (1-\delta)\mu] \leq \exp\left(-\delta^2\mu/2\right)$$

and

$$\Pr[X > (1+\delta)\mu] \leq \exp\left(-\delta^2\mu/3\right),$$

and for $a > 0$,

$$\Pr[X > \mu + a] \leq \exp(-2a^2/n)$$

Proof (Proof of Theorem 3.2). Let μ be the uniform distribution on $\mathcal{X} \times \mathcal{Y}$. We prove the theorem for $\alpha/\beta < \delta/6$, in which case we may assume $\gamma > \exp(-\delta|\mathcal{Y}|/6)$ since otherwise the right-hand side in the statement of Theorem 3.2 is non-positive.

Let $\mathcal{F}_B$ denote the set of functions that depend only on Bob's input, i.e., $f \in \mathcal{F}_B$ if there exists $f' \colon \mathcal{Y} \to \{0,1\}$ such that $f(x,y) = f'(y)$ for all x, y. Our family $\mathcal{F}$ will be the subset of $\mathcal{F}_B \times \mathcal{F}_B$ consisting of pairs of functions that are at most δ apart (with respect to the uniform distribution on $\mathcal{X} \times \mathcal{Y}$), i.e.,

$$\mathcal{F} \triangleq \{(f,g) \in \mathcal{F}_B \times \mathcal{F}_B \mid \delta_\mu(f,g) \leq \delta\}.$$

Note that the zero-error communication complexity of every function in the support of $\mathcal{F}$ is zero since Bob can correctly compute its value without any information from Alice. Thus, $\delta_\mu(\mathcal{F}) = \delta$

and $\mathsf{CC}_0^{\mu}(\mathcal{F}) = 0$.[8] So it remains to prove a lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta',\gamma}(\mathcal{F}))$.

We prove our lower bound by picking a distribution $\mathcal{D}_\rho$ supported mostly on $\mathcal{F}$ and by giving a lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta',\gamma}(\mathcal{D}_\rho))$. Let $\rho = \delta/2$. The distribution $\mathcal{D}_\rho$ samples (f, g) as follows. The function f is drawn uniformly at random from $\mathcal{F}_B$. Since $f \in \mathcal{F}_B$, there exists a function $f' \colon \mathcal{Y} \to \{0, 1\}$ such that $f(x, y) = f'(y)$ for all x, y. Then, g is chosen to be a "ρ-noisy copy" of f. Namely, we define a function $g' \colon \mathcal{Y} \to \{0, 1\}$ such that for every $y \in \mathcal{Y}$, $g'(y)$ is chosen to be equal to $f'(y)$ with probability $1 - \rho$ and equal to $1 - f'(y)$ with probability ρ. Then, for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we set $g(x, y) = g'(y)$.

By the Chernoff bound (Proposition 3.5), we have that

$$\Pr_{(f,g)\sim\mathcal{D}_\rho}[\delta(f, g) > \delta] \leq \exp(-\rho|\mathcal{Y}|/3) < \gamma.$$

So with probability at least $1 - \gamma$, the distribution $\mathcal{D}_\rho$ draws elements from $\mathcal{F}$. So, if a protocol solves $\textsc{Agree}_{\delta',\gamma}(\mathcal{F})$, then if $(f, g) \sim \mathcal{D}_\rho$ then with probability at least $(1 - \gamma) \cdot \gamma$ we would have that the functions (f, g) are from $\mathcal{F}$ and the protocol achieves agreement on a nearby function. We conclude that a protocol solving $\textsc{Agree}_{\delta',\gamma}(\mathcal{F})$ is also a protocol solving $\textsc{Agree}_{\delta',\gamma-\gamma^2}(\mathcal{D}_\rho)$.

We thus need to show a lower bound on the communication complexity of $\textsc{Agree}_{\delta',\gamma-\gamma^2}(\mathcal{D}_\rho)$. We now note that since this is a distributional problem, by Yao's min-max principle, if there is randomized protocol to solve $\textsc{Agree}_{\delta',\gamma-\gamma^2}(\mathcal{D}_\rho)$, then there is also a deterministic protocol for the same problem and with the same communication complexity. Thus, it suffices to lower-bound the deterministic communication complexity of $\textsc{Agree}_{\delta',\gamma-\gamma^2}(\mathcal{D}_\rho)$. Claim 3.6 below shows that any such protocol gives a deterministic protocol for Agreement-Distillation with $k = \Omega_{\delta'}(|\mathcal{Y}|)$. Combining this with Lemma 3.4 gives us the desired lower bound on $\mathsf{CC}(\textsc{Agree}_{\delta',\gamma-\gamma^2}(\mathcal{D}_\rho))$ and hence on $\mathsf{CC}(\textsc{Agree}_{\delta',\gamma}(\mathcal{F}))$. □

[8] Indeed, even the uncertain-communication complexity of $\mathcal{F}$ is zero, further highlighting the lack of need of agreement to solve uncertain-communication problems.

Claim 3.6. *Every protocol for* $\text{Agree}_{\delta',\gamma}(\mathcal{D}_\rho)$ *is also a protocol for* $\text{Agreement-Distillation}^k_{\gamma,\rho}$ *for* $k = (1 - H_b(\delta'')) \cdot |\mathcal{Y}|$*, where* $\delta'' = \delta'(1 + o(1))$ *and* $H_b(\cdot)$ *is the binary entropy function given by* $H_b(x) \triangleq -x \log x - (1-x)\log(1-x)$*, where* $o(1)$ *denotes a function that goes to* 0 *as* $|\mathcal{Y}|$ *grows.*

Proof. Suppose Alice and Bob wish to solve $\text{Agreement-Distillation}^k_{\gamma,\rho}$. They can sample a ρ-noisy pair of strings $(r, s) \in \{0,1\}^{|\mathcal{Y}|}$ and interpret them as functions $f', g' \colon \mathcal{Y} \to \{0,1\}$ or equivalently as functions $(f, g) \sim \mathcal{D}_\rho$ by letting $f, g \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be given by $f(x,y) = f'(y)$ and $g(x,y) = g'(y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. They can now simulate the protocol for $\text{Agree}_{\delta,\gamma}(f,g)$ and output $h_A \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ (on Alice's side) and $h_B \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ (on Bob's side). By definition of Agree, we have that $h_A = h_B$ with probability at least γ. So it suffices to show that $H_\infty(h_A), H_\infty(h_B) \geq k$.

The intuitive idea for establishing this is simple. In order to show that the min-entropy of h_A (symmetrically, h_B) is large, we need to argue that a given h_A cannot be a output by a correct protocol for $\text{Agree}_{\delta',\gamma}(\mathcal{D}_\rho)$ with too high a probability. We expect this to be true because a given h_A cannot be close within δ' to too many input functions f. In order to formally argue this, we define the real-valued function $h'_A \colon \mathcal{Y} \to [0,1]$ as $h'_A(y) := \mathbb{E}_{x \sim \mathcal{X}}[h_A(x,y)]$ for all $y \in \mathcal{Y}$. By the triangle inequality, we have that

$$\begin{aligned}
\delta(h'_A, f') &:= \mathbb{E}_{y\sim\mathcal{Y}}[|h'_A(y) - f'(y)|] \\
&= \mathbb{E}_{y\sim\mathcal{Y}}[|\mathbb{E}_{x\sim\mathcal{X}}[h_A(x,y) - f(x,y)]|] \\
&\leq \mathbb{E}_{(x,y)\sim\mathcal{X}\times\mathcal{Y}}[|h_A(x,y) - f(x,y)|] \\
&= \delta(h_A, f) \\
&\leq \delta'.
\end{aligned}$$

We now define a "randomized rounding" of h_A to be a random function $h' \colon \mathcal{Y} \to \{0,1\}$ such that independently for each $y \in \mathcal{Y}$, we have that $h'(y) = 1$ with probability $h'_A(y)$, and $h'(y) = 0$ with probability $1 - h'_A(y)$. Define S to be the set of all Boolean-valued functions $\tilde{f}' \colon \mathcal{Y} \to \{0,1\}$ such that $\delta(h'_A, \tilde{f}') \leq \delta'$. We now show that with probability $1 - o(1)$ over the random choice of h', at least a $1-o(1)$ fraction of the functions $\tilde{f}' \in S$ are such that $\delta(h', \tilde{f}') \leq \delta'(1+o(1))$. To see this, note that for any fixed $\tilde{f}' \in S$, we have that $\mathbb{E}_{h'}[\delta(h', \tilde{f}')] = \delta(h'_A, \tilde{f}') \leq \delta'$, and hence by the Chernoff bound (Proposition 3.5), $\Pr[\delta(h', \tilde{f}') > \delta'(1+o(1))] \leq o(1)$. This implies that

$$\mathbb{E}_{h'}\left[\Pr_{\tilde{f}' \sim S}[\delta(h', \tilde{f}') > \delta'(1+o(1))]\right] \leq o(1).$$

Thus, there exists a setting $h' \colon \mathcal{Y} \to \{0,1\}$ such that a $1-o(1)$ fraction of the functions in S are within a distance of $\delta'(1+o(1))$ from h'. Thus,

$$\begin{aligned} |S| &\leq |\{\tilde{f}' \colon \mathcal{Y} \to \{0,1\} \mid \delta(h', \tilde{f}') \leq \delta'(1+o(1))\}| \cdot (1+o(1)) \\ &\leq 2^{H_b(\delta'(1+o(1)))|\mathcal{Y}|} \cdot (1+o(1)), \qquad (3.7) \end{aligned}$$

where the last inequality follows from the fact that h' is a Boolean-valued function. Thus, we conclude that the right-hand side in ((3.7)) is also an upper bound on the number of functions $f \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ such that $f(x,y) = f'(y)$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$ for some function $f' \colon \mathcal{Y} \to \{0,1\}$ and that satisfy $\delta(h_A, f) \leq \delta$. Since the probability of sampling any such f is equal to $2^{-|\mathcal{Y}|}$, we get that probability of outputting any particular function $h_A \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ is at most $2^{-(1-H_b(\delta'(1+o(1)))|\mathcal{Y}|}$. This means that $H_\infty(h_A) \geq (1-H_b(\delta'(1+o(1)))|\mathcal{Y}|$. A similar lower bound applies to $H_\infty(h_B)$. Thus, we have that the outputs of the protocol for AGREE solve AGREEMENT-DISTILLATION$^k_{\gamma,\rho}$ with $k = (1-H_b(\delta'(1+o(1))))|\mathcal{Y}|$. □

4. One-way Uncertain-Communication Complexity

In this section, we prove Theorem 1.3 which we restate below (with a slight notational change — we use $\mathcal{X} \times \mathcal{Y}$ to denote the domain of the functions, as opposed to $\{0,1\}^n \times \{0,1\}^n$).

THEOREM 4.1 (restated). *There exists a constant c such that for all positive integers k and n and positive reals ϵ, δ, θ, for every* *distribution μ over $\mathcal{X} \times \mathcal{Y}$, and every family $\mathcal{F}$ of pairs of Boolean functions satisfying $\delta_\mu(\mathcal{F}) \leq \delta$ and $\mathsf{owCC}^\mu_\epsilon(\mathcal{F}) \leq k$, it holds that*

$$\mathsf{owCCU}^\mu_{\epsilon+2\delta+\theta}(\mathcal{F}) \leq c \cdot \frac{\left(k + \log\left(\frac{1}{\theta}\right)\right)}{\theta^2} \cdot \left(1 + \frac{I(X;Y)}{\theta^2}\right). \tag{4.2}$$

4.1. Overview of Protocol. We start with a high-level description of the protocol. Let μ be a distribution over an input space $\mathcal{X} \times \mathcal{Y}$. For any function $s \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and any $x \in \mathcal{X}$, we define the *restriction* of s to x to be the function $s_x \colon \mathcal{Y} \to \{0,1\}$ given by $s_x(y) = s(x,y)$ for any $y \in \mathcal{Y}$. We will consider a pair (X,Y) of random variables drawn from μ.

First, we consider the particular case of Theorem 1.3 where μ is a product distribution, i.e., $\mu = \mu_X \times \mu_Y$. Note that in this case, $I(X;Y) = 0$ in the right-hand side of (4.2). We will handle the case of general (not necessarily product) distributions later on.

The general idea is that given inputs (f, X), Alice can determine the restriction f_X, and she will try to describe it to Bob. For most values $x \in \mathcal{X}$, we have that f_x will be close (in δ_{μ_Y}-distance) to the function g_x. Bob will try to use the (yet unspecified) description given by Alice in order to determine some function B that is close to g_x. If he succeeds in doing so, he can output $B(Y)$ which would equal $g_x(Y)$ with high probability over Y.

We next explain how Alice will describe f_X, and how Bob will determine some function B that is close to g_X based on Alice's description. For the first part, we let Alice and Bob use shared randomness in order to sample $Y_1, \ldots, Y_m$, where the Y_i's are drawn independently with $Y_i \sim \mu_Y$, and m is a parameter to be chosen later. Alice's description of f_X will then be $(f_X(Y_1), \ldots, f_X(Y_m)) \in \{0,1\}^m$. Thus, the length of the communication is m bits and we need to show that setting m to be roughly $O(k)$ suffices. Before we explain this, we first need to specify what Bob does with Alice's message.

As a first cut, let us consider the following natural strategy: Bob picks an $\tilde{X} \in \mathcal{X}$ such that $g_{\tilde{X}}$ is close to f_X on $Y_1, \ldots, Y_m$, and sets $B = g_{\tilde{X}}$. It is clear that if $\tilde{X} = X$, then $B = g_{\tilde{X}} = g_X$, and for every $y \in \mathcal{Y}$, we would have $B(y) = g_X(y)$. Moreover, if $\tilde{X}$ is such that $g_{\tilde{X}}$ is close to g_X (which is itself close to f_X, for most values of X), then $B(Y)$ would now equal $g_X(Y)$ with high probability. It remains to deal with $\tilde{X}$ such that $g_{\tilde{X}}$ is far from g_X. Note that if we first fix any such $\tilde{X}$ and then sample $Y_1, \ldots, Y_m$, then with high probability, we would reveal that $g_{\tilde{X}}$ is far from g_X. This is because g_X is close to f_X (for most values of X), so $g_{\tilde{X}}$ should also be far from f_X. However, this idea alone cannot deal with all possible $\tilde{X}$ — using a naive union bound over all possible $\tilde{X} \in \mathcal{X}$ would require a failure probability of $1/|\mathcal{X}|$, which would itself require setting m to be roughly $\log |\mathcal{X}|$. Indeed, smaller values of m should not suffice since we have not yet used the fact that $\mathsf{CC}^{\mu}_{\epsilon}(g) \leq k$ — but we do so next.

Suppose that Π is a one-way protocol with k bits of communication. Then, note that Alice's message partitions $\mathcal{X}$ into 2^k sets, one corresponding to each message. Our modified strategy for Bob is to let him pick a representative from each set in this partition, and then set $B = g_{\tilde{X}}$ for an $\tilde{X}$ among the representatives for which $g_{\tilde{X}}$ and f are the closest on the samples $Y_1, \ldots, Y_m$. A simple analysis shows that the g_x's that lie inside the same set in this partition are close, and thus, if we pick $\tilde{X}$ to be the representative of the set containing X, then $g_{\tilde{X}}$ and f_X will be close on the sampled points. For another representative, once again if $g_{\tilde{X}}$ is close to g_X, then $g_{\tilde{X}}(Y)$ will equal $g_X(Y)$ with high probability. For a representative x' such that $g_{x'}$ is far from g_X (which is itself close to f_X), we can proceed as in the previous paragraph, and now the union bound works out since the total number of representatives is only 2^k.[9]

We now turn to the case of general (not necessarily product) distributions. In this case, we would like to run the above protocol with $Y_1, Y_2, \ldots, Y_m$ sampled independently from $\mu_{Y|x}$ (instead of μ_Y) where x is the particular realization of Alice's input. Note that Alice knows x and hence knows the distribution $\mu_{Y|x}$. Unfortunately, Bob does not know $\mu_{Y|x}$; he only knows μ_Y as a "proxy" for $\mu_{Y|x}$. While Alice and Bob cannot jointly sample such Y_i's without communicating (as in the product case), they can still run

[9] We note that a similar idea was used in a somewhat different context by Bar-Yossef *et al.* (2002) (following on Kremer *et al.* (1999)) in order to characterize one-way communication complexity of any function under product distributions in terms of its VC-dimension.

the rejection sampling protocol of Harsha *et al.* (2007) in order to agree on such samples while communicating at most $O(m \cdot I(X;Y))$ bits (see Section 4.2 for more details).

The outline of the rest of this section is the following. In Section 4.2, we describe the properties of the correlated sampling procedure that we will use. In Section 4.3, we give the formal proof of Theorem 1.3.

4.2. Rejection Sampling. We start by recalling two standard notions from information theory. Given two distributions P and Q, the *KL divergence* between P and Q is defined as $D(P||Q) \triangleq \mathbb{E}_{u \sim P}[\log(P(u)/Q(u))]$. Given a joint distribution μ of a pair (X, Y) of random variables with μ_X and μ_Y being the marginals of μ over X and Y respectively, the *mutual information* of X and Y is defined as $I(X;Y) \triangleq D(\mu||\mu_X\mu_Y)$. The following lemma summarizes the properties of the *rejection sampling* protocol of Harsha *et al.* (2007).

Lemma 4.3 (Rejection Sampling; Harsha *et al.* 2007). *Let P be a distribution known to Alice and Q be a distribution known to both Alice and Bob, with $D(P||Q)$ being finite. There exists a one-way public-coin protocol (with communication from Alice to Bob) such that at the end of the protocol, Alice and Bob output a sample from P such that the expected communication cost (over the public-randomness of the protocol) is at most $D(P||Q) + 2\log(D(P||Q) + 1) + O(1)$ bits.*

We will use the following corollary of Lemma 4.3.

Corollary 4.4. *Let μ be a distribution over (X, Y) with marginal μ_X over X, and assume that μ is known to both Alice and Bob. Fix $\epsilon > 0$ and let Alice be given a realization $x \sim \mu_X$. There is a* one-way *public-coin protocol that uses at most*

$$O(m \cdot I(X;Y)/\epsilon) + O(1/\epsilon)$$

bits of communication such that with probability at least $1 - \epsilon$ over the public coins of the protocol and the randomness of x, Alice and Bob agree on m samples $Y_1, Y_2, \ldots, Y_m$ i.i.d. $\sim \mu_{Y|x}$ at the end of the protocol.

Proof. When x is Alice's input, we can consider running the protocol in Lemma 4.3 on the distributions $P \triangleq \prod_{i=1}^{m} \mu_{Y_i|x}$ and $Q \triangleq \prod_{i=1}^{m} \mu_{Y_i}$. Note that each of P and Q is a distribution over tuples $(y_1, y_2, \ldots, y_m)$. Let Π be the resulting protocol transcript. The expected communication cost of Π is at most

$$\begin{aligned}\mathbb{E}_{x\sim\mu_X}[O(D(P||Q)) + O(1)] &= O(\mathbb{E}_{x\sim\mu_X}[D(P||Q)])) + O(1)\\ &= O(m \cdot I(X;Y))) + O(1),\end{aligned} \tag{4.5}$$

where the last equality follows from the fact that

$$\begin{aligned}\mathbb{E}_{x\sim\mu_X}[D(P||Q)] &= \mathbb{E}_{x\sim\mu_X}\left[\mathbb{E}_{y_1|x,\ldots,y_m|x}\left[\log\left(\frac{\prod_{i=1}^{m}\mu_{Y_i|x}(y_i)}{\prod_{i=1}^{m}\mu_{Y_i}(y_i)}\right)\right]\right]\\ &= \sum_{i=1}^{m}\mathbb{E}_{x\sim\mu_X}\left[\mathbb{E}_{y_1|x,\ldots,y_m|x}\left[\log\left(\frac{\mu_{Y_i|x}(y_i)}{\mu_{Y_i}(y_i)}\right)\right]\right]\\ &= \sum_{i=1}^{m}\mathbb{E}_{x\sim\mu_X}\left[\mathbb{E}_{y_i|x}\left[\log\left(\frac{\mu_{Y_i|x}(y_i)}{\mu_{Y_i}(y_i)}\right)\right]\right]\\ &= \sum_{i=1}^{m}\mathbb{E}_{(x,y)\sim\mu}\left[\log\left(\frac{\mu_{Y|x}(y)}{\mu_{Y}(y)}\right)\right]\\ &= m \cdot I(X;Y).\end{aligned}$$

By Markov's inequality applied to ((4.5)), we get that with probability at least $1-\epsilon$, the length of the transcript Π is at most

$$O(m \cdot I(X;Y)/\epsilon) + O(1/\epsilon) \qquad \text{bits.}$$

The statement now follows. □

4.3. Proof of Theorem 1.3. Recall that in the uncertain setting, Alice's input is (f, X) and Bob's input is (g, Y), where $(f, g) \in \mathcal{F}$, $(X, Y) \sim \mu$ and $\mathcal{F}$ is a family of pairs of Boolean functions satisfying $\mathsf{owCC}_{\epsilon}^{\mu}(\mathcal{F}) \le k$ and $\delta_{\mu}(\mathcal{F}) \le \delta$. Let Π be the one-way protocol for g in the standard setting that shows that $\mathsf{owCC}_{\epsilon}^{\mu}(g) \le k$. Note that Π can be described by an integer $L \le 2^k$ and functions $\pi \colon \mathcal{X} \to [L]$ and $\{B_i \colon \mathcal{Y} \to \{0,1\}\}_{i\in[L]}$, such that Alice's message on input X is $\pi(X)$, and Bob's output on message i from Alice and on input y is $B_i(Y)$. We use this notation below. We also set the parameter $m = \Theta\big((k + \log(1/\theta))/\theta^2\big)$, which is chosen such that $2^k \cdot \exp(-\theta^2 m/75) \le 2\theta/5$.

Protocol. Protocol 4.6 describes the protocol Π' we employ in the uncertain setting. Roughly speaking, the protocol works as follows. First, Alice and Bob run the one-way rejection sampling procedure given by Corollary 4.4 in order to sample $y_1, y_2 \ldots, y_m$ i.i.d. $\sim \mu_{Y|x}$. Then, Alice sends the sequence $(f_x(y_1), \ldots, f_x(y_m))$ to Bob. Bob enumerates over $i \in [L]$ and counts the fraction of $z \in \{y_1, \ldots, y_m\}$ for which $B_i(z) \neq f_x(z)$. For the index i which minimizes this fraction, Bob outputs $B_i(y)$ and halts.

PROTOCOL 4.6. The uncertain-communication protocol Π'.

The Setting: Let μ be a probability distribution over a message space $\mathcal{X} \times \mathcal{Y}$. Alice and Bob are given functions f and g, and inputs x and y, respectively, where $(f, g) \in \mathcal{F}$ and $(x, y) \sim \mu$ are realizations of the random pair (X, Y).

The Protocol:

1. Alice and Bob run one-way rejection sampling with error parameter set to $(\theta/10)^2$ in order to sample m values $Z = \{y_1, y_2, \ldots, y_m\} \subseteq \mathcal{Y}$ each sampled independently according to $\mu_{Y|x}$.
2. Alice sends $\{f_x(y_i)\}_{i \in [m]}$ to Bob.
3. For every $i \in [L]$, Bob computes $\mathsf{err}_i \triangleq \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}(B_i(y_j) \neq f_x(y_j))$. Let $i_{\min} \triangleq \operatorname{argmin}_{i \in [L]}\{\mathsf{err}_i\}$. Bob outputs $B_{i_{\min}}(y)$ and halts.

Analysis. Observe that by Corollary 4.4, the rejection sampling procedure requires $O(m \cdot I(X;Y)/\theta^2 + 1/\theta^2)$ bits of communication. Thus, the total communication of our protocol is at most

$$O(m \cdot I(X;Y)/\theta^2 + 1/\theta^2) + m \leq \frac{c\left(k + \log\left(\frac{1}{\theta}\right)\right)}{\theta^2} \cdot \left(1 + \frac{I(X;Y)}{\theta^2}\right)$$

bits for some absolute constant c, as promised. The next lemma establishes the correctness of the protocol.

Lemma 4.7. $\Pr_{\Pi',(x,y)\sim\mu}[B_{i_{\min}}(y) \neq g(x,y)] \leq \epsilon + 2\delta + \theta$, *where the probability is over both the internal randomness of the protocol* Π' *and over the randomness of the input-pair* (x,y).

Proof. We start with some notation. For $x \in \mathcal{X}$, let $\delta_x \triangleq \delta_{\mu_{Y|x}}(f_x, g_x)$ and let $\epsilon_x \triangleq \delta_{\mu_{Y|x}}(g_x, B_{\pi(x)})$. Note that by definition, $\delta = \mathbb{E}_{x\sim\mu_X}[\delta_x]$ and $\epsilon = \mathbb{E}_{x\sim\mu_X}[\epsilon_x]$. For $i \in [L]$, let $\gamma_{i,x} \triangleq \delta_{\mu_{Y|x}}(f_x, B_i)$. Recall the description of the (given) deterministic protocol Π by the positive integers integer $L \leq 2^k$ and functions $\pi\colon \mathcal{X} \to [L]$ and $\{B_i\colon \mathcal{Y} \to \{0,1\}\}_{i\in[L]}$, such that Alice's message on input x is $\pi(x)$, and Bob's output on message i from Alice and on input y is $B_i(y)$. Note that by the triangle inequality,

$$\gamma_{\pi(x),x} = \delta_{\mu_{Y|x}}(f_x, B_{\pi(x)}) \leq \delta_x + \epsilon_x. \tag{4.8}$$

In what follows, we will analyze the probability that $B_{i_{\min}}(y) \neq g(x,y)$ by analyzing the estimate err_i and the index $i_{\min}$ computed in the above protocol. Note that $\mathsf{err}_i = \mathsf{err}_i(x)$ computed above attempts to estimate $\gamma_{i,x}$, and that both err_i and $i_{\min}$ are functions of x.

Note that Corollary 4.4 guarantees that rejection sampling succeeds with probability at least $1-\theta^2/100$. Henceforth, we condition on the event that rejection sampling succeeds (we will account for the event where this does not happen at the end). By the Chernoff bound (Proposition 3.5), and using the definition of err_i in Algorithm Protocol 4.6 and the fact that $\mathbb{E}[\mathbb{1}(B_i(y_j) \neq f_x(y_j))] = \gamma_{i,x}$, we have for every x and $i \in [L]$

$$\Pr_{y_1,\ldots,y_m \text{ i.i.d.}\sim\mu_{Y|x}}\left[|\gamma_{i,x} - \mathsf{err}_i| > \frac{\theta}{5}\right] \leq \exp\left(-\frac{\theta^2 \cdot m}{75}\right).$$

By a union bound, we have for every $x \in X$,

$$\begin{aligned}\Pr_{y_1,\ldots,y_m\sim\mu_{Y|x}}\left[\exists i \in [L] \text{ s.t. } |\gamma_{i,x} - \mathsf{err}_i| > \frac{\theta}{5}\right] &\leq L \cdot \exp\left(-\frac{\theta^2 \cdot m}{75}\right) \\ &\leq \frac{2\theta}{5},\end{aligned}$$

where the last inequality follows from our choice of $m = \Theta\big((k + \log(1/\theta))/\theta^2\big)$.

Now assume that for all $i \in [L]$, we have that $|\gamma_{i,x} - \mathsf{err}_i| \leq \theta/5$, which we refer to below as the “Good Event.” Then, for $i_{\min}$, we have

$$\begin{aligned}
&\text{(since we assumed the Good Event)} & \gamma_{i_{\min},x} &\leq \mathsf{err}_{i_{\min}} + \theta/5 \\
&\text{(by definition of } i_{\min}\text{)} & &\leq \mathsf{err}_{\pi(x)} + \theta/5 \\
&\text{(since we assumed the Good Event)} & &\leq \gamma_{\pi(x),x} + 2\theta/5 \\
&\text{(by (4.8))} & &\leq \delta_x + \epsilon_x + 2\theta/5.
\end{aligned}$$

Let $W \subseteq \{0,1\}^n$ be the set of all x for which rejection sampling succeeds with probability at least $1 - \theta/10$ (over the internal randomness of the protocol). By Corollary 4.4 and an averaging argument, $\Pr_{x \sim \mu_X}[x \notin W] \leq \theta/10$. Denoting by $\mu_X | x \in W$ the conditional probability distribution of $x \sim \mu_X$ conditioned on the event that $x \in W$, we thus get,

$$\begin{aligned}
&\Pr_{\Pi',(x,y)\sim\mu}[B_{i_{\min}}(y) \neq f(x,y)] \\
&\leq \Pr_{x\sim\mu_X}[x \in W] \cdot \mathbb{E}_{x\sim\mu_X|x\in W}\left[\Pr_{\Pi,y\sim\mu_{Y|x}}[B_{i_{\min}}(y) \neq f(x,y)]\right] + \frac{\theta}{10} \\
&\leq \Pr_{x\sim\mu_X}[x \in W] \cdot \mathbb{E}_{x\sim\mu_X|x\in W}\left[\Pr_{y_1,\ldots,y_m,y\sim\mu_{Y|x}}[B_{i_{\min}}(y) \neq f(x,y)]\right] \\
&\qquad +\theta/5 \\
&= \Pr_{x\sim\mu_X}[x \in W] \cdot \mathbb{E}_{x\sim\mu_X|x\in W}\left[\gamma_{i_{\min},x}\right] + \theta/5 \\
&\leq \Pr_{x\sim\mu_X}[x \in W] \cdot \mathbb{E}_{x\sim\mu_X|x\in W}\left[\delta_x + \epsilon_x + 2\theta/5\right] + 3\theta/5 \\
&\leq \mathbb{E}_{x\sim\mu_X}\left[\delta_x + \epsilon_x\right] + \theta \\
&= \delta + \epsilon + \theta,
\end{aligned}$$

where the third inequality follows from the fact that the Good Event occurs with probability at least $1 - 2\theta/5$, and from the corresponding upper bound on $\gamma_{i_{\min},x}$. The other inequalities above follow from the definition of the set W and the fact that $\Pr_{x\sim\mu_X}[x \notin W] \leq \theta/10$. Finally, since $\delta(f,g) \leq \delta$, we have that Bob's output does not equal $g(x,y)$ (which is the desired output) with probability at most $\epsilon + 2\delta + \theta$. □

5. Lower Bound for Non-Product Distributions

In this section, we prove Theorem 1.2, or rather a slight strengthening of this theorem as stated below.

THEOREM 5.1. *There exist absolute constants $\alpha > 0$ and $\beta < \infty$ such that for positive integer n, every $\delta \in (0,1)$, and $\epsilon < 1/2 - 2^{-\beta\sqrt{\delta n}}$ the following holds: There exists a distribution μ supported on $\{0,1\}^n \times \{0,1\}^n$ and a function class $\mathcal{F}$ satisfying $\delta_\mu(\mathcal{F}) \leq \delta$ and $\mathsf{owCC}_0^\mu(\mathcal{F}) \leq 1$ such that*

$$\mathsf{CCU}_\epsilon^\mu(\mathcal{F}) \geq \alpha\sqrt{\delta n} - \log\left(\frac{2}{1/2-\epsilon}\right).$$

Note that Theorem 1.2 is the special case where δ and ϵ are absolute constants.

To prove Theorem 1.2 we start by defining the class of function pairs and distributions that will be used. Consider the parity functions on subsets of bits of the string $x \oplus y \in \{0,1\}^n$ (which is the coordinate-wise XOR of the strings $x, y \in \{0,1\}^n$). Specifically, for every $S \subseteq [n]$, let $\chi_S : \{0,1\}^n \to \{0,1\}$ be defined by $\chi_S(x) = \oplus_{i \in S} x_i$ for all $x \in \{0,1\}^n$, and let $f_S \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be defined as

$$f_S(x,y) \triangleq \chi_S(x \oplus y) = \oplus_{i\in S}(x_i \oplus y_i).$$

Let $q = q(n) > 0$ and define the class of pairs of Boolean functions

$$\mathcal{F}_q \triangleq \{(f_S, f_T) : |S \triangle T| \leq q \cdot n\}. \tag{5.2}$$

Next, we define a probability distribution μ_p on $\{0,1\}^n \times \{0,1\}^n$ where $p = p(n)$. We do so by giving a procedure to sample according to μ_p. To sample a pair $(X,Y) \sim \mu_p$, we draw $X \sim \{0,1\}^n$ (i.e., we draw X uniformly from $\{0,1\}^n$) and let Y be a p-noisy copy of X, i.e., $Y \sim N_p(X)$. Here, $N_p(x)$ is the distribution on $\{0,1\}^n$ that outputs $Y \in \{0,1\}^n$ such that, independently, for each $i \in [n]$, $Y_i = 1 - x_i$ with probability p, and $Y_i = x_i$ with probability $1 - p$. In other words, $\mu_p(x,y) = 2^{-n} \cdot p^{|x \oplus y|} \cdot (1-p)^{n-|x\oplus y|}$ for every $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ where the notation $|z|$ stands for the Hamming weight of z, for $z \in \{0,1\}^n$.

We will prove Lemmas Lemma 5.3 and Lemma 5.4 below about the function class $\mathcal{F}_q$ and the distribution μ_p. In words, Lemma 5.3 says that every pair of functions in $\mathcal{F}_q$ are (pqn)-close in δ_{μ_p}-distance, and every function in $\mathcal{F}_q$ has a one-way zero-error certain-communication protocol with a single bit of communication. Lemma 5.4 lower-bounds the uncertain-communication complexity of $\mathcal{F}_q$ under distribution μ_p.

Lemma 5.3. *For every n, $p, q \in [0,1]$ we have $\mathsf{owCC}_0^{\mu_p}(\mathcal{F}_q) \leq 1$ and $\delta_{\mu_p}(\mathcal{F}_q) \leq pqn$.*

Lemma 5.4. *There exist constants $\gamma, \tau > 0$ such that for every n, $p \in (0, 1/2)$, $q \in (0,1)$ and $\epsilon < 1/2$, it holds that:*

$$\mathsf{CCU}_\epsilon^{\mu_p}(\mathcal{F}_q) \geq \gamma \cdot \min\{p \cdot n, (q/2) \cdot n\} - \log\left(\frac{1}{1/2 - (\epsilon + \eta)}\right),$$

where $\eta = 2^{-\tau \cdot q \cdot n}$.

Note that applying Lemmas Lemma 5.3 and Lemma 5.4 with $\mathcal{F} = \mathcal{F}_q$, $\mu = \mu_p$ and $p = q = \sqrt{\delta/n}$ (where $\delta > 0$) implies Theorem 5.1.

In Section 5.1 below, we prove Lemma 5.3 which follows from two simple propositions. The main part of the rest of this section is dedicated to the proof of Lemma 5.4. The idea behind the proof of Lemma 5.4 is to reduce the problem of computing $\mathcal{F}_q$ under μ_p with uncertainty, from the problem of computing a related function in the standard distributional communication complexity model (i.e., *without* uncertainty) under a related distribution. We then use the *discrepancy method* to prove a lower bound on the communication complexity of the new problem. This task itself reduces to upper bounding the spectral norm of a specific communication matrix. The choice of our underlying distribution then implies a tensor structure for this matrix, which reduces the spectral norm computation to bounding the largest singular value of an explicit family of 4×4 matrices.

We point out that our lower bound in Lemma 5.4 is essentially tight up to a logarithmic factor. Namely, one can show using a simple one-way hashing protocol that for any constant $\epsilon > 0$, $\mathsf{owCCU}_{\epsilon}^{\mu_p}(\mathcal{F}_q) \leq O(r \cdot \log r)$ with $r \triangleq \min\{p \cdot n, (q \cdot n)/2\}$. More precisely, let us first assume that $p \leq (q/2)$, in which case $r = p \cdot n$. Then, with very high probability, x and y are within a Hamming distance of $2 \cdot r$. Thus, Bob can learn $x \oplus y$ (and thus deduce x) if Alice sends him a (one-way) message of $O(r \cdot \log r)$ bits. Specifically, when $r = \Theta(\sqrt{n})$, it can be seen that the one-way protocol for the "$(2 \cdot r)$-Hamming distance problem" (see e.g., Huang *et al.* (2006) and Blais *et al.* (2014)) reveals to Bob the coordinate-wise XOR of x and y. Hence, Bob can deduce x and output $\chi_T(x \oplus y)$ in order to solve the uncertain problem in Lemma 5.4. The case where $p > (q/2)$ is similar, except that an "r-Hamming distance protocol" is now applied to the pair (S, T) (instead of the pair (x, y)); this would allow Bob to deduccertaine S and, upon receving the bit $\chi_S(y)$ from Alice, he can output $\chi_S(x \oplus y) = \chi_S(x) \oplus \chi_S(y)$.

5.1. Proof of Lemma 5.3. Lemma 5.3 follows from Proposition 5.5 and Proposition 5.6 below. We first show that every two functions in $\mathcal{F}_q$ are close under the distribution μ_p.

PROPOSITION 5.5. *For every $(f, g) \in \mathcal{F}_q$, it holds that $\delta_{\mu_p}(f, g) \leq pqn$.*

PROOF. Any pair of functions $(f, g) \in \mathcal{F}_q$ is of the form $f = f_S$ and $g = f_T$ with $|S \triangle T| \leq q \cdot n$. Hence,

$$\begin{aligned} \Pr_{(x,y)\sim\mu}[f(x,y) \neq g(x,y)] &= \Pr_{(x,y)\sim\mu}[\chi_{S\triangle T}(x \oplus y) = 1] \\ &\leq 1 - (1-p)^{|S\triangle T|} \\ &\leq 1 - (1-p)^{qn} \\ &\leq pqn. \end{aligned}$$

□

Next, we show that there is a simple one-way communication protocol that allows Alice and Bob to compute f_S (for any $S \subseteq [n]$) with just a single bit of communication.

PROPOSITION 5.6. $\mathsf{owCC}(f_S) = 1$.

PROOF. Recall that $f_S(x,y) = \oplus_{i\in S}(x_i \oplus y_i)$. We write this as $f_S(x,y) = (\oplus_{i\in S}(x_i)) \oplus (\oplus_{i\in S}(y_i))$. This leads to the simple one-way protocol where Alice computes $b = \oplus_{i\in S}(x_i)$ and sends the single bit result of the computation to Bob. Bob can now compute $b \oplus (\oplus_{i\in S}(y_i)) = f_S(x,y)$ to obtain the value of f_S (with zero error). □

5.2. Proof of Lemma 5.4. In order to lower bound $\mathsf{CCU}_{\epsilon}^{\mu_p}(\mathcal{F}_q)$, we define a certain-communication problem in the distributional setting that can be reduced to the problem of computing $\mathcal{F}_q$ in the uncertain setting. The lower bound in Lemma 5.4 is then obtained by proving a lower bound on the communication complexity of the new problem which is defined as follows:

- **Inputs:** Alice's input is a pair (S,x) where $S \subseteq [n]$ and $x \in \{0,1\}^n$. Bob's input is a pair (T,y) such that $T \subseteq [n]$ and $y \in \{0,1\}^n$.
- **Function:** The goal is to compute the function F given by

$$F((S,x),(T,y)) \triangleq f_T(x,y) = \chi_T(x \oplus y).$$

- **Distribution:** Let $\mathcal{D}_q$ be a distribution on pairs of subsets (S,T) of $[n]$ defined by the following sampling procedure. To sample $(S,T) \sim \mathcal{D}_q$, we pick a subset $S \subseteq [n]$ uniformly at random, and we then sample T by letting its 0/1 indicator vector be a $(q/2)$-noisy copy of the 0/1 indicator vector of S. The joint distribution on the inputs of Alice and Bob is then described by $\nu_{p,q} = \mathcal{D}_q \otimes \mu_p$: we sample $(x,y) \sim \mu_p$ and independently sample $(S,T) \sim \mathcal{D}_q$.

Proposition 5.7 below – which follows from a simple Chernoff bound – shows that a protocol computing $\mathcal{F}_q$ under μ_p can also be used to compute the function F in the standard distributional model with $((S,x),(T,y)) \sim \nu_{p,q}$, and with the same amount of communication.

Proposition 5.7. *There exists $\tau > 0$ such that for every $\epsilon < 1/2$, it holds that $\mathsf{CCU}_{\epsilon}^{\mu_p}(\mathcal{F}_q) \geq \mathsf{CC}_{\epsilon+\eta}^{\nu_{p,q}}(F)$ with $\eta = 2^{-\tau \cdot q \cdot n}$.*

Proof. Since $((S,x),(T,y)) \sim \nu_{p,q}$, we have that $(x,y) \sim \mu_p$ and $(S,T) \sim \mathcal{D}_q$. Thus, it suffices to show that for $(S,T) \sim \mathcal{D}_q$, it holds that $|S \triangle T| \leq q \cdot n$ with probability at least $1-\eta$, where $\eta = 2^{-\tau \cdot q \cdot n)}$ for some universal constant $\tau > 0$. This follows from the definition of $\mathcal{D}_q$, the Chernoff bound (Proposition 3.5) and the fact that $\mathbb{E}_{(S,T)\sim\mathcal{D}_q}[|S \triangle T|] = (q \cdot n)/2$. □

In the rest of this section, we will prove the following lower bound on $\mathsf{CC}_{\epsilon}^{\nu_{p,q}}(F)$, which along with Proposition 5.7, implies Lemma 5.4:

Lemma 5.8. *There exists $\gamma > 0$ such that for every n, $p \in (0,1/2)$, $q \in (0,1)$ and $\epsilon < 1/2$, we have*

$$\mathsf{CC}_{\epsilon}^{\nu_{p,q}}(F) \geq \gamma \cdot \min\{p \cdot n, (q/2) \cdot n\} - \log\left(\frac{1}{1/2 - \epsilon}\right).$$

We first state and prove a proposition that allows us to eliminate one of the two parameters p and q.

Proposition 5.9. *For every positive integer n, $p \in (0,1/2)$, $q \in (0,1)$ and $\epsilon < 1/2$ we have: $\mathsf{CC}_{\epsilon}^{\nu_{p,q}}(F) \geq \mathsf{CC}_{\epsilon}^{\nu_{r,2r}}(F)$ where $r \triangleq \min(p, q/2)$.*

Proof. We use the fact that Alice and Bob can perturb their inputs (using private randomness) to reduce the correlations among them. Specifically we use the fact that if y is a p-noisy copy of x and z is a η-noisy copy of y, then z is a $(p(1-\eta)+\eta(1-p))$-noisy copy of x. Below we show how to use this formally in a reduction.

Suppose $((S,x),(T,y)) \sim \nu_{r,2r}$ and Alice has (S,x) and Bob has (T,y) and the goal is to compute $F((S,x),(T,y)) = \chi_T(x \oplus y)$. Suppose Π is a protocol with communication complexity k, that ϵ-computes F on $\nu_{p,q}$.

If $q/2 > p = r$, then Alice samples S' η-noisily from S for $\eta = (q/2 - r)/(1-2r)$, so that $(S',T) \sim \mathcal{D}_q$. Alice and Bob can now compute $\Pi((S',x),(T,y))$ using k bits of communication.

By the correctness of Π we have that their output disagrees with $F((S',x),(T,Y))$ with probability at most ϵ. But then we have $F((S,x),(T,y)) = F((S',x),(T,y))$ since F does not depend on S, and so Bob can simply output the output of Π to get a protocol that ϵ-computes F on $\nu_{r,2r}$.

Now we turn to the case that $p \geq q/2 = r$. In this case Bob samples y' η-noisily from y, for $\eta = (p-r)/(1-2r)$, to get y' which is an r-noisy copy of x. By simulating $\Pi((S,x),(T,y'))$, Bob can ϵ-compute $\chi_T(x \oplus y')$. Now using the fact that $\chi_T(x \oplus y) = \chi_T(x \oplus y') \oplus \chi_T(y' \oplus y)$ we have that if Bob outputs $\Pi((S,x),(T,y')) \oplus \chi_T(y' \oplus y)$ then he gets a protocol that is correct with probability at least $1-\epsilon$.

Thus in either case Π can be converted to a protocol with the same communication that ϵ-computes F on $\nu_{r,2r}$. □

So to prove Lemma 5.8 from now we will set $q = 2p$ and prove a lower bound of $\gamma \cdot p \cdot n - \log(1/(1/2-\epsilon))$ on the communication complexity. So henceforth, we denote $\nu_p \triangleq \nu_{p,2p}$. The proof will use the *discrepancy bound* which is a well-known method for proving lower bounds on distributional communication complexity in the standard model.

Definition 5.10 (Discrepancy Kushilevitz & Nisan 1997, Definition 3.27). *Let F and ν_p be as above and let R be any rectangle (i.e., any set of the form $R = C \times D$ where $C, D \subseteq \{0,1\}^{2n}$). Denote*

$$\textsc{Disc}_{\nu_p}(R,F) \triangleq \Big| \Pr_{\nu_p}\big[F((S,x),(T,y)) = 0 \text{ and } ((S,x),(T,y)) \in R\big] - \Pr_{\nu_p}\big[F((S,x),(T,y)) = 1 \text{ and } ((S,x),(T,y)) \in R\big]\Big|.$$

The discrepancy of F according to ν_p is

$$\textsc{Disc}_{\nu_p}(F) \triangleq \max_R \textsc{Disc}_{\nu_p}(R,F),$$

where the maximum is over all rectangles R.

The next known proposition relates distributional communication complexity to discrepancy.

Proposition 5.11 (Kushilevitz & Nisan 1997, Proposition 3.28). *For every $\epsilon < 1/2$, it holds that* $\mathsf{CC}_{\epsilon}^{\nu_p}(F) \geq \log((1-2\epsilon)/\mathrm{Disc}_{\nu_p}(F))$.

We will prove the following lemma.

Lemma 5.12. $\mathrm{Disc}_{\nu_p}(F) \leq 2^{-\gamma \cdot p \cdot n}$ *for some absolute constant* $\gamma > 0$.

Note that Lemma 5.12 and Proposition 5.11 and Proposition 5.9 put together immediately imply Lemma Lemma 5.8. The proof of Lemma 5.12 uses some standard facts about the spectral properties of matrices and their tensor powers that we next recall. Let $A \in \mathbb{R}^{d \times d}$ be a real square matrix. Then, $v \in \mathbb{R}^d$ is said to be an eigenvector of A with eigenvalue $\lambda \in \mathbb{R}$ if $Av = \lambda v$. If A is furthermore (symmetric) positive semi-definite, then all its eigenvalues are real and non-negative. We can now define the spectral norm of a (not necessarily symmetric) matrix.

Definition 5.13. *The spectral norm of a matrix* $A \in \mathbb{R}^{d \times d}$ *is given by* $\|A\| \triangleq \sqrt{\lambda_{max}(A^T A)}$, *where* $\lambda_{max}(A^T A)$ *is the largest eigenvalue of* $A^T A$.

Also, recall that given a matrix $A \in \mathbb{R}^{d \times d}$ and a positive integer t, the tensor power matrix $A^{\otimes t} \in \mathbb{R}^{d^t \times d^t}$ is defined by $(A^{\otimes t})_{(i_1,\ldots,i_t),(j_1,\ldots,j_t)} \triangleq \prod_{\ell=1}^{t} A_{i_\ell, j_\ell}$ for every $(i_1, \ldots, i_t), (j_1, \ldots, j_t) \in \{1, \ldots, d\}^t$. We will use the following standard fact which in particular says that the spectral norm is multiplicative with respect to tensoring.

Fact 5.14 (e.g., Laub 2005). *For any matrix* $A \in \mathbb{R}^{d \times d}$, *vector* $u \in \mathbb{R}^d$, *scalar* $c \in \mathbb{R}$ *and positive integer* t, *we have*

(i) $\|cA\| = |c| \cdot \|A\|$.

(ii) $\|A^{\otimes t}\| = \|A\|^t$.

(iii) $\|Au\|_2 \leq \|A\| \cdot \|u\|_2$, *where for any vector* $w \in \mathbb{R}^d$, $\|w\|_2$ *denotes the* Euclidean norm *of* w, *i.e.*, $\|w\|_2 \triangleq \sqrt{\sum_{i=1}^{d} w_i^2}$.

The next lemma upper bounds the spectral norm of an explicit family of 4 × 4 matrices that will be used in the proof of Lemma 5.12. Looking ahead, it is crucial for our purposes that the coefficient multiplying a on the right-hand side of (5.16) is a constant smaller than 2.

Lemma 5.15. *Let $a \in (0,1)$ be a real number and $N \triangleq N(a) \triangleq \begin{bmatrix} 1 & a & a & -a^2 \\ a & 1 & -a^2 & a \\ a & a^2 & 1 & -a \\ a^2 & a & -a & 1 \end{bmatrix}$. Then,*

$$\|N\| \leq 1 + \sqrt{2} \cdot a + a^2 + \frac{a^4}{2} + \frac{a^5}{\sqrt{2}}. \tag{5.16}$$

The proof of Lemma 5.15 is deferred to the end of this section. We are now ready to prove Lemma 5.12.

Proof (Proof of Lemma 5.12). Fix any rectangle $R = C \times D$ where $C, D \subseteq \{0,1\}^{2n}$. We wish to show that $\text{Disc}_{\nu_p}(R, F) \leq 2^{-\gamma \cdot p \cdot n}$. First, note that $\text{Disc}_{\nu_p}(R, F) = |1_C M 1_D|$ where 1_C and 1_D are the 0/1 indicator vectors of C and D (respectively) and M is the $2^{2n} \times 2^{2n}$ real matrix defined by[10]

$$\begin{aligned} M_{((S,x),(T,y))} &\triangleq \nu_p((S,T),(x,y)) \cdot (-1)^{\chi_T(x \oplus y)} \\ &= \frac{1}{2^{2n}} (1-p)^{2n} (-1)^{\langle T, x \oplus y \rangle} (\frac{p}{1-p})^{|S \oplus T| + |x \oplus y|} \end{aligned}$$

for every $S, x, T, y \in \{0,1\}^n$. Letting $a \triangleq p/(1-p)$, we can write

$$M_{((S,x),(T,y))} = \frac{1}{2^{2n}} (1-p)^{2n} (N^{\otimes n})_{((S,x),(T,y))}$$

with $N = N(a)$ being the 4 × 4 real matrix defined by[11]

$$N_{((S_1,x_1),(T_1,y_1))} \triangleq (-1)^{T_1(x_1 \oplus y_1)} a^{|S_1 \oplus T_1| + |x_1 \oplus y_1|} \tag{5.17}$$

[10] We here use the symbols S and T to denote both subsets of $[n]$ and the corresponding 0/1 indicator vectors.

[11] In (5.17), $T_1(x_1 \oplus y_1)$ denotes the product of the bit T_1 and the bit $(x_1 \oplus y_1)$. Moreover, since $(S_1 \oplus T_1)$ is a single bit, its Hamming weight $|S_1 \oplus T_1|$ is the same as its bit-value, and similarly for $(x_1 \oplus y_1)$.

for all $S_1, x_1, T_1, y_1 \in \{0,1\}$. Using the third property listed in Fact 5.14, we get

$$\begin{aligned} \textsc{Disc}_{\nu_p}(R,F) &= |1_C^\top M 1_D| \leq \|1_C\|_2 \cdot \|M\| \cdot \|1_D\|_2 \\ &\leq \sqrt{2^{2n}} \cdot \|M\| \cdot \sqrt{2^{2n}} = 2^{2n} \cdot \|M\| \end{aligned} \tag{5.18}$$

We now use the first two properties listed in Fact 5.14 to relate $\|M\|$ to $\|N\|$ as follows:

$$\|M\| = \|\frac{1}{2^{2n}}(1-p)^{2n} N^{\otimes n}\| = \frac{1}{2^{2n}}(1-p)^{2n}\|N\|^n. \tag{5.19}$$

Using (5.17), we can check that

$$N = N(a) = \begin{bmatrix} 1 & a & a & -a^2 \\ a & 1 & -a^2 & a \\ a & a^2 & 1 & -a \\ a^2 & a & -a & 1 \end{bmatrix}.$$

Applying Lemma 5.15 with $a = p/(1-p)$ and p sufficiently small (e.g., less than $1/10$), we get

$$\|N\| \leq 1 + \sqrt{2} \cdot (\frac{p}{1-p}) + O(p^2). \tag{5.20}$$

Combining (5.18),(5.19) and (5.20) above, we conclude that

$$\begin{aligned} \textsc{Disc}_{\nu_p}(R,F) &\leq (1-p)^{2n} \cdot \big(1 + \sqrt{2} \cdot (\frac{p}{1-p}) + O(p^2)\big)^n \\ &= \Big[(1-p) \cdot \big(1 + p \cdot (\sqrt{2} - 1) + O(p^2)\big)\Big]^n \\ &= \Big[1 - p \cdot (2 - \sqrt{2}) + O(p^2)\Big]^n \\ &\leq 2^{-\gamma \cdot p \cdot n} \end{aligned}$$

for some absolute constant $\gamma > 0$. □

We conclude this section by proving Lemma 5.15.

Proof (Proof of Lemma 5.15). One can verify that

$$N^T N = \begin{bmatrix} (a^2+1)^2 & 2a(a^2+1) & 2a(1-a^2) & 0 \\ 2a(a^2+1) & (a^2+1)^2 & 0 & 2a(1-a^2) \\ 2a(1-a^2) & 0 & (a^2+1)^2 & -2a(a^2+1) \\ 0 & 2a(1-a^2) & -2a(a^2+1) & (a^2+1)^2 \end{bmatrix}.$$

Assuming that $a \in (0,1)$, one can also verify that $N^T N$ has as eigenvectors

$$v_1 \triangleq \begin{bmatrix} \frac{\sqrt{2(a^4+1)}}{1-a^2} \\ \frac{a^2+1}{1-a^2} \\ 1 \\ 0 \end{bmatrix}, v_2 \triangleq \begin{bmatrix} \frac{a^2+1}{1-a^2} \\ \frac{\sqrt{2(a^4+1)}}{1-a^2} \\ 0 \\ 1 \end{bmatrix}$$

$$\text{with eigenvalue } \lambda_1(a) \triangleq 2a^2 + a^4 + 2a\sqrt{2(a^4+1)} + 1,$$

and

$$v_3 \triangleq \begin{bmatrix} \frac{\sqrt{2(a^4+1)}}{a^2-1} \\ \frac{a^2+1}{1-a^2} \\ 1 \\ 0 \end{bmatrix}, v_4 \triangleq \begin{bmatrix} \frac{a^2+1}{1-a^2} \\ \frac{\sqrt{2(a^4+1)}}{a^2-1} \\ 0 \\ 1 \end{bmatrix}$$

$$\text{with eigenvalue } \lambda_2(a) \triangleq 2a^2 + a^4 - 2a\sqrt{2(a^4+1)} + 1.$$

Note that for any value of $a \in (0,1)$, the vectors v_1, v_2, v_3 and v_4 are linearly independent and each of the eigenvalues $\lambda_1(a)$ and $\lambda_2(a)$ has multiplicity 2. Moreover, we have that $\lambda_1(a) \geq \lambda_2(a)$. Hence,

$$\|N\| = \sqrt{\lambda_1(a)} = \sqrt{2a^2 + a^4 + 2a\sqrt{2(a^4+1)} + 1}.$$

Applying twice the fact that $\sqrt{1+x} \leq 1 + x/2$ for any $x \geq -1$, we get that

$$\begin{aligned}\|N\| &= \sqrt{1+2a^2+a^4+2a\sqrt{2}\sqrt{1+a^4}} \\ &\leq 1+a^2+\frac{a^4}{2}+a\sqrt{2}\sqrt{1+a^4} \\ &\leq 1+a^2+\frac{a^4}{2}+a\sqrt{2}(1+\frac{a^4}{2}) \\ &= 1+a\sqrt{2}+a^2+\frac{a^4}{2}+\frac{a^5}{\sqrt{2}}.\end{aligned}$$

□

6. The Need to Work with Approximations

For completeness, we exhibit a class $\mathcal{F}$ of pairs of Boolean-valued functions such that for every $(f, g) \in \mathcal{F}$, the functions f and g are very close with respect to the uniform distribution, the zero-error communication complexity of each of f and g in the standard model is a *single bit*, but the *zero-error* communication in the uncertain model is quite large. Formally, we prove the following:

THEOREM 6.1. *For every $\delta \in (0,1)$, there exists a class $\mathcal{F}$ of pairs of Boolean-valued functions over the domain $\{0,1\}^n$ such that*

(i) $\delta_\nu(\mathcal{F}) \leq \delta$.

(ii) $\mathsf{CC}_0(\mathcal{F}) = 1$.

(iii) $\mathsf{CCU}_0^\nu(\mathcal{F}) = \Omega(n - \log(1/\delta))$,

where ν is the uniform distribution on $\{0,1\}^n \times \{0,1\}^n$.

To prove Theorem 6.1, we will need the following lower bound on the well-studied INDEXING function. Recall that in the INDEXING$_m$ problem with parameter m, Alice is given an element $x \in [m]$, Bob is given a function $h\colon [m] \to \{0,1\}$, and they are required to output $h(x)$.[12] The next theorem asserts that the two-way communication complexity of INDEXING$_m$ is $\Omega(\log m)$ bits. Note that

[12] The standard definition of INDEXING terms x the "index" and views h as a vector in $\{0,1\}^m$. Our version is equivalent and a little more convenient notationally.

this bound is essentially tight as Alice can send her input to Bob using $\log m$ bits of communication.

THEOREM 6.2. *There is a constant $\epsilon > 0$ such that*

$$\mathsf{CC}_\epsilon(\textsc{Indexing}_m) = \Omega(\log m).$$

Theorem 6.2 follows from the well-known fact that the one-way communication complexity *from Bob to Alice* of $\textsc{Indexing}_n$ (a.k.a, the "hard direction") is $\Omega(n)$ bits (Kushilevitz & Nisan 1997, Exercise 4.20) and the generic fact that there is at most an exponential gap between one-way communication complexity and two-way communication complexity (Kushilevitz & Nisan 1997, Exercise 4.21).

We are now ready to prove Theorem 6.1.

PROOF (Proof of Theorem 6.1). We first describe the class of functions we work with. Let $T \subseteq \{0,1\}^n$ be a set of size $\delta \cdot 2^n$. Let $\mathcal{F}' = \{g' : \{0,1\}^n \to \{0,1\} \mid g'(x) = 0,\ \forall x \notin T\}$. We now define $\mathcal{F}$ in terms of $\mathcal{F}'$ as follows:

$$\mathcal{F} = \{(0,g) \mid \exists g' \in \mathcal{F}' \text{ s.t. } g(x,y) = g'(x),\ \forall (x,y)\}.$$

So the first function f is always the 0 function, and the second function g depends only on x and is always 0 if $x \notin T$.

Since $f \neq g$ only when $x \in T$ and this happens with probability δ, we have

$$\delta_\nu(f,g) \triangleq \Pr_{(X,Y)\sim\nu}[f(X,Y) \neq g(X,Y)] \leq \Pr_{(X,Y)\sim\nu}[X \in T] = \delta.$$

We conclude that $\delta_\nu(\mathcal{F}) \triangleq \max_{(f,g)\in\mathcal{F}}\{\delta_\nu(f,g)\} \leq \delta$ yielding Part (1) of the Theorem. Part (2) is immediate from the fact that $g(x,y) = g'(x)$ for every $(f,g) \in \mathcal{F}$ and so, in the certain-communication setting, Alice can compute $g'(x)$ and send it to Bob.

We now turn to Part (3) for which we give a reduction from $\textsc{Indexing}_m$ for $m = \delta \cdot 2^n$. Suppose $\mathsf{CCU}_0^\nu(\mathcal{F}) \leq k$ and so there is a protocol Π that communicates k bits such that if Alice is given (f,x) and Bob (g,y) with $(f,g) \in \mathcal{F}$, the protocol outputs $g(x,y)$. We show $k = \Omega(\log m) = \Omega(n - \log(1/\delta))$. (Note that since Π is a zero error protocol, we have that the output of Π is correct on all valid inputs, and so we can ignore the distribution on (x, y) below.) Associate $[m]$ with the set T, so that Alice's input is an element $x \in T$ and Bob's input is a function $h\colon T \to \{0,1\}$, and their goal is to compute $h(x)$. Let $g'\colon \{0,1\}^n \to \{0,1\}$ be given by $g'(x) = 0$ if $x \notin T$ and $g'(x) = h(x)$ otherwise. Let $g(x, y) = g'(x)$. Alice can map her input to the pair $(0, x)$ and Bob to the pair $(g, 0)$, and now they have inputs to our uncertain communication problem on $\mathcal{F}$. Running Π on this pair produces as output $g(x) = h(x)$ (since $x \in T$) which is the desired output of $\textsc{Indexing}_m$. Applying Theorem 6.2 we conclude $k = \Omega(\log m)$ and this yields Part (3). □

Acknowledgements

We would like to thank Oded Goldreich for very helpful feedback, suggestions and encouragement. We would like to thank Oded Goldreich for pointing out the need for Section 6 and Avi Wigderson for helpful feedback regarding the proof therein. We would like to thank Brendan Juba for illuminating discussions about previous attempts to studying the role of context in artificial intelligence. Finally, we would like to thank the anonymous reviewers for helpful comments and pointers.

A preliminary version of this paper appeared in the Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016, pages 2072–2085.

Badih Ghazi's research was supported in part by NSF STC Award CCF 0939370 and NSF Award CCF-1217423. Ilan Komargodski's work was done partially while he was an intern at Microsoft Research New England, and supported in part by grants from the I-CORE Program of the Planning and Budgeting Committee, the Israel Science Foundation, BSF and the Israeli Ministry of Science and Technology. Pravesh Kothari's work was done partly while he was an intern at Microsoft Research New England. Madhu Sudan's work was done in part while he was at Microsoft Research New England, and supported in part by a Simons Investigator Award and NSF Award CCF 1565641.

References

Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar & D. Sivakumar (2002). Information Theory Methods in Communication Complexity. In *17th Annual IEEE Conference on Computational Complexity*, 93–102.

Mohammad Bavarian, Dmitry Gavinsky & Tsuyoshi Ito (2014). On the role of shared randomness in simultaneous communication. In *International Colloquium on Automata, Languages, and Programming*, 150–162. Springer.

Eric Blais, Joshua Brody & Badih Ghazi (2014). The Information Complexity of Hamming Distance. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* 465.

Andrej Bogdanov & Elchanan Mossel (2011). On Extracting Common Random Bits From Correlated Sources. *IEEE Transactions on Information Theory* **57**(10), 6351–6355.

Clément Louis Canonne, Venkatesan Guruswami, Raghu Meka & Madhu Sudan (2015). Communication with Imperfectly Shared Randomness. In *Innovations in Theoretical Computer Science, ITCS*, 257–262.

Badih Ghazi, Pritish Kamath & Madhu Sudan (2016). Communication complexity of permutation-invariant functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, 1902–1921. SIAM.

Badih Ghazi & Madhu Sudan (2017). The Power of Shared Randomness in Uncertain Communication. In *International Colloquium on Automata, Languages and Programming (ICALP)*, 49:1–49:14.

Oded Goldreich, Brendan Juba & Madhu Sudan (2012). A theory of goal-oriented communication. *J. ACM* **59**(2), 8.

Elad Haramaty & Madhu Sudan (2014). Deterministic compression with uncertain priors. In *Innovations in Theoretical Computer Science, ITCS*, 377–386.

PRAHLADH HARSHA, RAHUL JAIN, DAVID MCALLESTER & JAIKUMAR RADHAKRISHNAN (2007). The communication complexity of correlation. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, 10–23. IEEE.

WEI HUANG, YAOYUN SHI, SHENGYU ZHANG & YUFAN ZHU (2006). The communication complexity of the Hamming distance problem. *Information Processing Letters* **99**(4), 149–153.

BRENDAN JUBA, ADAM TAUMAN KALAI, SANJEEV KHANNA & MADHU SUDAN (2011). Compression without a common prior: an information-theoretic justification for ambiguity in language. In *Innovations in Computer Science, ICS*, 79–86.

BRENDAN JUBA & MADHU SUDAN (2008). Universal semantic communication I. In *40th Annual ACM Symposium on Theory of Computing*, 123–132.

BRENDAN JUBA & MADHU SUDAN (2011). Efficient Semantic Communication via Compatible Beliefs. In *Innovations in Computer Science, ICS*, 22–31.

BRENDAN JUBA & RYAN WILLIAMS (2013). Massive online teaching to bounded learners. In *Innovations in Theoretical Computer Science, ITCS*, 1–10.

ILAN KREMER, NOAM NISAN & DANA RON (1999). On Randomized One-Round Communication Complexity. *Computational Complexity* **8**(1), 21–49.

EYAL KUSHILEVITZ & NOAM NISAN (1997). *Communication complexity.* Cambridge University Press.

ALAN J LAUB (2005). *Matrix analysis for scientists and engineers.* Siam.

MICHAEL MITZENMACHER & ELI UPFAL (2005). *Probability and computing: Randomized algorithms and probabilistic analysis.* Cambridge University Press.

ANDREW CHI-CHIH YAO (1979). Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *11h Annual ACM Symposium on Theory of Computing*, 209–213.

Andrew Chi-Chin Yao (1977). Probabilistic computations: Toward a unified measure of complexity. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, 222–227. IEEE.

Manuscript received 16 January 2017

Badih Ghazi
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge MA 02139, USA
badih@mit.edu

Ilan Komargodski
Department of Computer Science and Applied Mathematics,
Weizmann Institute of Science
Rehovot 76100, Israel
ilan.komargodski@weizmann.ac.il

Pravesh K. Kothari
Princeton University and IAS
Princeton, NJ, USA
kothari@cs.princeton.edu

Madhu Sudan
Harvard John A. Paulson School of Engineering and Applied Sciences
33 Oxford Street
Cambridge, MA 02138, USA
madhu@cs.harvard.edu