Electronic Colloquium on Computational Complexity, Revision 8 of Report No. 190 (2013)

Toward Better Formula Lower Bounds: An Information Complexity Approach to the KRW Composition Conjecture

Dmitry Gavinsky[*] Or Meir[†] Omri Weinstein[‡] Avi Wigderson[§]

May 29, 2014

Abstract

One of the major open problems in complexity theory is proving super-logarithmic lower bounds on the depth of circuits (i.e., $\mathbf{P} \not\subseteq \mathbf{NC}^1$). This problem is interesting for two reasons: first, it is tightly related to understanding the power of parallel computation and of small-space computation; second, it is one of the first milestones toward proving super-polynomial circuit lower bounds.

Karchmer, Raz, and Wigderson [KRW95] suggested to approach this problem by proving the following conjecture: given two boolean functions f and g, the depth complexity of the composed function $g \circ f$ is roughly the sum of the depth complexities of f and g. They showed that the validity of this conjecture would imply that $\mathbf{P} \not\subseteq \mathbf{NC}^1$.

As a starting point for studying the composition of functions, they introduced a relation called "the universal relation", and suggested to study the composition of universal relations. This suggestion proved fruitful, and an analogue of the KRW conjecture for the universal relation was proved by Edmonds et. al. [EIRS01]. An alternative proof was given later by Håstad and Wigderson [HW93]. However, studying the composition of functions seems more difficult, and the KRW conjecture is still wide open.

In this work, we make a natural step in this direction, which lies between what is known and the original conjecture: we show that an analogue of the conjecture holds for the composition of a function with a universal relation. We also suggest a candidate for the next step and provide initial results toward it.

Our main technical contribution is developing an approach based on the notion of *information complexity* for analyzing KW relations – communication problems that are closely related to questions on circuit depth and formula complexity. Recently, information complexity has proved to be a powerful tool, and underlined some major progress on several long-standing open problems in communication complexity. In this work, we develop general tools for analyzing the information complexity of KW relations, which may be of independent interest.

[*] Institute of Mathematics, Academy of Sciences, Žitna 25, Praha 1, Czech Republic. Partially funded by the grant P202/12/G061 of GA ČR and by RVO: 67985840. Part of this work was done while DG was visiting the CQT at the National University of Singapore, and was partially funded by the Singapore Ministry of Education and the NRF.

[†] Institute for Advanced Study, Princeton, NJ, USA. ormeir@ias.edu. This material is based upon work supported by the National Science Foundation under agreement No. DMS-1128155, and is also partially supported by NSF grant CCF-0832797. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[‡] Department of Computer Science, Princeton University, Princeton, NJ, USA. oweinste@cs.princeton.edu. Research supported by a Simons Fellowship in Theoretical Computer Science.

[§] Institute for Advanced Study, Princeton, NJ, USA. Partially supported from NSF grant CCF-0832797

ISSN 1433-8092

1 Introduction

One of the holy grails of complexity theory is showing that **NP** cannot be computed by polynomial-size circuits, namely, that $\mathbf{NP} \not\subseteq \mathbf{P}/\mathrm{poly}$. Unfortunately, it currently seems that even finding a function in **NP** that cannot be computed by circuits of linear size is beyond our reach. Thus, it makes sense to try to prove lower bounds against weaker models of computation, in the hope that such study would eventually lead to lower bounds against general circuits.

This paper focuses on (de-Morgan) formulas, which are one such weaker model. Intuitively, formulas model computations that cannot store intermediate results. Formally, they are defined as circuits with AND, OR, and NOT gates that have fan-out 1, or in other words, their underlying graph is a tree.

For our purposes, it is useful to note that formulas are polynomially related to circuits[1] of depth $O(\log n)$: It is easy to show that circuits of depth $O(\log n)$ can be converted into formulas of polynomially-related size. On the other hand, every formula of size s can be converted into a formula of depth $O(\log s)$ and size $\mathrm{poly}(s)$ [Spi71, Bre74, BB94]. In particular, the complexity class[2] $\mathbf{NC}^1$ can be defined both as the class of polynomial-size formulas, and as the class of polynomial-size circuits of depth $O(\log n)$.

It is a major open problem to find an explicit function that requires formulas of super-polynomial size, that is, to prove that $\mathbf{P} \not\subseteq \mathbf{NC}^1$. In fact, even proving that[2] $\mathbf{NEXP} \not\subseteq \mathbf{NC}^1$ would be a big breakthrough. The state-of-the-art in this direction is the work of Håstad [Hås98], which provided an explicit function whose formula complexity is $n^{3-o(1)}$ (following the work of Andreev [And87]). Improving over this lower bound is an important challenge.

One strategy for separating **P** from $\mathbf{NC}^1$ was suggested by Karchmer, Raz, and Wigderson [KRW95]. They made a conjecture on the depth complexity of composition, and showed that this conjecture implies that $\mathbf{P} \not\subseteq \mathbf{NC}^1$. In order to introduce their conjecture, we need some notation:

Definition 1.1 (Composition)**.** Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ be boolean functions. Their composition $g \circ f : (\{0,1\}^n)^m \to \{0,1\}$ is defined by

$$(g \circ f)(x_1, \ldots, x_m) \stackrel{\mathrm{def}}{=} g(f(x_1), \ldots, f(x_m)),$$

where $x_1, \ldots, x_m \in \{0,1\}^n$.

Definition 1.2 (Depth complexity)**.** Let $f : \{0,1\}^n \to \{0,1\}$. The depth complexity of f, denoted $\mathsf{D}(f)$, is the smallest depth of a circuit of fan-in 2 that computes f using AND, OR and NOT gates.

Conjecture 1.3 (The KRW conjecture [KRW95])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ be non-constant functions. Then*[3]

$$\mathsf{D}(g \circ f) \approx \mathsf{D}(g) + \mathsf{D}(f). \tag{1}$$

[1] All the circuits in this paper are assumed to have constant fan-in.

[2] In this paper, $\mathbf{NC}^1$ always denotes the *non-uniform* version of $\mathbf{NC}^1$, which is sometimes denoted $\mathbf{NC}^1/\mathrm{poly}$.

[3] The meaning of "approximate equality" in Equation 1 is left vague, since there are a few variations that could be useful, some of which are considerably weaker than strict equality. In particular, proving either of the following lower bounds would imply that $\mathbf{P} \not\subseteq \mathbf{NC}^1$:

$$\begin{aligned}\mathsf{D}(g \circ f) &\geq \varepsilon \cdot \mathsf{D}(g) + \mathsf{D}(f)\\ \mathsf{D}(g \circ f) &\geq \mathsf{D}(g) + \varepsilon \cdot \mathsf{D}(f).\end{aligned}$$

It is also sufficient to prove the first inequality for a random g, or the second inequality for a random f.

As noted above, [KRW95] showed that this conjecture could be used to prove that $\mathbf{P} \not\subseteq \mathbf{NC}^1$: the basic idea is that one could apply $O(\log n)$ compositions of a random function $f : \{0,1\}^{\log n} \to \{0,1\}$, thus obtaining a new function over n bits that is computable in polynomial time yet requires depth $\tilde{\Omega}(\log^2 n)$. The key point here is that a random function on $\log n$ bits has depth complexity $\log n - o(\log n)$, and can be described explicitly using n bits. An interesting feature of this argument is that it does not seem to fall[4] into the framework of "natural proofs" of [RR97].

In this paper, we make a natural step toward proving the KRW conjecture, using a new information-theoretic approach[5]. We also suggest a candidate for the next step, and provide some initial results toward it. The rest of this introduction is organized as follows: In Section 1.1, we review the background relevant to our results. In Section 1.2, we describe our main result and our techniques. In Section 1.3, we describe our candidate for the next step, and our initial results in this direction.

1.1 Background

1.1.1 Karchmer-Wigderson relations

Karchmer and Wigderson [KW90] observed an interesting connection between depth complexity and communication complexity: for every boolean function f, there exists a corresponding communication problem R_f, such that the depth complexity of f is equal to the deterministic[6] communication complexity of R_f. The communication problem R_f is often called the Karchmer-Wigderson relation of f, and we will refer to it as a KW relation for short. In fact, a stronger statement is implicit in [KW90]:

Fact 1.4 ([KW90])**.** *For every formula ϕ that computes f, there exists a deterministic protocol Π_ϕ for R_f, whose underlying tree is exactly the underlying tree of f, and vice versa.*

A corollary of Fact 1.4 that is particularly useful for us is the following: the formula size of f is exactly the minimal number of distinct transcripts in every protocol that solves R_f.

The communication problem R_f is defined as follows: Alice gets an input $x \in f^{-1}(0)$, and Bob gets as input $y \in f^{-1}(1)$. Clearly, it holds that $x \neq y$. The goal of Alice and Bob is to find a coordinate i such that $x_i \neq y_i$. Note that there may be more than one possible choice for i, which means that R_f is a relation rather than a function.

This connection between functions and KW relations allows us to study the formula and depth complexity of functions using techniques from communication complexity. In the past, this approach has proved very fruitful in the setting of *monotone* formulas [KW90, GS91, RW92, KRW95], and in particular [KRW95] used it to prove a monotone analogue of the KRW conjecture.

During the last decade, a new kind of information-theoretic techniques have emerged in the area of communication complexity [CSWY01, JRS03, BJKS04, JKS03, DW07, HJMR10, BBCR10, BR11, Bra12, BW12, KLL$^+$12], which were called "interactive information complexity" by [BBCR10, Bra12]. While these techniques have existed for some time, in the last few years they have drawn much interest, and a significant and rapid progress in their study is being made. These techniques have been especially useful in attacking another "economy of scale" problem, namely, the direct sum problem conjecture. One of the contributions of this work is showing how some of those ideas can be applied in the setting of KW relations (see Section 3).

[4] More specifically, it seems that this argument violates the largeness property, because it only proves a lower bound for a specific, artificially constructed function, rather than for a random function.

[5] We note that the works [KW90, EIRS01] on the KRW conjecture also use a (different) information-theoretic argument.

[6] In this paper, we always refer to *deterministic* communication complexity, unless stated explicitly otherwise.

1.1.2 KW relations and the KRW conjecture

In order to prove the KRW conjecture, one could study the KW relation that corresponds to the composition $g \circ f$. Let us describe how the KW relation $R_{g \circ f}$ looks like. Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$. For every $m \times n$ matrix X, let us denote by $f(X)$ the vector in $\{0,1\}^m$ obtained by applying f to each row of X. In the KW relation $R_{g \circ f}$, Alice and Bob get as inputs $m \times n$ matrices X, Y, respectively, such that $f(X) \in g^{-1}(0)$ and $f(Y) \in g^{-1}(1)$, and their goal is to find an entry (j, i) such that $X_{j,i} \neq Y_{j,i}$.

Let us denote the (deterministic) communication complexity of a problem R by $\mathsf{C}(R)$. Clearly, it holds that

$$\mathsf{C}(R_{g \circ f}) \leq \mathsf{C}(R_g) + \mathsf{C}(R_f). \tag{2}$$

This upper bound is achieved by the following protocol: For every $j \in [m]$, let X_j denote the j-th row of X, and same for Y. Alice and Bob first use the optimal protocol of g on inputs $f(X)$ and $f(Y)$, and thus find an index $j \in [m]$ such that $f(X_j) \neq f(Y_j)$. Then, they use the optimal protocol of f on inputs $f(X_j)$ and $f(Y_j)$ to find a coordinate i on which the j-th rows differ, thus obtaining an entry (j, i) on which X and Y differ.

The KRW conjecture says that the above protocol is essentially optimal. One intuition for that conjecture is the following: the best way for Alice and Bob to solve $R_{g \circ f}$ is to solve R_f on some row j such that $f(X_j) \neq f(Y_j)$, since otherwise they are not using the guarantee they have on X and Y. However, in order to do that, they must find such a row j, and to this end they have to solve R_g. Thus, they have to transmit $\mathsf{C}(R_g)$ bits in order to find j, and another $\mathsf{C}(R_f)$ bits to solve f on the j-th row.

This intuition was made rigorous in the proof of the monotone version of the KRW conjecture [KRW95], who used it to prove the monotone version of $\mathbf{P} \not\subseteq \mathbf{NC}^1$. A similar intuition underlies our argument, as well as the works of [EIRS01, HW93] that are to be discussed later.

1.1.3 The universal relation and its composition

Since proving the KRW conjecture seems difficult, [KRW95] suggested studying a simpler problem as a starting point. To describe this simpler problem, we first need to define a communication problem called the universal relation, and its composition with itself. The universal relation R_{U_n} is a communication problem in which Alice and Bob get as inputs $x, y \in \{0,1\}^n$ with the sole guarantee that $x \neq y$, and their goal is to find a coordinate i such that $x_i \neq y_i$. The universal relation R_{U_n} is universal in the sense that every KW relation reduces to it, and indeed, it is not hard to prove that $\mathsf{C}(R_{\mathrm{U}_n}) \geq n$.

The composition of two universal relations R_{U_m} and R_{U_n}, denoted $R_{\mathrm{U}_m \circ \mathrm{U}_n}$, is defined as follows. Alice gets as an input an $m \times n$ matrix X and a string $a \in \{0,1\}^m$, and Bob gets as an input an $m \times n$ matrix Y and a string $b \in \{0,1\}^m$. Their inputs satisfy the following conditions:

1. $a \neq b$.
2. for every $j \in [n]$ such that $a_j \neq b_j$, it holds that $X_j \neq Y_j$.

Their goal, as before, is to find an entry on which X and Y differ. The vectors a and b are analogues of the vectors $f(X)$ and $f(Y)$ in the KW relation $R_{g \circ f}$.

To see why $R_{\mathrm{U}_m \circ \mathrm{U}_n}$ is a good way to abstract the KRW conjecture, observe that $R_{\mathrm{U}_m \circ \mathrm{U}_n}$ is a universal version of composition problems $R_{g \circ f}$, in the sense that every composition problem $R_{g \circ f}$ reduces to $R_{\mathrm{U}_m \circ \mathrm{U}_n}$. Moreover, the protocol described above for $R_{g \circ f}$ also works for $R_{\mathrm{U}_m \circ \mathrm{U}_n}$: Alice and Bob first apply the optimal protocol for R_{U_m} to a and b to find j, and then apply the optimal protocol for R_{U_n} to X_j and Y_j. Thus, a natural variant of the KRW conjecture for this protocol would be that this protocol is optimal for $R_{\mathrm{U}_m \circ \mathrm{U}_n}$. Following this reasoning, [KRW95] suggested to prove that

$$\mathsf{C}(R_{\mathrm{U}_m \circ \mathrm{U}_n}) \approx \mathsf{C}(R_{\mathrm{U}_m}) + \mathsf{C}(R_{\mathrm{U}_n}) \geq m + n \tag{3}$$

as a first step toward proving the KRW conjecture. This challenge was met[7] by [EIRS01] up to a small additive loss, and an alternative proof was given later in [HW93]. Since then, there has been no further progress on the KRW conjecture for about two decades.

1.2 Our main result: The composition of a function with the universal relation

Summing up, the KRW conjecture is about the composition of two functions $R_{g \circ f}$, but it was only known how to prove it for the composition of two universal relations $R_{U_m \circ U_n}$. In this work we go a step further: We prove an analogue of the KRW conjecture for relations of the form $R_{g \circ \mathrm{U}_n}$, where $g \in \{0,1\}^m \to \{0,1\}$ is an arbitrary function; and where $R_{g \circ \mathrm{U}_n}$ is a problem that can be naturally viewed as the composition of g with the universal relation.

We define the communication problem $R_{g \circ \mathrm{U}_n}$ as follows. Alice gets as an input an $m \times n$ matrix X and a string $a \in g^{-1}(0)$, and Bob gets as an input an $m \times n$ matrix Y and a string $b \in g^{-1}(1)$. Their inputs are guaranteed to satisfy Condition 2 of $R_{\mathrm{U}_m \circ \mathrm{U}_n}$, i.e., for every $j \in [n]$ such that $a_j \neq b_j$, it holds that $X_j \neq Y_j$. Clearly, their inputs also satisfy $a \neq b$, as in Condition 1 of $R_{\mathrm{U}_m \circ \mathrm{U}_n}$. The goal of Alice and Bob, as usual, is to find an entry on which X and Y differ.

Note that $R_{g \circ U_n}$ is universal, in the sense that for any $f : \{0,1\}^n \to \{0,1\}$, the communication problem $R_{g \circ f}$ reduces to $R_{g \circ \mathrm{U}_n}$. An analogue of the KRW conjecture for $R_{g \circ U_n}$ would be

$$\mathsf{C}(R_{g \circ U_n}) \approx \mathsf{C}(R_g) + \mathsf{C}(R_{\mathrm{U}_n}) \geq \mathsf{C}(R_g) + n. \tag{4}$$

We prove the following closely related result.

Theorem 1.5. *Let $m, n \in \mathbb{N}$, and let $g : \{0,1\}^m \to \{0,1\}$ be a non-constant function. Then,*

$$\mathsf{C}(R_{g \circ \mathrm{U}_n}) \geq \Omega\left(\mathsf{C}(R_g)\right) + n - O(1 + \frac{m}{n}) \cdot \log m.$$

In fact, we obtain Theorem 1.5 as a corollary of the following theorem, which gives a tighter bound in terms of formula complexity. Let $\mathsf{L}(g)$ denote the formula complexity of g, and recall that $\log \mathsf{L}(g) \geq \Omega\left(\mathsf{C}(R_g)\right)$ due to the correspondence between formula size and circuit depth. We have the following result.

Theorem 1.6 (Main theorem)**.** *Let $m, n \in \mathbb{N}$, and let $g : \{0,1\}^m \to \{0,1\}$ be a non-constant function. Then,*

$$\mathsf{C}(R_{g \circ \mathrm{U}_n}) \geq \log \mathsf{L}(g) + n - O(1 + \frac{m}{n}) \cdot \log m,$$

Moreover, the same lower bound applies to the logarithm of the number of leaves of any protocol for $R_{g \circ \mathrm{U}_n}$ (which is the "formula complexity" of $R_{g \circ \mathrm{U}_n}$).

There is a good reason why the formula complexity $\mathsf{L}(g)$ appears in Theorem 1.6, as will be made clear in the following discussion on our techniques.

Remark 1.7. In the target application of the KRW conjecture, namely the proof that $\mathbf{P} \not\subseteq \mathbf{NC}^1$, the parameters can be chosen such that $m \ll n$, so the loss of $O(1 + \frac{m}{n}) \cdot \log m$ in Theorem 1.6 is not very important.

[7] In fact, they only consider the case where $m = n$, but their argument should generalize to the case where $m \neq n$.

Remark 1.8. We note that Theorem 1.6 also implies a lower bound on the composition $R_{U_m \circ U_n}$ of two universal relations, thus giving a yet another proof for the results of [EIRS01, HW93]. In fact, our techniques can be used to give a simpler proof for those results.

1.2.1 Our techniques

As mentioned above, our techniques use some ideas from the information complexity literature. In particular, we use a complexity measure called *(external) information cost*, which measures the amount of information that the protocol leaks on the inputs of Alice and Bob. Our starting point is the observation that (the logarithm of) the size of a formula ϕ for any function f can be reinterpreted as the information cost of the corresponding protocol for R_f.

To see why this is helpful, consider the KW relation $R_{g \circ \mathrm{U}_n}$. Intuitively, we would like to argue that in order to solve $R_{g \circ \mathrm{U}_n}$, Alice and Bob must solve R_g (incurring a cost of $\mathsf{C}(R_g)$), and also solve the universal relation on one of the rows their matrices (incurring a cost of n). Such an argument requires decomposing the communication of Alice and Bob into communication "about" R_g and communication "about" R_{U_n}. However, it is not clear how to do that, because Alice and Bob may "talk" simultaneously about R_g and R_{U_n} (e.g. by sending the XOR of a bit of a and a bit of X).

On the other hand, when considering the *information* transmitted by Alice and Bob, such a decomposition comes up naturally: the information that Alice and Bob transmit can be decomposed, using the chain rule, into the information they transmit on the strings a, b (which are inputs of R_g) and the information they transmit on the matrices X and Y (which consist of inputs of R_{U_n}). We now derive the required lower bound

$$\mathsf{C}(R_{g \circ \mathrm{U}_n}) \geq \log \mathsf{L}(g) + n - O(1 + \frac{m}{n}) \cdot \log m,$$

as follows: the information about a and b contributes $\log \mathsf{L}(g)$ (which is the information cost of R_g); and the information about X and Y contributes n (which is the information cost of R_{U_n}). Of course, implementing this argument is far from trivial, and in particular, we do not know how to extend this argument to the full KRW conjecture, i.e., KW relations of the form $R_{g \circ f}$.

This is reminiscent of a similar phenomenon in the literature about the direct sum problem in communication complexity (e.g., [BBCR10]): the direct sum problem asks whether solving k independent instances of a function is k times harder than solving a single instance. The reason that information complexity is useful for studying this question is that there, too, the information transmitted by the protocol can be decomposed, using the chain rule, to the information about each of the independent instances.

This suggests that information complexity may be the "right" tool to study the KRW conjecture. In particular, since in the setting of KW relations, the information cost is analogous to the formula size, the "correct" way to state the KRW conjecture may be using formula size:

$$\mathsf{L}(g \circ f) \approx \mathsf{L}(g) \cdot \mathsf{L}(f).$$

Interestingly, the KRW conjecture is supported by the works of [And87, Hås98], which prove that

$$\mathsf{L}(g \circ \oplus_m) = \mathsf{L}(g) \cdot \tilde{\Omega}\left(m^2\right) = \tilde{\Omega}\left(\mathsf{L}(g) \cdot \mathsf{L}(\oplus_m)\right),$$

where $\oplus_m$ is the parity function of m bits and g is an arbitrary function, and where the second equality follows from [Khr72].

We note that an additional contribution of this work is developing some basic generic tools for working with information complexity in the setting of KW relations (see Sections 3 and 5.2).

1.2.2 On hard distributions

One significant difference between our work and previous works on information complexity and direct sum (e.g. [BBCR10]) is the following: In order to define the information complexity of a communication problem, one must specify a distribution on the inputs. The reason is information-theoretic notions such as entropy are only defined with respect to a distribution. The previous works use distributions that are *protocol independent*, that is, they first choose a distribution μ, and then prove that every protocol π for the problem must have a large information cost with respect to μ.

In the setting of KW relations, this is impossible: for every distribution μ there exists a protocol π that has a small information cost with respect to μ (as discussed in Section 3.2). This can be derived from the fact that KW relations have very efficient randomized protocols. Therefore, the only way to apply information-complexity techniques to KW relations is to use *protocol-dependent* distributions, that is, to tailor a different distribution for each protocol.

1.2.3 A combinatorial proof

After discovering the proof of the main result using information complexity, we found that the same proof can be rephrased as a combinatorial "double counting" argument, without making any reference to information theory. We believe that both formulations give useful perspectives on the proof. Therefore, while most of the paper is focused on the information-complexity approach, we also provide a self-contained combinatorial proof in Section 6.

1.3 A candidate for the next step: The composition $\oplus_m \circ f$

In order to make further progress toward the KRW conjecture, we would like to replace the universal relation by a function. One possible approach to this question would be to start with compositions $g \circ f$ where g is a some known simple function. Perhaps the simplest such example is the composition $\vee_m \circ f$, where $\vee_m$ is the disjunction of m bits, and f is an arbitrary function. For this example, an analogue of the KRW conjecture is already known, that is,

$$\mathsf{L}(\vee_m \circ f) = \mathsf{L}(\vee_m) \cdot \mathsf{L}(f) = m \cdot \mathsf{L}(f)$$

(see, e.g., [Weg87, Chapter 10], and also discussion in Section 3.2.1 below). The next simplest example would be $\oplus_m \circ f$, where $\oplus_m$ is the parity of m bits. For this example, an analogue of the KRW conjecture would be

$$\mathsf{L}(\oplus_m \circ f) \approx \mathsf{L}(\oplus_m) \cdot \mathsf{L}(f) = m^2 \cdot \mathsf{L}(f), \tag{5}$$

where the second equality follows from [Khr72]. We therefore suggest the following conjecture as a next step toward the KRW conjecture:

Conjecture 1.9. *For every function $f : \{0,1\}^n \to \{0,1\}$ and every $m \in \mathbb{N}$, it holds that*

$$\mathsf{L}(\oplus_m \circ f) = \tilde{\Omega}\left(m^2 \cdot \mathsf{L}(f)\right).$$

We note that Conjecture 1.9 is not only interesting as a step toward the KRW conjecture, but is also interesting on its own right. In particular, if Conjecture 1.9 is proved, it will yield an alternative proof of the state-of-the-art lower bound of of $\tilde{\Omega}(n^3)$ by [Hås98]. The lower bound is proved as follows: consider the following function F, which gets as input the truth table of a function $f : \{0,1\}^n \to \{0,1\}$ and $m \overset{\text{def}}{=} \frac{2^n}{n}$ instances for f:

$$F(f, x_1, \ldots, x_m) \overset{\text{def}}{=} (\oplus_m \circ f)(x_1, \ldots, x_m) = \bigoplus_{j=1}^{m} f(x_j).$$

The input length of F is $N = 2^{n+1}$. Since the function f could be fixed to be maximally hard function with formula complexity $2^n/\log n$, Conjecture 1.9 implies that this function has formula complexity

$$\mathsf{L}(F) = \tilde{\Omega}\left(m^2 \cdot \mathsf{L}(f)\right) = \tilde{\Omega}\left(\left(\frac{2^n}{n}\right)^2 \cdot \frac{2^n}{\log n}\right) = \tilde{\Omega}(N^3),$$

as required. We note that this construction of a hard function F is a twist of an argument of Andreev [And87], which was also used by [Hås98]. In particular, the only difference is that [And87] uses $f \circ \oplus_m$, rather than $\oplus_m \circ f$, in the definition of F. However, this is a significant difference: when the parity function is "at the bottom" (as in $f \circ \oplus_m$), one can easily apply random restrictions to the function, but it is not clear how to do it when the parity function is "at the top" (as in $\oplus_m \circ f$)

In this work, we provide two preliminary results toward proving Conjecture 1.9:

- **A lower bound for $R_{\oplus_m \circ U_n}$:** A natural first step toward proving Conjecture 1.9 would be to prove a corresponding lower bound on $R_{\oplus_m \circ \mathrm{U}_n}$, the composition of parity with the universal relation. Though in principle we could apply our main theorem with $g = \oplus_m$, in this case it would not give a meaningful lower bound (due to the error term of $O(1 + \frac{m}{n}) \cdot \log m$).
  One contribution of this work is proving the following *almost tight* analogue of Conjecture 1.9 for $R_{\oplus_m \circ \mathrm{U}_n}$:

  Theorem 1.10. *For every $m, n \in \mathbb{N}$ it holds that*

  $$\mathsf{C}(R_{\oplus_m \circ U_n}) \geq 2 \log m + n - O(\log \log m).$$

  Moreover, the same lower bound applies to the logarithm of the number of leaves of any protocol for $R_{\oplus_m \circ U_n}$ (which is the "formula complexity" of $R_{\oplus_m \circ \mathrm{U}_n}$).

- **A candidate hard distribution:** We would like to use information-complexity techniques in order to study the KW relation $R_{\oplus_m \circ f}$. In order to define the information complexity of a protocol, we must first define an appropriate distribution over the inputs. We would therefore like to find a "hard distribution" over the inputs of $R_{\oplus_m \circ f}$ that will have a large information cost. As discussed in Section 1.2.2, this requires tailoring a different hard distribution for each protocol.
  We propose a candidate hard distribution for each protocol for $R_{\oplus_m \circ f}$. While we do not know how to prove that this distribution is indeed hard, it has some appealing and non-trivial properties that we expect a hard distribution to have (see Section 5.1 for further discussion).

Another open problem. An additional natural step toward the KRW conjecture would be to prove a lower bound for the relation $R_{U_m \circ f}$, defined as follows: Let $f : \{0,1\}^n \to \{0,1\}$. Alice and Bob get $m \times n$ matrix X and Y, respectively, and they are guaranteed that $f(X) \neq f(Y)$.

Their goal, as always, is to find an entry on which X and Y differ. The natural conjecture for this relation would be

$$\mathsf{C}(R_{U_m \circ f}) \approx \mathsf{C}(R_f) + n.$$

Proving such a lower bound would complement our result, and is perhaps a less ambitious goal than Conjecture 1.9.

Organization of this paper

In Section 2, we review the required preliminaries. Then, in Section 3, we discuss how one can apply techniques from information complexity in the setting of KW relations, and make some useful general observations. Next, in Section 4, we prove our main result (Theorem 1.6), as well as the lower bound for the special case $R_{\oplus_m \circ U_n}$ (Theorem 1.10). In Section 5, we construct our candidate hard distribution for the KW relation $R_{\oplus_m \circ f}$. Finally, in Section 6, we provide a self-contained proof of our main result that uses a combinatorial double-counting argument rather than an information-theoretic argument.

2 Preliminaries

We reserve bold letters for random variables, and calligraphic letters for sets. We use $[n]$ to denote the set $\{1, \ldots, n\}$. For a function $f : \mathbb{N} \to \mathbb{N}$, we denote

$$\begin{aligned} \tilde{O}(f) &\overset{\text{def}}{=} O(f \cdot \log^{O(1)} f) \\ \tilde{\Omega}(f) &\overset{\text{def}}{=} \Omega(f / \log^{O(1)} f). \end{aligned}$$

We denote the set of $m \times n$ binary matrices by $\{0,1\}^{m \times n}$. For every binary $m \times n$ matrix X, we denote by $X_j \in \{0,1\}^n$ the j-th row of X. Throughout the paper, we denote by $\oplus_m$ the parity function over m bits.

2.1 Formulas

Definition 2.1. A formula ϕ is a binary tree, whose leaves are identified with literals of the forms x_i and $\neg x_i$, and whose internal vertices are labeled as AND ($\wedge$) or OR ($\vee$) gates. A formula ϕ computes a binary function $f : \{0,1\}^n \to \{0,1\}$ in the natural way. The size of a formula is the number of its *leaves* (which is the same as the number of its wires up to a factor of 2). We note that a single input coordinate x_i can be associated with many leaves.

Definition 2.2. The formula complexity of a boolean function $f : \{0,1\}^n \to \{0,1\}$, denoted $\mathsf{L}(f)$, is the size of the smallest formula that computes f. The depth complexity of f, denoted $\mathsf{D}(f)$, is the smallest depth of a formula that computes f.

The following theorem establishes a tight connection between the formula complexity and the depth complexity of a function.

Theorem 2.3 ([BB94], following [Spi71, Bre74])**.** *For every $\alpha > 1$ the following holds: For every formula ϕ of size s, there exists an equivalent formula ϕ' of depth $O(\log s)$ and size s^{α}. The constant in the Big-O notation depends on α.*

Remark 2.4. Note that we define here the depth complexity of a function by as the depth of a *formula* that computes f, while in the introduction we defined it as the depth of a *circuit* that computes f. However, for our purposes, this distinction does not matter, since every circuit of depth $O(\log n)$ can be transformed into a formula of the same depth and of polynomial size.

2.2 Communication complexity

Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be sets, and let $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. The communication problem [Yao79] that corresponds to R is the following: two players, Alice and Bob, get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively. They would like to communicate and find $z \in \mathcal{Z}$ such that $(x, y, z) \in R$. At each round, one of the players sends a bit that depends on her/his input and on the previous messages, until they find z. The communication complexity of R is the minimal number of bits that is transmitted by a protocol that solves R. More formally, we define a protocol as a binary tree, in which every vertex represents a possible state of the protocol, and every edge represents a message that moves the protocol from one state to another:

Definition 2.5. A (deterministic) protocol that solves a relation $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is a rooted binary tree with the following structure:

- Every node of the tree is labeled by a rectangle $\mathcal{X}_v \times \mathcal{Y}_v$ where $\mathcal{X}_v \subseteq \mathcal{X}$ and $\mathcal{Y}_v \subseteq \mathcal{Y}$. The root is labeled by the rectangle $\mathcal{X} \times \mathcal{Y}$. Intuitively, the rectangle $\mathcal{X}_v \times \mathcal{Y}_v$ is the set of pairs of inputs that lead the players to the node v.

- Each internal node v is *owned* by Alice or by Bob. Intuitively, v is owned by Alice if at state v, it is Alice's turn to speak, and same for Bob.

- Every edge of the tree is labeled by either 0 or 1.

- For every internal node v that is owned by Alice, the following holds: Let v_0 and v_1 be the children of v associated with the out-going edges labeled with 0 and 1, respectively. Then,

  - $\mathcal{X}_v = \mathcal{X}_{v_0} \cup \mathcal{X}_{v_1}$, and $\mathcal{X}_{v_0} \cap \mathcal{X}_{v_1} = \emptyset$.
  - $\mathcal{Y}_v = \mathcal{Y}_{v_0} = \mathcal{Y}_{v_1}$.

  Intuitively, when the players are at the vertex v, Alice transmits 0 if her input is in $\mathcal{X}_{v_0}$ and 1 if her input is in $\mathcal{X}_{v_1}$. An analogous property holds for notes owned by Bob, while changing the roles of $\mathcal{X}$ and $\mathcal{Y}$.

- For each leaf ℓ, there exists a value z such that $\mathcal{X}_\ell \times \mathcal{Y}_\ell \times \{z\} \subseteq R$. Intuitively, z is the output of the protocol at ℓ.

Definition 2.6. The communication complexity of a protocol Π, denoted $\mathsf{C}(\Pi)$, is the the depth of the protocol tree. In other words, it is the maximum number of bits that can be transmitted in an invocation of the protocol on any pair of inputs (x, y). For a relation R, we denote by $\mathsf{C}(R)$ the minimal communication complexity of a (deterministic) protocol that solves R.

Definition 2.7. Given a protocol Π, the transcript $\Pi(x, y)$ is the string that consists of the messages of Alice and Bob in the protocol when they get the inputs x and y, respectively. More formally, observe that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, there is a unique leaf ℓ such that $(x, y) \in \mathcal{X}_\ell \times \mathcal{Y}_\ell$. The transcript $\Pi(x, y)$ is the string that is obtained by concatenating the labels of the edges on the path from the root to the leaf ℓ . We will sometimes identify $\Pi(x, y)$ with the leaf ℓ itself.

We now define a notion of protocol size that is analogous to the notion of formula size.

Definition 2.8. We define the size of a protocol Π to be its number of leaves. Note that this is also the number of distinct transcripts of the protocol. We define the protocol size of a relation R, denoted $\mathsf{L}(R)$, as the size of the smallest protocol that solves it.

We will sometimes invoke a protocol Π on inputs that are random variables $\mathbf{x}, \mathbf{y}$. In such a case, the transcript is a random variable as well. With some abuse of notation, we will use $\Pi \stackrel{\text{def}}{=} \Pi(\mathbf{x}, \mathbf{y})$ to denote this random transcript.

2.3 Karchmer-Wigderson relations

In this section, we define KW relations formally, and give a sketch of the correspondence between KW relations and formulas. In addition, in Section 2.3.1, we introduce a useful generalization of KW relations, which we call "relaxed KW problems".

Definition 2.9. Let $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ be two disjoint sets. The KW relation $R_{\mathcal{X},\mathcal{Y}} \subseteq \mathcal{X} \times \mathcal{Y} \times [n]$ is defined by

$$R_{\mathcal{X},\mathcal{Y}} \stackrel{\text{def}}{=} \{(x, y, i) : x_i \neq y_i\}$$

Intuitively, $R_{\mathcal{X},\mathcal{Y}}$ corresponds to the communication problem in which Alice gets $x \in \mathcal{X}$, Bob gets $y \in \mathcal{Y}$, and they would like to find a coordinate $i \in [n]$ such that $x_i \neq y_i$ (note that $x \neq y$ since $\mathcal{X} \cap \mathcal{Y} = \emptyset$).

Definition 2.10. Let $f : \{0,1\}^n \to \{0,1\}$ be a non-constant function. The KW relation of f, denoted R_f, is defined by $R_f \stackrel{\text{def}}{=} R_{f^{-1}(0), f^{-1}(1)}$.

Definition 2.11. Let $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ be two disjoint sets. We say that a formula ϕ separates $\mathcal{X}$ and $\mathcal{Y}$ if $\phi(\mathcal{X}) = 0$ and $\phi(\mathcal{Y}) = 1$.

Theorem 2.12 (Implicit in [KW90]). *Let $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ be two disjoint sets. Then, for every formula ϕ that separates $\mathcal{X}$ and $\mathcal{Y}$, there exists a protocol Π_ϕ that solves $R_{\mathcal{X},\mathcal{Y}}$, whose underlying tree is the same as the underlying tree of ϕ. In the other direction, for every protocol Π that solves $R_{\mathcal{X},\mathcal{Y}}$ there exists a formula ϕ_Π that separates $\mathcal{X}$ and $\mathcal{Y}$, whose underlying is tree the same as the underlying tree of Π.*

Proof. For the first direction, let ϕ be a formula such that separates $\mathcal{X}$ and $\mathcal{Y}$. We construct Π_ϕ by induction: if ϕ is of size 1, then ϕ is a single literal of the form x_i or $\neg x_i$. This implies that all the strings in $\mathcal{X}$ differ from all the strings in $\mathcal{Y}$ on the coordinate i. Therefore, we define Π_ϕ as the protocol in which the players do not interact, and always output i. Note that the protocol tree Π_ϕ indeed has the same structure as the tree of ϕ.

Next, assume that $\phi = \phi_0 \wedge \phi_1$ (if $\phi = \phi_0 \vee \phi_1$ the construction is analogous). Let us denote by $\mathcal{X}_0$ and $\mathcal{X}_1$ the sets of strings x such that $\phi_0(x) = 0$ and $\phi_1(x) = 0$, respectively, and observe that $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$. Moreover, observe that $\phi_0(\mathcal{Y}) = \phi_1(\mathcal{Y}) = 1$. We now define Π_ϕ as follows: Alice sends Bob a bit b such that her input belongs to $\mathcal{X}_b$, and then they execute the protocol Π_{ϕ_b}. It is easy to see that Π_ϕ indeed solves $R_{\mathcal{X},\mathcal{Y}}$, and that the protocol tree of Π_ϕ has the same structure as the tree of ϕ. This concludes the first direction.

For the second direction, let Π be a protocol that solves $R_{\mathcal{X},\mathcal{Y}}$. Again, we construct ϕ_Π by induction: If Π is of size 1, then it consists of a single leaf that is labeled with some coordinate i. This implies that all the strings in $\mathcal{X}$ differ from all the strings in $\mathcal{Y}$ on the coordinate i. If for all $x \in \mathcal{X}$ it holds that $x_i = 0$, we define ϕ_Π to be the literal x_i, and otherwise we define it to be the literal $\neg x_i$. Note that the tree of ϕ_Π indeed has the same structure as the tree of Π.

Next, assume that Alice speaks first at Π (if Bob speaks first, the construction is analogous). Let us denote $\mathcal{X}_0$ and $\mathcal{X}_1$ the sets of strings x on which Alice sends the bit 0 and 1 as her first message, respectively. Let Π_0 and Π_1 be the residual protocols obtained from Π by conditioning on Alice's message, and note that by induction there exist formulas ϕ_{Π_0} and ϕ_{Π_1} such that ϕ_{Π_b}

separates $\mathcal{X}_b$ and $\mathcal{Y}$. We now define $\phi_{\Pi} \stackrel{\text{def}}{=} \phi_{\Pi_0} \wedge \phi_{\Pi_1}$. It is easy to see that ϕ_{Π} indeed separates $\mathcal{X}$ and $\mathcal{Y}$, and to see that the tree of ϕ_{Π} has the same structure as the tree of Π. This concludes the second direction. ■

Corollary 2.13. *Let $f : \{0,1\}^n \to \{0,1\}$.Then, for every formula ϕ for f, there exists a protocol Π_ϕ that solves R_f whose underlying tree is the same as the underlying tree of ϕ. In the other direction, for every protocol Π that solves R_f there exists a formula ϕ_Π for f whose underlying tree the same as the underlying tree of Π.*

Corollary 2.14. *For every non-constant $f : \{0,1\}^n \to \{0,1\}$, it holds that $\mathsf{D}(f) = \mathsf{C}(R_f)$, and $\mathsf{L}(f) = \mathsf{L}(R_f)$.*

2.3.1 Relaxed Karchmer-Wigderson problems

In this section, we introduce the notion of "relaxed KW problems". Intuitively, these are KW relations that only require that the players *"almost"* find a coordinate i such that $x_i \neq y_i$. This relaxation turns out to be useful at a certain point in our proof, where we want to argue that the players have to "almost" solve a KW relation.

A bit more formally, given a boolean function $f : \{0,1\}^n \to \{0,1\}$ and a number $t \in \mathbb{N}$, the relaxed KW problem $R_f(t)$ is a communication problem in which Alice wants to find a set $\mathcal{I}$ of size less than t such that $x|_{\mathcal{I}} \neq y|_{\mathcal{I}}$. This relaxes the definition of KW relations in two ways:

1. Unlike a standard KW relation, Alice is not required to know a particular coordinate i such that $x_i \neq y_i$. Instead, she only need to isolate it to a "small" set $\mathcal{I}$. The parameter t measures the amount of uncertainty that Alice about the coordinate i.
2. Moreover, unlike a standard KW relation, we do not require that at the end of the protocol, both players know the set $\mathcal{I}$. Instead, we only require that Alice knows the set $\mathcal{I}$.

The second relaxation above implies that a "relaxed KW problem" can not be defined as a relation, in the same way we defined communication problems until this point. This leads us to the following definition of the relaxed KW problem.

Definition 2.15. Let $f : \{0,1\}^n \to \{0,1\}$ be a non-constant function and let $t \in \mathbb{N}$. Let Π be a protocol whose root is labeled by the rectangle $f^{-1}(0) \times f^{-1}(1)$. We say that Π solves the relaxed KW problem $R_f(t)$ if it satisfies the following requirement:

- For every leaf ℓ of Π that is labeled by a rectangle $\mathcal{X}_\ell \times \mathcal{Y}_\ell$, and for every $x \in \mathcal{X}_\ell$, there exists a set $\mathcal{I} \subseteq [n]$, $|\mathcal{I}| < t$, such that $x|_{\mathcal{I}} \neq y|_{\mathcal{I}}$ for every $y \in \mathcal{Y}_\ell$.

Remark 2.16. Note that in Definition 2.15, the fact that $\mathcal{I}$ is determined by both ℓ and x means that Alice knows the set $\mathcal{I}$, but Bob does not necessarily know it.

Remark 2.17. It is tempting to guess that $R_f(1)$ is the same as R_f, but it is not: in the communication problem R_f, Bob is required to know i at the end of the protocol, while in $R_f(1)$, he is not.

Remark 2.18. Definition 2.15 is inspired by the definition of k-limit by [HJP95, Definition 2.1].

We now prove the following easy proposition, which says that the relaxed KW problem $R_f(t)$ is not much easier than te original KW relation R_f.

Proposition 2.19. *Let $f : \{0,1\}^n \to \{0,1\}$, and let $t \in \mathbb{N}$. Then,*

$$\begin{aligned} \mathsf{C}(R_f(t)) &\geq \mathsf{C}(R_f) - t \cdot (\log n + 2) \\ \mathsf{L}(R_f(t)) &\geq 2^{-t \cdot (\log n + 2)} \cdot \mathsf{L}(R_f). \end{aligned}$$

Proof. We prove the proposition by reducing R_f to $R_f(t)$. Let Π be a protocol for $R_f(t)$. We show that there exists a protocol Π' for R_f such that

$$\begin{aligned} \mathsf{C}(\Pi') &\leq \mathsf{C}(R_f) + t \cdot (\log n + 2) \\ \mathsf{L}(R_f(t)) &\leq 2^{t \cdot (\log n + 2)} \cdot \mathsf{L}(R_f). \end{aligned}$$

The protocol Π' for R_f is defined as follows: When Alice and Bob get inputs x and y, respectively, they invoke the protocol Π on their inputs, thus reaching a leaf ℓ. By Definition 2.15, there exists a set $\mathcal{I} \subseteq [n]$, $|\mathcal{I}| < t$, such that $x|_{\mathcal{I}} \neq y'|_{\mathcal{I}}$ for every y' that is supported by ℓ. Alice now sends the set $\mathcal{I}$ and the string $x|_{\mathcal{I}}$ to Bob, and Bob replies with $y|_{\mathcal{I}}$. At this point, they both know a coordinate on which x and y differ, and the protocol ends.

The correctness of the protocol Π' is easy to verify. To analyze its communication complexity and size, observe that after reaching the leaf ℓ, Alice and Bob transmit at most

$$|\mathcal{I}| \cdot \log n + 2 \cdot |\mathcal{I}| < t \cdot (\log n + 2)$$

bits: $|\mathcal{I}| \cdot \log n$ bits for transmitting the set $\mathcal{I}$ itself, and another $2 \cdot |\mathcal{I}|$ bits for transmitting $a|_{\mathcal{I}}$ and $b|_{\mathcal{I}}$. This implies that the protocol tree of Π' can be obtained from the protocol tree of Π by replacing each leaf of Π with a binary tree that has at most $2^{t \cdot (\log n + 2)}$ leaves and is of depth at most $t \cdot (\log n + 2)$. The required upper bounds on $\mathsf{C}(\Pi')$ and $\mathsf{L}(\Pi')$ follow. ■

2.4 The universal relation and its compositions

In this section, we define the universal relation and its compositions formally. *We stress that the following definitions are slightly different than the ones given in the introduction:* In the definition given in the introduction, the players were promised that $x \neq y$. For example, in the following definition, they are not given this promise, but are allowed to reject if the promise does not hold. This modification was suggested by [HW93].

Definition 2.20. The universal relation R_{U_n} is defined as follows:

$$R_{\mathrm{U}_n} \overset{\text{def}}{=} \{(x,y,i) : x \neq y \in \{0,1\}^n, i \in [n], x_i \neq y_i\} \cup \{(x,x,\perp) : x \in \{0,1\}^n\}.$$

This corresponds to the communication problem in which Alice and Bob get strings x and y, respectively, and are required to output a coordinate i on which x and y differ, or the special rejection symbol $\perp$ if $x = y$.

We use Definition 2.20 rather than the definition of the introduction because it is more convenient to work with. For example, using Definition 2.20, it is trivial to prove a lower bound on the communication complexity of this relation: The easiest way to see it is to note that the task of checking whether two strings are equal reduces to R_{U_n}, and the communication complexity of this task is well known to be at least n.

We note, however, that the difference between Definition 2.20 and the definition of the introduction does not change the communication complexity of R_{U_n} substantially. To see it, suppose that there is a protocol Π that solves R_{U_n} under the promise that $x \neq y$. Then, there is a protocol

Π' that solves R_{U_n} without this promise using two more bits: Given inputs x and y which may be equal, the players invoke the protocol Π. Suppose Π outputs a coordinate i. Now, the players check whether $x_i \neq y_i$ by exchanging two more bits. If they find that $x_i = y_i$, they reject, and otherwise they output i.

We turn to define the composition of universal relations. The composition relation corresponds to the following communication problem:

- Alice gets a matrix $X \in \{0,1\}^{m\times n}$ and a string $a \in \{0,1\}^m$.
- Bob gets a matrix $Y \in \{0,1\}^{m\times n}$ and a string $b \in \{0,1\}^m$.
- They should find an entry (j,i) such that $X_{j,i} \neq Y_{j,i}$, or may reject if $a = b$ or if there exists an index $j \in [m]$ such that $a_j \neq b_j$ but $X_j = Y_j$.

Note that here, too, we do not make promises on the inputs, but rather allow the players to reject if the promises do not hold, and this has no substantial effect on the complexity.

Definition 2.21. The universal composition relation $R_{\mathrm{U}_m\circ\mathrm{U}_n}$ is defined as follows.

$$\begin{aligned} R_{\mathrm{U}_m\circ\mathrm{U}_n} \overset{\text{def}}{=} \; & \left\{((X,a),(Y,b),(j,i)) : X,Y \in \{0,1\}^{m\times n}, a,b \in \{0,1\}^n, X_{j,i} \neq Y_{j,i}\right\} \\ & \cup \left\{((X,a),(Y,b),\perp) : X,Y \in \{0,1\}^{m\times n}, a,b \in \{0,1\}^n, a = b\right\} \\ & \cup \left\{((X,a),(Y,b),\perp) : X,Y \in \{0,1\}^{m\times n}, a,b \in \{0,1\}^n, \exists j : \; a_j \neq b_j, X_j = Y_j\right\}. \end{aligned}$$

The following bound was proved in [HW93], improving on an earlier bound of [EIRS01]:

Theorem 2.22 ([HW93])**.** *It holds that that* $\mathsf{L}(R_{U_n\circ U_n}) \geq (1-o(1))\cdot 2^{2n-1}$*, and that* $\mathsf{C}(R_{U_n\circ U_n}) \geq 2n-1$ *for sufficiently large* n*.*

2.5 Information theory

We use basic concepts from information theory. For a more thorough overview to the field, including proofs to the claims presented below, we refer the reader to [CT91].

Definition 2.23 (Entropy)**.** The entropy of a random variable $\mathbf{x}$ is

$$H(\mathbf{x}) \overset{\text{def}}{=} \mathbb{E}_{x\sim\mathbf{x}}\left[\log\frac{1}{\Pr\left[\mathbf{x}=x\right]}\right] = \sum_x \Pr\left[\mathbf{x}=x\right]\cdot\log\frac{1}{\Pr\left[\mathbf{x}=x\right]}.$$

The *conditional entropy* $H(\mathbf{x}|\mathbf{y})$ is defined to be $\mathbb{E}_{\mathbf{y}}[H(\mathbf{x}|\mathbf{y}=y)]$.

Fact 2.24. $H(\mathbf{x})$ *is upper bounded by the logarithm of the support of* $\mathbf{x}$*, and equality is achieved by the uniform distribution over this support.*

We turn to define the notion of mutual information between two variables $\mathbf{x}$ and $\mathbf{y}$, which measures how much information $\mathbf{x}$ gives on $\mathbf{y}$ and vice versa. Intuitively, the information that $\mathbf{x}$ gives on $\mathbf{y}$ is captured by how much the uncertainty about $\mathbf{y}$ decreases when $\mathbf{x}$ becomes known.

Definition 2.25 (Mutual Information)**.** The mutual information between two random variables $\mathbf{x},\mathbf{y}$, denoted $I(\mathbf{x}:\mathbf{y})$ is defined as

$$I(\mathbf{x}:\mathbf{y}) \overset{\text{def}}{=} H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}). \tag{6}$$

The second equality in Equation 6 follows from the chain rule, to be discussed next. Similarly, for a random variable $\mathbf{z}$, the conditional mutual information $I(\mathbf{x};\mathbf{y}|\mathbf{z})$ is defined as

$$I(\mathbf{x}:\mathbf{y}|\mathbf{z}) \overset{\text{def}}{=} H(\mathbf{x}|\mathbf{z}) - H(\mathbf{x}|\mathbf{y},\mathbf{z}) = H(\mathbf{y}|\mathbf{z}) - H(\mathbf{y}|\mathbf{x},\mathbf{z}).$$

Fact 2.26 (The Chain Rule)**.** *Let* $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}$ *be random variables. Then*

$$\begin{aligned} H(\mathbf{x}, \mathbf{y}) &= H(\mathbf{x}) + H(\mathbf{y}|\mathbf{x}) \\ H(\mathbf{x}, \mathbf{y}|\mathbf{w}) &= H(\mathbf{x}|\mathbf{w}) + H(\mathbf{y}|\mathbf{x}, \mathbf{w}) \\ I(\mathbf{x}, \mathbf{y} : \mathbf{z}) &= I(\mathbf{x} : \mathbf{z}) + I(\mathbf{y} : \mathbf{z}|\mathbf{x}) \\ I(\mathbf{x}, \mathbf{y} : \mathbf{z}|\mathbf{w}) &= I(\mathbf{x} : \mathbf{z}|\mathbf{w}) + I(\mathbf{y} : \mathbf{z}|\mathbf{x}, \mathbf{w}). \end{aligned}$$

We use the following fact, which shows that if conditioning a uniformly distributed random variable on an event E decreases the entropy of $\mathbf{x}$ by much, then the event E must have small probability.

Fact 2.27. *Let* $\mathbf{x}$ *be a random variable that is uniformly distributed over a set* $\mathcal{X}$*, and let* E *be an event. If* $H(\mathbf{x}|E) \leq \log|\mathcal{X}| - t$*, then* $\Pr[E] \leq 2^{-t}$*.*

Proof. It holds that

$$\begin{aligned} \log|\mathcal{X}| - t &\geq H(\mathbf{x}|E) \\ &= \mathbb{E}_{x\sim\mathbf{x}|E}\left[\log\frac{1}{\Pr[\mathbf{x} = x|E]}\right] \\ \text{(Bayes' rule)} &\geq \mathbb{E}_{x\sim\mathbf{x}|E}\left[\log\frac{\Pr[E]}{\Pr[\mathbf{x} = x]}\right] \\ (\mathbf{x} \text{ is uniformly distributed over } \mathcal{X}) &= \mathbb{E}_{x\sim\mathbf{x}|E}\left[\log\frac{\Pr[E]}{1/|\mathcal{X}|}\right] \\ &= \log|\mathcal{X}| + \log\Pr[E]. \end{aligned}$$

It follows that $\log\Pr[E] \leq -t$ or that $\Pr[E] \leq 2^{-t}$. ■

2.6 Information complexity

In this paper we use the concept of Information Complexity, first defined in [CSWY01, BBCR10]. The main notion we use is the (external) information cost of a protocol, which captures what an external observer learns about both players' inputs from seeing the transcript of the protocol.

Definition 2.28 (External Information Cost)**.** Let μ be a distribution over pairs of inputs (x, y). The (external) information cost of a protocol Π over μ is given by

$$\mathsf{IC}_\mu(\Pi) \overset{\text{def}}{=} I_\mu(\Pi : \mathbf{x}, \mathbf{y}),$$

where on the right hand side, Π denotes the transcript $\Pi(\mathbf{x}, \mathbf{y})$.

We note that in this paper, we only consider deterministic protocols. In this special case, the external information cost is equal to the entropy $H_\mu(\Pi)$. The reason is that the transcript is a deterministic function of the inputs, which implies that $H(\Pi|\mathbf{x}, \mathbf{y}) = 0$ and therefore

$$\mathsf{IC}_\mu(\Pi) \overset{\text{def}}{=} I(\Pi : \mathbf{x}, \mathbf{y}) = H(\Pi) - H(\Pi|\mathbf{x}, \mathbf{y}) = H(\Pi).$$

However, it is often more useful to think of the information cost as $I(\Pi : \mathbf{x}, \mathbf{y})$ rather than as $H(\Pi)$.

Another common measure of information complexity is the internal information cost of a protocol, which captures the information that the players learn about each other's inputs from the protocol, given their prior knowledge:

Definition 2.29 (Internal Information Cost)**.** The internal information cost of a protocol Π over μ is given by:

$$\mathsf{IC}_{\mu}^{\text{int}}(\Pi) := I_{\mu}(\Pi : \mathbf{x}|\mathbf{y}) + I_{\mu}(\Pi : \mathbf{y}|\mathbf{x}).$$

Remark 2.30. Our notation is slightly different than the notation of previous works on information complexity. Previous works denoted by $\mathsf{IC}_{\mu}(\Pi)$ the *internal* information cost, and by $\mathsf{IC}_{\mu}^{\text{ext}}(\Pi)$ the external information cost. The reason for the difference is that previous works used mainly the internal cost, while we use mainly the external cost.

The following lemma describes the relationship between the internal information cost of a protocol Π, its external information cost, and its communication complexity.

Lemma 2.31 ([BR11])**.** *For any protocol* Π *and distribution* μ*, it holds that*

$$\mathsf{IC}_{\mu}^{\text{int}}(\Pi) \leq \mathsf{IC}_{\mu}(\Pi) \leq \mathsf{C}(\Pi).$$

Proof sketch. For the second inequality, note that since one bit of communication can never reveal more than one bit of information, both the external and internal information cost are upper bounded by the communication complexity of a protocol. Here is a sketch of the first inequality.

Let $(\mathbf{x}, \mathbf{y}) \sim \mu$, and let $\Pi = \Pi(\mathbf{x}, \mathbf{y})$. Suppose Alice speaks first, and denote the (random) bit she sends by Π_1. We show that $\mathsf{IC}_{\mu}^{\text{int}}(\Pi_1) \leq \mathsf{IC}_{\mu}(\Pi_1)$. Using the chain rule (Fact 2.26), the external information of Π_1 can be written as

$$I(\Pi_1; \mathbf{x}, \mathbf{y}) = I(\Pi_1; \mathbf{y}) + I(\Pi_1; \mathbf{x}|\mathbf{y}) \geq I(\Pi_1; \mathbf{x}|\mathbf{y}) = I(\Pi_1; \mathbf{x}|\mathbf{y}) + I(\Pi_1; \mathbf{y}|\mathbf{x}),$$

where the last equality follows since $I(\Pi_1; \mathbf{y}|\mathbf{x}) = 0$, as Alice's message Π_1 is independent of $\mathbf{y}$ given her input $\mathbf{x}$. Proceeding by induction on the number of bits of the protocol using the chain rule finishes the proof. ■

2.7 A combinatorial lemma

In this section, we state and prove a combinatorial lemma that will be used in Sections 4 and 6. The motivation for this lemma comes from the following question in communication complexity, which will be encountered in the latter sections: suppose Alice and Bob get as inputs $x, y \in \Sigma^m$ for some finite alphabet Σ. They would like to verify that their inputs agree on at least h coordinates. We wish to prove that Alice and Bob must transmit at least $h \cdot \log |\Sigma|$ bits.

This communication problem motivates the definition of the following property of sets of strings.

Definition 2.32. Let Σ be a finite alphabet, let $h, m \in \mathbb{N}$, and let $\mathcal{S} \subseteq \Sigma^m$. We say that $\mathcal{S}$ satisfies the h-agreement property if every two strings in $\mathcal{S}$ agree on at least h coordinates.

Now, in order to prove the lower bound on the above communication problem, we need an upper bound on the size of sets that satisfy the h-agreement property.

The most straightforward way to construct a set that satisfies the h-agreement property is to fix a set of coordinates $\mathcal{I} \subseteq [m]$ of size h, and take all the strings whose restriction to $\mathcal{I}$ is some fixed string. A set $\mathcal{S}$ constructed this way will be of size $|\Sigma|^{m-h}$. The following theorem says that this is the optimal way of constructing such a set.

Theorem 2.33 ([FT99, Corollary 1])**.** *Let* $\mathcal{S} \subseteq \Sigma^m$ *be a set that satisfies the* h*-agreement property, and suppose that* $|\Sigma| \geq h + 1$*. Then* $|\mathcal{S}| \leq |\Sigma|^{m-h}$*.*

The proof of [FT99] is quite non-trivial. For completeness, we provide a simple proof of the following weaker lemma, which is still sufficient for our purposes. Our proof generalizes the proof of [ADFS04, Claim 4.1], who considered the case $h = 1$ (following [GL74, Theorem 1]).

Lemma 2.34. *Let $\mathbb{F}$ be a finite field, let $m \leq |\mathbb{F}|$, and let $\mathcal{S} \subseteq \mathbb{F}^m$ be a set that satisfies the h-agreement property. Then $|\mathcal{S}| \leq |\mathbb{F}|^{m-h}$.*

Proof. We start with some notation. Let $H \subseteq \mathbb{F}$ be an arbitrary set of size m, and let us identify strings in $\mathbb{F}^m$ with functions $f : H \to \mathbb{F}$. Furthermore, let C be the set of such functions that are univariate polynomials of degree at most $h - 1$. Observe that $|C| = |\mathbb{F}|^h$, so the number of distinct cosets $x + C$ is $|\mathbb{F}|^{m-h}$.

Now, for the sake of contradiction, let us assume that $|\mathcal{S}| > |\mathbb{F}|^{m-h}$. By the pigeonhole principle, there exist two distinct strings $x, y \in \mathcal{S}$ such that $x+C = y+C$. Equivalently, it holds that $x-y \in C$, that is, $x - y$ is a non-zero univariate polynomial of degree at most $h - 1$. But, such a polynomial has at most $h-1$ roots, and therefore x and y may agree on at most $h-1$ coordinates, contradicting the assumption that $x, y \in \mathcal{S}$. ■

Remark 2.35. We note that in the proof of of Lemma 2.34, we could have replaced C with any MDS code of message length h.

3 On the Information Complexity of KW Relations

In this section, we make general observations regarding how one can use information complexity to analyze KW relations, and to prove formula lower bounds.

3.1 Information complexity and formula size

To see how information complexity is related to proving formula lower bounds, we first recall the following corollary of the connection between KW relations and functions (Corollary 2.14): for every boolean function f, it holds that $\mathsf{L}(f) = \mathsf{L}(R_f)$. In other words, the formula complexity of f is equal to the number of leaves in the smallest protocol tree that solves R_f.

Now, we have the following easy observation, which shows that one can prove lower bounds on the size of protocols by proving lower bounds on their information complexity.

Claim 3.1. *Let Π be a protocol. Then, for every distribution μ on the inputs, it holds that $\log \mathsf{L}(\Pi) \geq \mathsf{IC}_\mu(\Pi)$. Moreover, there exists a distribution μ over the inputs of the protocol such that $\log \mathsf{L}(\Pi) = \mathsf{IC}_\mu(\Pi)$. We refer to the later distribution μ as a hardest distribution of Π.*

Proof. We start by recalling that for deterministic protocols

$$\mathsf{IC}_\mu(\Pi) = H_\mu(\Pi),$$

(see discussion after Definition 2.28). The first part follows immediately by combining the latter equality with the fact that the entropy of a random variable is always upper bounded by the logarithm of its support's size (see [CT91]).

For the second part, we construct a hardest distribution μ for Π as follows: To sample a pair of inputs $(\mathbf{x}, \mathbf{y})$ from μ, pick a leaf ℓ of Π uniformly at random, and pick an arbitrary pair of inputs from $\mathcal{X}_\ell \times \mathcal{Y}_\ell$. Now, observe that the random variable $\Pi(\mathbf{x}, \mathbf{y})$ is uniform over the leaves of Π. Since the entropy of a uniform distribution is equal to the logarithm of its support's size, we get that

$$\mathsf{IC}_\mu(\Pi) = H_\mu(\Pi) = \log \mathsf{L}(\Pi),$$

as required. ■

This leads to the following corollary, which relates formula complexity and information complexity.

Corollary 3.2. *Let f be a boolean function, and let $s \in \mathbb{N}$. Then, $\mathsf{L}(f) \geq s$ if and only if for every protocol Π that solves R_f there exists a distribution μ, such that $\mathsf{IC}_\mu(\Pi) \geq \log s$.*

3.1.1 Example: Formula complexity of parity

We now give an example that shows how Corollary 3.2 can be useful for proving formula lower bounds. In particular, we use Corollary 3.2 to give a simple proof of the lower bound of [Khr72] for the parity function over n bits.

Theorem 3.3 ([Khr72]). *The parity function on n bits requires formulas of size n^2.*

Proof. Let $R_{\oplus_n}$ be the KW relation of the parity function on n bits. In this communication problem, Alice gets an n bit string x of even Hamming weight, Bob gets an n bit string of odd Hamming weight, and they would like to find a coordinate i such that $x_i \neq y_i$. Let Π be a protocol that solves $R_{\oplus_n}$. We show that there exists a distribution μ such that $\mathsf{IC}_\mu(\Pi) \geq 2\log n$, and this will imply the required lower bound.

We choose the distribution μ as follows. Let $\mathbf{x}$ and $\mathbf{y}$ be uniformly distributed string in $\{0,1\}^n$ of even and odd weights, respectively, such that $\mathbf{x}$ and $\mathbf{y}$ differ on a unique coordinate $\mathbf{i} \in [n]$, which is uniformly distributed over $[n]$. In other words, $(\mathbf{x}, \mathbf{y})$ is a uniformly distributed edge of the n-dimensional boolean hypercube.

Now, in order to lower-bound $\mathsf{IC}_\mu(\Pi)$, we use the fact that the external information cost is lower-bounded by the internal information cost (Lemma 2.31):

$$\begin{aligned} \mathsf{IC}_\mu(\Pi) &\geq \mathsf{IC}_\mu^{\text{int}}(\Pi) \\ &\overset{\text{def}}{=} I(\Pi : \mathbf{x}|\mathbf{y}) + I(\Pi : \mathbf{y}|\mathbf{x}). \end{aligned}$$

We claim that each of the terms $I(\Pi : \mathbf{x}|\mathbf{y})$ and $I(\Pi : \mathbf{y}|\mathbf{x})$ is lower-bounded by $\log n$, and this will prove the required lower bound. By symmetry, it suffices to analyze the first term $I(\Pi : \mathbf{x}|\mathbf{y})$. This term is the information that Bob gains from the interaction on the input of Alice. We show that this quantity is $\log n$ bits: intuitively, the reason is that Bob learns the coordinate $\mathbf{i}$ from the interaction. Formally, note that given $\mathbf{y}$, the string $\mathbf{x}$ determines $\mathbf{i}$, and vice versa. This implies that

$$\begin{aligned} I(\Pi : \mathbf{x}|\mathbf{y}) &= H(\mathbf{x}|\mathbf{y}) - H(\mathbf{x}|\mathbf{y}, \Pi) \\ \text{(Since } \mathbf{x} \text{ and } \mathbf{i} \text{ determine each other)} &= H(\mathbf{i}|\mathbf{y}) - H(\mathbf{i}|\mathbf{y}, \Pi) \\ \text{(Since the transcript } \Pi \text{ determines } \mathbf{i}\text{)} &= H(\mathbf{i}|\mathbf{y}) - 0 \\ \text{(Since } \mathbf{i} \text{ is uniform over } [n] \text{, even conditioned on } \mathbf{y}\text{)} &= \log n, \end{aligned}$$

as required. ■

We note that the above proof is just a reformulation of the previous proofs of [Khr72] and [KW90], but putting those proofs in terms of information complexity makes the proof particularly short, simple, and natural. We also note that the lower bound of Theorem 3.3 is tight, since there is a simple protocol that solves $R_{\oplus_n}$ by binary search.

3.1.2 Example: Protocol size of the universal relation

As an another example, we show how a lower bound on the size of the universal relation can be proved using information complexity. This serves as a good warm-up toward the proof of our main result in Section 4.

Claim 3.4. $\mathsf{L}(R_{\mathrm{U}_n}) \geq 2^n$.

Proof. The following argument is based on a similar argument of [HW93]. We construct a distribution μ such that every protocol Π for R_{U_n} satisfies $\mathsf{IC}_\mu(\Pi) \geq n$. We choose μ to be the distribution that outputs a pair $(\mathbf{x}, \mathbf{x})$, where $\mathbf{x} \in \{0,1\}^n$ is a uniformly distributed string. It holds that

$$\begin{aligned} \mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{x}) \\ &= H(\mathbf{x}) - H(\mathbf{x}|\Pi) \\ &= n - H(\mathbf{x}|\Pi). \end{aligned}$$

It remains to prove that $H(\mathbf{x}|\Pi) = 0$. We prove that for every fixed transcript π in the support of $\Pi(\mathbf{x}, \mathbf{x})$ it holds that $H(\mathbf{x}|\Pi = \pi) = 0$.

Fix a transcript π in the support of $\Pi(\mathbf{x}, \mathbf{x})$, and observe that π is a transcript in which the protocol outputs the rejection symbol $\perp$. Let $\mathcal{X}_\pi \times \mathcal{Y}_\pi$ be the rectangle of π, when viewed as a leaf of the protocol tree. Now, suppose that $H(\mathbf{x}|\Pi = \pi) > 0$. This implies that there exist at least two distinct strings $x, x' \in \{0,1\}^n$ that are in the support of $\mathbf{x}|\Pi = \pi$. Hence, it holds that $(x, x), (x', x') \in \mathcal{X}_\pi \times \mathcal{Y}_\pi$.

Now, since $\mathcal{X}_\pi \times \mathcal{Y}_\pi$ is a rectangle, we get that $(x, x') \in \mathcal{X}_\pi \times \mathcal{Y}_\pi$. This means that if we give x and x' as inputs to Alice and Bob, respectively, the resulting transcript will be π. However, in the latter case Alice and Bob get distinct strings, and thus are not allowed to output $\perp$. We reached a contradiction, and therefore $H(\mathbf{x}|\Pi = \pi) = 0$. ■

3.2 On hard distributions

There is an interesting difference between the way Corollary 3.2 is stated and the way we used it in the proofs of Theorem 3.3 and Claim 3.4: In Corollary 3.2, the distribution μ is *protocol dependent*, that is, the choice of μ depends on the protocol Π. On the other hand, in the proofs of Theorem 3.3 and Claim 3.4, the distribution μ is *protocol independent*. Moreover, all the previous works on interactive information complexity use protocol-independent distributions (e.g. [BJKS04, Bra12]).

This raises the question whether we can usually work with protocol independent distributions in the context of KW relations as well. Unfortunately, the answer is negative: the following proposition shows that the best lower bound one can prove for a KW relation using a protocol-independent distribution is at most $2\log n + O(1)$. This means that one cannot use protocol-independent distributions to prove formula lower bounds that are better than the lower bound for parity (up to a constant factor).

Proposition 3.5. *There exists a universal constant c such that the following holds: For every KW relation R_f and every distribution μ on inputs for R_f, there exists a protocol Π_μ such that*

$$\mathsf{IC}_\mu(\Pi) \leq 2\log n + O(1).$$

In order to prove Proposition 3.5, we first introduce some background.

Definition 3.6. Let Π be a protocol, and let μ be a distribution over inputs for Π. We denote by

$$\overline{\mathsf{C}}_{\mu}(\Pi) \overset{\text{def}}{=} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mu}\left[|\Pi(\mathbf{x},\mathbf{y})|\right]$$

the expected length of the transcript of Π. For a relation R, we denote by $\overline{\mathsf{C}}_{\mu}(R)$ the minimal value of $\overline{\mathsf{C}}_{\mu}(\Pi)$ over all protocols for R.

Definition 3.7. A (zero error) public coin protocol Π is a protocol in which the players share a random string, which they can use in order to solve the relation. Formally, we define a public coin protocol as a distribution over deterministic protocols. We define by $\mathsf{RC}(\Pi)$ the expected length of the transcript of Π for the *worst* pair of inputs, where the expectation is over the choice of the deterministic protocol. For a relation R, we denote by $\mathsf{RC}(R)$ the minimal value of $\mathsf{RC}(\Pi)$ over all protocols for R.

Fact 3.8 (Corollary of Yao's min-max theorem)**.** *For every relation R and distribution μ, it holds that $\overline{\mathsf{C}}_{\mu}(R) \leq \mathsf{RC}(R)$.*

We now go back to proving Proposition 3.5.

Proof of Proposition 3.5 Let $f : \{0,1\}^n \to \{0,1\}$ and let R_f its KW relation. We begin by observing that for every μ and Π it holds that

$$\mathsf{IC}_{\mu}(\Pi) \leq \overline{\mathsf{C}}_{\mu}(\Pi).$$

The reason is that the transcript can be thought of as a prefix-free code, and it is well known that the expected length of a prefix-free code is lower-bounded by its entropy (see, e.g., [CT91, Theorem 5.3.1]). It therefore suffices to prove that for every distribution μ, there is a protocol Π such that $\overline{\mathsf{C}}_{\mu}(\Pi) \leq 2\log n + O(1)$. By Fact 3.8, we can prove the latter claim by proving that $\mathsf{RC}(R_f) \leq 2\log n + O(1)$. A non-tight version of this upper bound was proved by Karchmer (see [RW89, Theorem 1.4]).

For completeness, we sketch a tight variant of Karchmer's protocol, which borrows ideas from [Bra12]: The players use the shared random string to choose an $n\times n$ invertible matrix A. Clearly, $A\cdot x \neq A\cdot y$. Next, the players find the first coordinate $j \in [n]$ such that $(A\cdot x)_j \neq (A\cdot y)_j$ as follows: Alice sends $(A\cdot x)_1$ and Bob replies with $(A\cdot y)_1$. If $(A\cdot x)_1 \neq (A\cdot y)_1$, then they found j. Otherwise, they proceed with $(A\cdot x)_2$ and $(A\cdot y)_2$, etc.

Now, the players know $j \in [n]$ such that $(A\cdot x)_j \neq (A\cdot y)_j$, and this means that x and y differ on the parity of the coordinates in the support of the j-th row of A. They therefore solve the KW relation of parity on those coordinates.

We turn to analyze the expected communication complexity of this protocol. We first observe that finding j requires $O(1)$ bits of communication in expectation. Basically, the reason is that conditioned on Alice and Bob reaching $(A\cdot x)_k$ and $(A\cdot y)_k$, the probability that $(A\cdot x)_k \neq (A\cdot y)_k$ is at least $\frac{1}{2}$. Hence, the probability that j will not be found after k iterations is at least $\left(\frac{1}{2}\right)^k$. Since in each such iteration they communicate 2 bits, the total expected communication complexity is at most $2\cdot\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k = 4$.

After finding j, the players solve the KW relation of parity over $\leq n$ bits. This can be done by transmitting at most $2\log n$ bits using binary search. Summing up, we get that the expected communication complexity of the protocol is at most $2\log n + 4$ bits. ■

Proposition 3.5 shows that if we want to prove lower bounds that are better than $2\log n + O(1)$, we have to tailor a different distribution for each protocol. One example for how such tailoring can be done is the construction of a hardest distribution μ_Π in Corollary 3.2. However, this construction is in some sense "trivial" and not interesting. We now provide a more interesting example: constructing hard distributions for the composition $\vee_m \circ f$. A considerably more sophisticated example is considered in Section 5, where we suggest a candidate construction of hard distributions for the composition $\oplus_m \circ f$.

Remark 3.9. There is an interesting contrast between Proposition 3.5, and the fact that the universal relation R_{U_n} does have a protocol-independent hard distribution, as we saw in the proof of Claim 3.4. The reason that R_{U_n} does have a protocol-independent hard distribution is that Karchmer's argument fails when the players are allowed to get identical inputs: specifically, in this case, $A \cdot x = A \cdot y$ and therefore j is never found. If we modify the definition of R_{U_n} such that the players are guaranteed to get distinct inputs, then the relation ceases to have a protocol-independent hard-distribution, and one has to construct protocol-dependent distributions instead.

3.2.1 Example: Information complexity of $\vee_m \circ f$

Let $\vee_m$ denote the disjunction function over m bits, and let $f : \{0,1\}^n \to \{0,1\}$. The composition $\vee_m \circ f$ is defined by invoking f on m distinct instances, and computing the disjunction of the results. The following result is well known (see, e.g., [Weg87, Chapter 10])

Theorem 3.10. $\mathsf{L}(\vee_m \circ f) = m \cdot \mathsf{L}(f) = \mathsf{L}(\vee_m) \cdot \mathsf{L}(f)$.

Proof. Let $\overline{0} \in \{0,1\}^n$ be the all-zeroes string, and assume, without loss of generality, that $f(\overline{0}) = 0$. Let ϕ be a formula that computes $\vee_m \circ f$. Now, observe that if one fixes all the m instances of f to $\overline{0}$ except for the first instance, then ϕ becomes a formula that computes f on the first instance, and therefore has at least $\mathsf{L}(f)$ leaves that are associated with the first instance of f. By repeating this argument for each of the m instances, and noting that leaves of ϕ that correspond to different instances must be distinct, it follows that ϕ must have at least $m \cdot \mathsf{L}(f)$ leaves. ■

We now recast the proof of Theorem 3.10 in terms of information complexity, as an example for how one can construct protocol-dependent hard distributions. It also serves as a simple example for how the chain rule could be used to analyze composition relations $R_{g\circ f}$, which is a recurring theme in this work.

Let $R_{\vee_m \circ f}$ be the KW relation of $\vee_m \circ f$, and fix a protocol Π for $R_{\vee_m \circ f}$. We show that there exists a distribution μ such that $\mathsf{IC}_\mu(\Pi) \geq \log m + \log \mathsf{L}(f)$. We denote inputs to $R_{\vee_m \circ f}$ by $m \times n$ matrices X, Y whose rows are inputs to f. The distribution μ samples a pair of inputs $(\mathbf{X}, \mathbf{Y})$ by the following process:

1. Choose a uniformly distributed $\mathbf{j} \in [m]$.
2. Let $\Pi_{\mathbf{j}}$ be the protocol for R_f obtained from Π by fixing $X_k = Y_k = \overline{0}$ for all $k \in [m] - \{\mathbf{j}\}$.
3. Sample a pair $(\mathbf{x}_j, \mathbf{y}_j)$ of inputs for R_f from a hardest distribution for $\Pi_{\mathbf{j}}$, which exists by Corollary 3.2.
4. Output the pair $(\mathbf{X}, \mathbf{Y})$ of inputs for $R_{\vee_m \circ f}$, where $\mathbf{X}_{\mathbf{j}} = \mathbf{x}_{\mathbf{j}}$, $\mathbf{Y}_{\mathbf{j}} = \mathbf{y}_{\mathbf{j}}$, and $\mathbf{X}_k = \mathbf{Y}_k = \overline{0}$ for all $k \in [m] - \{\mathbf{j}\}$.

It remains to prove that $\mathsf{IC}_\mu(\Pi) \geq \log m + \log \mathsf{L}(f)$. Intuitively, the bound follows because:

- The protocol reveals the index $\mathbf{j}$, which is $\log m$ bits of information.

- Conditioned on $\mathbf{j}$, the players still have to solve R_f on a hardest distribution for $\Pi_\mathbf{j}$, which reveals $\log \mathsf{L}(f)$ bits of information.

We turn to the formal proof, starting with the following observations:

- The index $\mathbf{j}$ is determined by the pair $(\mathbf{X}, \mathbf{Y})$, since this is the only row on which the matrices differ.
- The index $\mathbf{j}$ is determined by the transcript $\Pi(\mathbf{X}, \mathbf{Y})$, since the protocol Π finds a entry on which $\mathbf{X}$ and $\mathbf{Y}$ differ, and this entry must belong to the j-th row.

Now, it holds that

$$\begin{aligned}
\mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{X}, \mathbf{Y}) \\
\text{(Since } \mathbf{j} \text{ is determined by } \mathbf{X}, \mathbf{Y}) &= I(\Pi : \mathbf{X}, \mathbf{Y}, \mathbf{j}) \\
\text{(The chain rule)} &= I(\Pi : \mathbf{j}) + I(\Pi : \mathbf{X}, \mathbf{Y} | \mathbf{j}) \\
\text{(Conditioned on } \mathbf{j}\text{, the protocol } \Pi \text{ behaves like } \Pi_\mathbf{j}) &= I(\Pi : \mathbf{j}) + I(\Pi_\mathbf{j} : \mathbf{X}_\mathbf{j}, \mathbf{Y}_\mathbf{j}) \; (7) \\
\text{(Since } \Pi \text{ determines } \mathbf{j}) &= \log m + I(\Pi_\mathbf{j} : \mathbf{X}_\mathbf{j}, \mathbf{Y}_\mathbf{j}) \\
\text{(Since } \mathbf{X}_\mathbf{j}, \mathbf{Y}_\mathbf{j} \text{ were drawn from a hardest distribution for } \Pi_\mathbf{j}) &= \log m + \log \mathsf{L}(f).
\end{aligned}$$

This concludes the argument. We note a particularly interesting feature of this proof: Equality 7 decomposes the information that protocol transmits about $R_{\vee_m \circ f}$ into information about $R_{\vee_m}$ and information about R_f. This is exactly the kind of decomposition we would like to have for every composition relation $R_{g \circ f}$. We also note the role that the chain rule plays in deriving this decomposition.

3.3 External versus internal information cost

As discussed in Section 2.6, the literature on information complexity has two notions of information cost, namely, an external cost and an internal cost. So far we used mostly the external cost IC, and used the internal cost IC^{int} only to derive the lower bound on the parity function in Section 3.1.1. On the other hand, most previous works on information complexity used mainly the internal cost. This raises the question how useful is the notion of internal information cost to the study of KW relations.

The next proposition shows that internal information cost cannot be used to prove lower bounds beyond $2 \log n$, which makes the example of parity optimal. Recall that the internal information cost is defined by

$$\mathsf{IC}^{\text{int}}_\mu(\Pi) \overset{\text{def}}{=} I(\Pi : \mathbf{x} | \mathbf{y}) + I(\Pi : \mathbf{y} | \mathbf{x}).$$

Proposition 3.11. *There exists a protocol Π that solves every KW relation R_f, such that*

$$\mathsf{IC}^{\text{int}}_\mu(\Pi) \leq 2 \log n$$

for every distribution μ.

Proof. The following argument is somewhat similar to [Bra12, Prop. 3.21], but is simpler. On inputs x and y, the protocol Π works iteratively as follows: in the i-th iteration, Alice and Bob send x_i and y_i. If $x_i \neq y_i$, Alice and Bob halt and output i. Otherwise, they proceed to the next iteration.

The correctness of Π is easy to see. It remains to show that for every distribution μ, it holds that $\mathsf{IC}_{\mu}^{\text{int}}(\Pi) \le 2\log n$. We prove that $I(\Pi : \mathbf{x}|\mathbf{y}) \le \log n$. A similar argument holds for $I(\Pi : \mathbf{y}|\mathbf{x})$, and by taking their sum it follows that $\mathsf{IC}_{\mu}^{\text{int}}(\Pi) \le 2\log n$. Intuitively, it holds that $I(\Pi : \mathbf{x}|\mathbf{y}) \le \log n$ because the only thing that Bob learns is the first coordinate i on which $\mathbf{x}$ and $\mathbf{y}$ differ.

Formally, let μ be a distribution over pairs $(\mathbf{x}, \mathbf{y})$ of inputs. Let $\mathbf{i}$ be the first coordinate on which $\mathbf{x}$ and $\mathbf{y}$ differ, and note that the transcript $\Pi(\mathbf{x}, \mathbf{y})$ determines $\mathbf{i}$. It holds that

$$\begin{aligned} I(\Pi : \mathbf{x}|\mathbf{y}) &\le H(\Pi|\mathbf{y}) \\ \text{(Since } \Pi \text{ determines } \mathbf{i}) &= H(\Pi, \mathbf{i}|\mathbf{y}) \\ \text{(The chain rule)} &= H(\mathbf{i}|\mathbf{y}) + H(\Pi|\mathbf{i}, \mathbf{y}) \\ \text{(The entropy is upper bounded by the logarithm of the support size)} &\le \log n + H(\Pi|\mathbf{i}, \mathbf{y}). \end{aligned}$$

It remains to prove that $H(\Pi|\mathbf{i}, \mathbf{y}) = 0$. To see it, observe that $\mathbf{y}$ and $\mathbf{i}$ together determine the transcript $\Pi(\mathbf{x}, \mathbf{y})$. It follows that $I(\Pi : \mathbf{x}|\mathbf{y}) \le \log n$ and therefore $\mathsf{IC}_{\mu}^{\text{int}}(\Pi) \le 2\log n$, as required. ■

4 The Composition of a Function with the Universal Relation

In this section, we prove our main result, namely, a lower bound on the complexity of the relation $R_{g \circ \mathrm{U}_n}$. We start by defining $R_{g \circ \mathrm{U}_n}$ formally. Let $g : \{0,1\}^m \to \{0,1\}$. The relation $R_{g \circ \mathrm{U}_n}$ corresponds to the following communication problem: Alice gets as an input a matrix $X \in \{0,1\}^{m \times n}$ and a string $a \in g^{-1}(0)$. Bob gets a matrix $Y \in \{0,1\}^{m \times n}$ and a string $b \in g^{-1}(1)$. Their goal is to find an entry (j, i) on which X and Y differ, but they are allowed to reject if there exists an index $j \in [m]$ such that $a_j \ne b_j$ but $X_j = Y_j$. Formally,

Definition 4.1. Let $g : \{0,1\}^m \to \{0,1\}$, and let $n \in \mathbb{N}$. The relation $R_{g \circ \mathrm{U}_n}$ is defined by

$$\begin{aligned} R_{g \circ \mathrm{U}_n} \overset{\text{def}}{=} & \left\{((X,a),(Y,b),(j,i)) : X, Y \in \{0,1\}^{m \times n}, a \in g^{-1}(0), b \in g^{-1}(1), X_{j,i} \ne Y_{j,i}\right\} \\ & \cup \left\{((X,a),(Y,b),\perp) : X, Y \in \{0,1\}^{m \times n}, a \in g^{-1}(0), b \in g^{-1}(1), \exists j : a_j \ne b_j, X_j = Y_j\right\}. \end{aligned}$$

Theorem (1.6, main theorem, restated)**.** *Let $m, n \in \mathbb{N}$, and let $g : \{0,1\}^m \to \{0,1\}$ be a non-constant function. Then,*

$$\mathsf{C}(R_{g \circ U}) \ge \log \mathsf{L}(R_{g \circ \mathrm{U}_n}) \ge \log \mathsf{L}(g) + n - O(1 + \frac{m}{n}) \cdot \log m.$$

The rest of this section is organized as follows. First, in Section 4.1, we consider the special case $R_{\oplus_m \circ \mathrm{U}_n}$, and prove a lower bound that is tighter than the main theorem. Then, in Section 4.2, we prove the main theorem itself. The special case $R_{\oplus_m \circ \mathrm{U}_n}$ serves as a warm-up toward the proof of the main theorem, and as discussed in the introduction, it is also a step toward proving Conjecture 1.9. The following definition will be useful for both proofs.

Definition 4.2. Let ℓ be a leaf of Π and let $\mathcal{X}_\ell \times \mathcal{Y}_\ell$ be its corresponding rectangle.

- We say that the leaf ℓ supports a matrix $X \in \{0,1\}^{m \times n}$ if X can be given as an input to both players at ℓ. Formally, ℓ supports X if there exist $a, b \in \{0,1\}^m$ such that $(X, a) \in \mathcal{X}_\ell$ and $(X, b) \in \mathcal{Y}_\ell$. We also say that X is supported by ℓ and a, or by ℓ and b. Note that the leaf ℓ must be a leaf that outputs $\perp$.

- We say that the leaf ℓ supports $a \in g^{-1}(0)$ if a can be given as input to Alice at ℓ. Formally, ℓ supports a if there exists a matrix $X \in \{0,1\}^{m \times n}$ such that $(X, a) \in \mathcal{X}_\ell$. A similar definition applies to strings $b \in g^{-1}(1)$.

4.1 Complexity of $R_{\oplus_m \circ \mathrm{U}_n}$

The relation $R_{\oplus_m \circ \mathrm{U}_n}$ corresponds to the following communication problem: Alice gets a matrix $X \in \{0,1\}^{m \times n}$ and a string $a \in \{0,1\}^m$ of even weight. Bob gets a matrix $Y \in \{0,1\}^{m \times n}$ and a string $b \in \{0,1\}^m$ of odd weight. Their goal is to find an entry on which X and Y differ, and they are allowed to reject if there is an index $j \in [m]$ such that $a_j \neq b_j$ but $X_j = Y_j$. We prove the following result:

Theorem (1.10, restated)**.** *For every $m, n \in \mathbb{N}$ it holds that*

$$\mathsf{C}(R_{\oplus_m \circ U_n}) \geq \log \mathsf{L}(R_{\oplus_m \circ \mathrm{U}_n}) \geq 2 \log m + n - O(\log \log m).$$

We note that only the second inequality requires a proof, whereas the first inequality is trivial since a binary tree of depth c has at most 2^c leaves. Fix a protocol Π for $R_{\oplus_m \circ \mathrm{U}_n}$. We analyze the external information cost of Π with respect to the distribution μ that is sampled as follows:

1. Choose a uniformly distributed matrix $\mathbf{X} \in \{0,1\}^{m \times n}$.
2. Choose uniformly distributed strings $\mathbf{a}, \mathbf{b} \in \{0,1\}^m$ of even and odd weights, respectively, which differ on a unique coordinate $\mathbf{j}$ that is uniformly distributed over $[m]$. In other words, $(\mathbf{a}, \mathbf{b})$ is a uniformly distributed edge of the hypercube.
3. Give the input $(\mathbf{X}, \mathbf{a})$ to Alice, and $(\mathbf{X}, \mathbf{b})$ to Bob.

Note that the distribution μ can be thought of as a combination of the hard distribution for $\oplus_m$ (as we saw in Section 3.1.1) and of m independent copies of the hard distribution for the universal relation (as we saw in Section 3.1.2).

We now prove the lower bound of Theorem 1.10 on the information cost $\mathsf{IC}_\mu(\Pi)$. The intuition for this proof is the following: on inputs drawn from μ, Alice and Bob always reject and output $\perp$. In order for them to output $\perp$, they must be convinced that they agree on the $\mathbf{j}$-th row of their matrices, where $\mathbf{j}$ is the unique coordinate such that $\mathbf{a_j} \neq \mathbf{b_j}$. In particular:

1. Alice and Bob must be convinced that they agree on at least one row of their matrices. We show that this requires them to transmit at least n bits of information (see Lemma 4.3 below).
2. Alice and Bob either find the coordinate $\mathbf{j}$, or not. We consider the two cases separately:
   (a) If they find $\mathbf{j}$, then they must transmit about $2 \log m$ bits of information, since this is the information complexity of $\oplus_m$ on the distribution $(\mathbf{a}, \mathbf{b})$.
   (b) If they do not find $\mathbf{j}$, then at the end of the protocol there are multiple possibilities for the value of $\mathbf{j}$. In such a case, Alice and Bob must be convinced that they agree on all the corresponding rows in their matrices - otherwise, they are not allowed to output $\perp$. However, this requires them to transmit n bits for each possible value of $\mathbf{j}$, and for most matrices $\mathbf{X}$ they cannot afford it, unless $\log \mathsf{L}(\Pi) > 2 \log m + n$.

The formal proof goes as follows. With some abuse of notation, we denote by $\Pi = \Pi\left((\mathbf{X}, \mathbf{a}), (\mathbf{X}, \mathbf{b})\right)$ the random transcript of the protocol on μ. It holds that

$$\begin{aligned} \mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{X}, \mathbf{a}, \mathbf{b}) \\ \text{(The chain rule)} &= I(\Pi : \mathbf{X}) + I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X}). \end{aligned}$$

Thus, we decomposed the information cost of Π into information about $\mathbf{X}$ (which corresponds to R_{U_n}), and information about $\mathbf{a}, \mathbf{b}$ (which correspond to $R_{\oplus_m}$). We would now like to show that the first term contributes about n (the information complexity of the R_{U_n}), and that the second term contributes $2\log m$ (the information complexity of $R_{\oplus_m}$). The following two lemmas state the precise bounds.

Lemma 4.3. *$I(\Pi : \mathbf{X}) \geq n$.*

Lemma 4.4. *$I(\Pi : \mathbf{a}, \mathbf{b}|\mathbf{X}) \geq 2\log m - O(\log\log m)$.*

We prove lemmas 4.3 and 4.4 in the next two subsections.

4.1.1 Proof of Lemma 4.3

We prove that $I(\Pi : \mathbf{X}) \geq n$. We note that the following proof only uses the facts that Π is a protocol for a relation of the form $R_{g \circ \mathrm{U}_n}$, and that $\mathbf{X}$ was chosen uniformly at random. In particular, the proof does not use the fact that $g = \oplus_m$, or the precise form of the distribution $\mathbf{a}, \mathbf{b}|\mathbf{X}$. Thus, we will be able to use this lemma again in Section 4.2 below.

As discussed above, the intuition for the lower bound $I(\Pi : \mathbf{X}) \geq n$ is that by the end of the protocol, Alice and Bob must be convinced that their matrices agree on at least one row, and we will show that this requires transmitting n bits of information. By the definition of mutual information, it holds that

$$\begin{aligned} I(\Pi : \mathbf{X}) &= H(\mathbf{X}) - H(\mathbf{X}|\Pi) \\ &= m \cdot n - H(\mathbf{X}|\Pi). \end{aligned}$$

Thus, it suffices to prove that $H(\mathbf{X}|\Pi) \leq (m-1) \cdot n$. We prove the following stronger claim: for every fixed transcript π in the support of Π, the number of matrices that are supported by π is at most $2^{(m-1)\cdot n}$.

Fix a transcript π, and let $\mathcal{T}$ be the set of matrices X that are supported by π (see Definition 4.2). We prove the following claim on $\mathcal{T}$, which is equivalent to saying that Alice and Bob must be convinced that their matrices agree on at least one row.

Claim 4.5. *Every two matrices X, X' in $\mathcal{T}$ agree on at least one row.*

Proof. We use a standard "fooling set" argument. Let $\mathcal{X}_\pi \times \mathcal{Y}_\pi$ denote the rectangle that corresponds to π. Suppose, for the sake of contradiction, that there exist $X, X' \in \mathcal{T}$ that do not agree on any row. By definition of $\mathcal{T}$, it follows that there exist strings $a, b \in \{0,1\}^m$ of even and odd weights, respectively such that $(X, a) \in \mathcal{X}_\pi$ and $(X', b) \in \mathcal{Y}_\pi$. In particular, this means that if we give to Alice and Bob the inputs (X, a) and (X', b), respectively, the resulting transcript of the protocol will be π.

However, this is a contradiction: on the one hand, π is a transcript on which the protocol outputs $\perp$, since it was generated by the distribution μ. On the other hand, the players are not allowed to output $\perp$ on inputs (X, a), (X', b), since X and X' differ on all their rows, and in particular differ on the all the rows j for which $a_j \neq b_j$. The claim follows. ■

Finally, we observe that Claim 4.5 is just another way of saying that $\mathcal{T}$ satisfies the 1-agreement property (Definition 2.32), when viewed as a set of strings in $\mathbb{F}^m$ over the alphabet $\mathbb{F} = \{0,1\}^n$. Therefore, Lemma 2.34 implies that $|\mathcal{T}| \leq 2^{(m-1)\cdot n}$, as required.

4.1.2 Proof of Lemma 4.4

We turn to prove that $I(\Pi : \mathbf{a}, \mathbf{b}|\mathbf{X}) \geq 2\log m - O(\log\log m)$. The intuition for the proof is the following. Either the transcript $\Pi = \Pi\left((\mathbf{X}, \mathbf{a})(\mathbf{X}, \mathbf{b})\right)$ reveals the coordinate $\mathbf{j}$ on which Alice and Bob differ, or it does not. We show that in the first ("good") case, Π reveals almost $2\log m$ bits of information, and that the second ("bad") case rarely happens.

Formally, we say that a transcript π is bad if, for $t \overset{\text{def}}{=} \log\log m + 2$, it holds that either

$$H(\mathbf{j}|\mathbf{a}, \Pi = \pi) > t, \text{ or } H(\mathbf{j}|\mathbf{b}, \Pi = \pi) > t,$$

which intuitively means that π does not reveal $\mathbf{j}$ to one of the players. Otherwise, we say that π is good. Since external information is lower-bounded by internal information (Lemma 2.31), it holds that

$$\begin{aligned}
I(\Pi : \mathbf{a}, \mathbf{b}|\mathbf{X}) &\geq I(\Pi : \mathbf{a}|\mathbf{b}, \mathbf{X}) + I(\Pi : \mathbf{b}|\mathbf{a}, \mathbf{X}) \\
\text{(Since } \mathbf{j} \text{ and } \mathbf{a} \text{ determine each other conditioned on } \mathbf{b}) &= I(\Pi : \mathbf{j}|\mathbf{b}, \mathbf{X}) + I(\Pi : \mathbf{j}|\mathbf{a}, \mathbf{X}) \\
\text{(By the definition of mutual information)} &= H(\mathbf{j}|\mathbf{b}, \mathbf{X}) - H(\mathbf{j}|\mathbf{b}, \mathbf{X}, \Pi) \\
&\quad + H(\mathbf{j}|\mathbf{a}, \mathbf{X}) - H(\mathbf{j}|\mathbf{a}, \mathbf{X}, \Pi) \\
\text{(Since } \mathbf{j} \text{ is independent of } \mathbf{a}, \mathbf{X} \text{ or } \mathbf{b}, \mathbf{X}) &= 2 \cdot H(\mathbf{j}) - H(\mathbf{j}|\mathbf{b}, \mathbf{X}, \Pi) - H(\mathbf{j}|\mathbf{a}, \mathbf{X}, \Pi) \\
&= 2\log m - H(\mathbf{j}|\mathbf{b}, \mathbf{X}, \Pi) - H(\mathbf{j}|\mathbf{a}, \mathbf{X}, \Pi) \\
\text{(Since removing conditioning does not decrease entropy)} &\geq 2\log m - H(\mathbf{j}|\mathbf{b}, \Pi) - H(\mathbf{j}|\mathbf{a}, \Pi).
\end{aligned}$$

Now, if both $H(\mathbf{j}|\mathbf{a}, \Pi)$ and $H(\mathbf{j}|\mathbf{b}, \Pi)$ are at most t, then we are done. We prove that this is indeed the case, by proving that bad transcripts π occur with low probability, and therefore do not contribute too much to $H(\mathbf{j}|\mathbf{a}, \Pi)$ and $H(\mathbf{j}|\mathbf{b}, \Pi)$. In other words, the transcript Π is usually good, and almost reveals $\mathbf{j}$.

Fix a bad transcript π, and assume without loss of generality that $H(\mathbf{j}|\mathbf{a}, \Pi = \pi) > t$. We prove the following claim, which says that if $\mathbf{j}$ has not been revealed, a lot of information must have been revealed on $\mathbf{X}$.

Claim 4.6. *It holds that* $H(\mathbf{X}|\mathbf{a}, \Pi = \pi) \leq m \cdot n - 2^{H(\mathbf{j}|\mathbf{a}, \Pi = \pi)} \cdot n$.

Proof. We start by proving the claim for fixed values of $\mathbf{a}$. Fix a string a that is supported by π. Let $\mathcal{J}$ be the support of $\mathbf{j}|\mathbf{a} = a, \Pi = \pi$, and let $\mathcal{T}$ be the set of matrices that are supported by π and a. We prove that

$$H(\mathbf{X}|\mathbf{a} = a, \Pi = \pi) \leq (m - |\mathcal{J}|) \cdot n \leq m \cdot n - 2^{H(\mathbf{j}|\mathbf{a} = a, \Pi = \pi)} \cdot n. \tag{8}$$

To this end, we show that all the matrices in $\mathcal{T}$ must agree on all the rows whose indices are in $\mathcal{J}$. Let $j \in \mathcal{J}$. We show that all the matrices in $\mathcal{T}$ agree on the j-th row using a standard fooling set argument. Let $\mathcal{X}_\pi \times \mathcal{Y}_\pi$ be the rectangle associated with π. By the definition of $\mathcal{J}$, there exists a string b^j of even weight that differs from a only on the coordinate j, such that $(Y, b^j) \in \mathcal{Y}_\pi$ for some matrix Y. We claim that the matrices $X \in \mathcal{T}$ agree with Y on its j-th row.

To see it, let $X \in \mathcal{T}$, and observe that if we give the input (X, a) to Alice, and (Y, b^j) to Bob, the resulting transcript will be π. However, π is a transcript that outputs $\perp$, and since j is the only coordinate on which a and b^j differ, the protocol is only allowed to output $\perp$ if the players agree on the j-th row of their matrices, that is, $X_j = Y_j$.

It follows that all the matrices in $\mathcal{T}$ agree on all the rows in $\mathcal{J}$, and therefore $\mathcal{T} \leq 2^{(m-|\mathcal{J}|)\cdot n}$. Inequality 8 now follows by noting that $|\mathcal{J}| \geq 2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}$. To derive the claim, we average over $\mathbf{a}$ and use the convexity of the function 2^x:

$$\begin{aligned} H(\mathbf{X}|\mathbf{a},\Pi=\pi) &= \mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}\left[H(\mathbf{X}|\mathbf{a}=a,\Pi=\pi)\right] \\ &\leq \mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}\left[m\cdot n - 2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}\cdot n\right] \\ &= m\cdot n - \mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}\left[2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}\right]\cdot n \\ (2^x \text{ is convex}) &\leq m\cdot n - 2^{\mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}[H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)]}\cdot n \\ &= m\cdot n - 2^{H(\mathbf{j}|\mathbf{a},\Pi=\pi)}\cdot n, \end{aligned}$$

as required. ■

We now use Claim 4.6 to show that π only occurs with low probability. To this end, we show that conditioning on $\Pi=\pi$ decreases the entropy of the pair $(\mathbf{X},\mathbf{a})$ by much, and then use Fact 2.27 to deduce that the event $\Pi=\pi$ must have low probability. Observe that

$$\begin{aligned} H(\mathbf{X},\mathbf{a}|\Pi=\pi) &= H(\mathbf{a}|\Pi=a) + H(\mathbf{X}|\mathbf{a},\Pi=\pi) \\ &\leq m + m\cdot n - 2^{H(\mathbf{j}|\mathbf{a},\Pi=\pi)}\cdot n \\ (\text{By assumption on } \pi) &\leq m + m\cdot n - 2^t\cdot n \\ &< m + m\cdot n - 4\cdot n\cdot\log m. \end{aligned}$$

Now, when not conditioning on $\Pi=\pi$, the pair $(\mathbf{X},\mathbf{a})$ is uniformly distributed over the set of all pairs (X,a), which is of size $2^{m+m\cdot n}$. Hence, by Fact 2.27, it holds that

$$\Pr\left[\Pi=\pi\right] \leq 2^{-4\cdot n\cdot\log m} \leq m^{-4}\cdot 2^{-n}.$$

This shows that the probability of a fixed bad transcript π is at most $m^{-4}\cdot 2^{-n}$. We now apply union bound over all bad transcripts, and deduce that

$$\Pr\left[\Pi \text{ is bad}\right] \leq \mathsf{L}(\Pi)\cdot m^{-4}\cdot 2^{-n}.$$

We may assume that $\mathsf{L}(\Pi) \leq 2^{2\log m+n}$, since otherwise the theorem we are trying to prove would follow immediately. It follows that

$$\begin{aligned} \Pr\left[\Pi \text{ is bad}\right] &\leq m^2\cdot 2^n\cdot m^{-4}\cdot 2^{-n} \\ &\leq m^{-2}. \end{aligned}$$

We conclude that

$$\begin{aligned} I(\Pi:\mathbf{a},\mathbf{b}|\mathbf{X}) &\geq 2\log m - (H(\mathbf{j}|\mathbf{b},\Pi) + H(\mathbf{j}|\mathbf{a},\Pi)) \\ &\geq 2\log m \\ &\quad - \Pr\left[\Pi \text{ is not bad}\right]\cdot\mathbb{E}\left[H(\mathbf{j}|\mathbf{b},\Pi) + H(\mathbf{j}|\mathbf{a},\Pi)|\,\Pi \text{ is not bad}\right] \\ &\quad - \Pr\left[\Pi \text{ is bad}\right]\cdot\mathbb{E}\left[H(\mathbf{j}|\mathbf{b},\Pi) + H(\mathbf{j}|\mathbf{a},\Pi)|\,\Pi \text{ is bad}\right] \\ &\geq 2\log m - \Pr\left[\Pi \text{ is not bad}\right]\cdot(2\cdot t) - \Pr\left[\Pi \text{ is bad}\right]\cdot(2\log m) \\ &\geq 2\log m - 2\cdot t - m^{-2}\cdot(2\log m) \\ &= 2\log m - O(\log\log m), \end{aligned}$$

and Lemma 4.4 follows as required.

4.2 Complexity of $R_{g \circ \mathrm{U}_n}$

We turn to prove our lower bound for a general function g, namely,

Theorem (1.6, main theorem, restated)**.** *Let $m, n \in \mathbb{N}$, and let $g : \{0,1\}^m \to \{0,1\}$ be a non-constant function. Then,*

$$\mathsf{C}(R_{g \circ U}) \geq \log \mathsf{L}(R_{g \circ \mathrm{U}_n}) \geq \log \mathsf{L}(g) + n - O(1 + \frac{m}{n}) \cdot \log m.$$

Again, only the second inequality requires a proof, whereas the first inequality is trivial since a binary tree of depth c has at most 2^c leaves. Fix a protocol Π for $R_{g \circ \mathrm{U}_n}$. We define a distribution μ on inputs for Π, and prove a lower bound for $\mathsf{IC}_\mu(\Pi)$.

4.2.1 Proof outline

Generally, the proof will follow the lines of the proof for $R_{\oplus_m \circ \mathrm{U}_n}$: we construct the distribution μ with random variables $\mathbf{X} \in \{0,1\}^{m \times n}$, $\mathbf{a} \in g^{-1}(0)$, $\mathbf{b} \in g^{-1}(1)$, and give to Alice and Bob the inputs $(\mathbf{X}, \mathbf{a})$ and $(\mathbf{X}, \mathbf{b})$, respectively. We observe that for inputs drawn from μ, Alice and Bob always reject and output $\perp$. To do that, they have to be convinced that there exists some $j \in [m]$ such that $a_j \neq b_j$, and such that their matrices agree on their j-th row. In particular:

1. Alice and Bob must be convinced that they agree on at least one row of their matrices. As in the case of $R_{\oplus_m \circ \mathrm{U}_n}$, this requires them to transmit at least n bits of information.
2. Alice and Bob either find a coordinate j such that $\mathbf{a}_j \neq \mathbf{b}_j$, or not. We consider the two cases separately:

   (a) If they find such a coordinate j, then they must solve the KW relation R_g, and therefore they must transmit about $\log L(g)$ bits of information.

   (b) If they do not find such a coordinate j, then at the end of the protocol there are multiple possibilities for coordinates j on which they may differ, and Alice and Bob must be convinced that they agree on all the corresponding rows in their matrices - otherwise, they are not allowed to output $\perp$. However, this requires them to transmit n bits for each such row, and for most matrices $\mathbf{X}$ they cannot afford it, unless $\log \mathsf{L}(\Pi)$ is large.

However, there are two issues that we need to deal with in order to implement this approach:

- In Item 2, it is no longer clear what does it mean "to find j such that $\mathbf{a}_j \neq \mathbf{b}_j$". In the case of $R_{\oplus_m \circ \mathrm{U}_n}$, the coordinate $\mathbf{j}$ was unique, and therefore we could tell whether the players know it by looking at the entropies $H(\mathbf{j}|\mathbf{a}, \Pi)$ and $H(\mathbf{j}|\mathbf{b}, \Pi)$. However, for the general case of $R_{g \circ \mathrm{U}_n}$, there might be multiple coordinates j such that $\mathbf{a}_j \neq \mathbf{b}_j$, and thus there is no single random variable whose entropy can be measured. Therefore, instead about defining this case as "Alice and Bob find j", we define it as "Alice and Bob essentially solve the relaxed KW problem $R_g(t)$" (see Section 2.3.1 for the definition of $R_g(t)$).
- In Item 2a above, we would like to argue that if the players solve R_g, they must transmit at least $\log L(g)$ bits of information. However, this is only true if the players solve R_g on a hardest distribution of R_g.
  It is not clear how to define such a hardest distribution: In general a hard distribution is protocol dependent, and therefore one needs to construct it with respect to a specific protocol for R_g. However, here we do not have a protocol that solves R_g, but only a protocol that

solves $R_{g \circ \mathrm{U}_n}$.

To resolve this issue, we extract sub-trees of Π that can "play the role" of a protocol for R_g, and construct the hard distribution with respect to those sub-trees. More specifically, for every matrix $X \in \{0,1\}^{m \times n}$, we consider the sub-tree T_X of Π that consists of the leaves of Π that support X. As we show below, in the case where Alice and Bob find a coordinate j such that $a_j \neq b_j$, the sub-tree T_X can be treated as a protocol for the relaxed KW problem $R_g(t)$. We therefore sample $(\mathbf{a}, \mathbf{b})$ from a hardest distribution for $T_{\mathbf{X}}$, and the analysis can proceed as before.

There is also a more technical difference between the following proof and the proof for $R_{\oplus_m \circ \mathrm{U}_n}$: in the proof for $R_{\oplus_m \circ \mathrm{U}_n}$, we distinguished the Cases 2a and 2b above by distinguishing "good" and "bad" transcripts, which were transcripts that, respectively, reveal and do not reveal $\mathbf{j}$. We then showed that bad transcripts occurred with low probability. On the other hand, in the following proof, we distinguish the Cases 2a and 2b above by distinguishing "good" and "bad" *matrices*, which are matrices for which, intuitively, the sub-tree T_X solves R_g or does not solve R_g, respectively. We then show that bad matrices occur with low probability.

4.2.2 Construction of the distribution μ

We begin the formal proof by constructing the distribution μ, with respect to which we will analyze the information cost of Π. To this end, we first define the sub-tree T_X of a matrix X.

Definition 4.7. Let $X \in \{0,1\}^{m \times n}$ be a matrix. Then, the sub-tree of X, denoted T_X, is the sub-tree of Π that consists of the leaves that support X. Note that all those leaves output $\perp$.

The distribution μ is sampled as follows:

1. Choose a uniformly distributed matrix $\mathbf{X} \in \{0,1\}^{m \times n}$.
2. Choose a uniformly distributed leaf ℓ of $T_{\mathbf{X}}$, and let $\mathcal{X}_\ell \times \mathcal{Y}_\ell$ denote its rectangle.
3. Choose an arbitrary pair $(\mathbf{a}, \mathbf{b})$ such that $(\mathbf{X}, \mathbf{a}) \in \mathcal{X}_\ell$ and $(\mathbf{X}, \mathbf{b}) \in \mathcal{Y}_\ell$.
4. Give the input $(\mathbf{X}, \mathbf{a})$ to Alice, and $(\mathbf{X}, \mathbf{b})$ to Bob.

Note that indeed, for a given choice of $\mathbf{X}$, the pair $(\mathbf{a}, \mathbf{b})$ is sampled from a distribution that is constructed in the same way we constructed hardest distributions in Claim 3.1. We proceed to analyze the information cost of Π with respect to μ:

$$\begin{aligned} \mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{X}, \mathbf{a}, \mathbf{b}) \\ &= I(\Pi : \mathbf{X}) + I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X}). \end{aligned}$$

We lower-bound each of the terms $I(\Pi : \mathbf{X})$ and $I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X})$ separately. The term $I(\Pi | \mathbf{X})$ is at least n, by Lemma 4.3 from the analysis of $R_{\oplus_m \circ \mathrm{U}_n}$: as noted there, the proof for that lemma did not use the fact that $g = \oplus_m$. The rest of this section focuses on proving that $I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X}) \geq \log \mathsf{L}(g) - O(\frac{m \cdot \log m}{n})$.

To this end, we define good and bad matrices X, which intuitively are matrices for which the protocol solves R_g and does not solve R_g, respectively. We will then show that for good matrices X it must hold that $I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X} = X) \geq \log \mathsf{L}(g) - O(1 + \frac{m}{n}) \cdot \log m$, while bad matrices X only occur with low probability and therefore do not affect much $I(\Pi : \mathbf{a}, \mathbf{b} | \mathbf{X})$.

4.2.3 Lower-bounding $I(\mathbf{a},\mathbf{b}|\mathbf{X})$

We turn to define good and bad matrices X. We start with the following auxilary definition of the protocol Π_X, which can be thought of as the protocol for R_g that is obtained from Π by fixing the players' matrices to be X.

Definition 4.8. Let $X \in \{0,1\}^{m\times n}$. Let Π_X be the protocol that is obtained from Π as follows: in the protocol tree of Π, we replace each rectangle $\mathcal{X}_v \times \mathcal{Y}_v$ with the rectangle $\mathcal{X}'_v \times \mathcal{Y}'_v$ defined by

$$\begin{aligned}\mathcal{X}'_v &\overset{\text{def}}{=} \{a : (X,a) \in \mathcal{X}_v\}\\ \mathcal{Y}'_v &\overset{\text{def}}{=} \{b : (X,b) \in \mathcal{Y}_v\}.\end{aligned}$$

Then, we remove all vertices whose rectangles are empty, and merge all pairs of vertices that have identical rectangles.

Definition 4.9. Let $t \overset{\text{def}}{=} \lceil \frac{6m}{n} \rceil + 2$. A matrix $X \in \{0,1\}^{m\times n}$ is good if Π_X is a protocol that solves the relaxed KW problem $R_g(t)$ (see Definition 2.15). Otherwise, we say that X is bad.

Next, we have the following lemma, which shows that whenever $\mathbf{X}$ is good, the protocol must transmit a lot of information.

Lemma 4.10. *For every good matrix X, it holds that $I(\Pi : \mathbf{a},\mathbf{b}|\mathbf{X} = X) \ge \log \mathsf{L}(g) - t\cdot(\log m + 2)$.*

Proof. We start by noting that

$$I(\Pi : \mathbf{a},\mathbf{b}|\mathbf{X} = X) \overset{\text{def}}{=} H(\Pi|\mathbf{X} = X) - H(\Pi|\mathbf{a},\mathbf{b},\mathbf{X} = X) = H(\Pi|\mathbf{X} = X),$$

where the second equality holds since the transcript Π is determined by $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{X}$. Thus, it suffices to lower-bound the entropy $H(\Pi|\mathbf{X} = X)$.

Next, observe that by the definition of μ, it holds that conditioned on $\mathbf{X} = X$, the transcript $\Pi = \Pi\left((\mathbf{X},\mathbf{a}),(\mathbf{X},\mathbf{b})\right)$ is distributed uniformly over the leaves of T_X. Therefore, it suffices to prove that the tree T_X has at least $2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(g)$ leaves.

Finally, observe that the set of leaves of T_X is exactly the set of leaves of the protocol Π_X. Hence, it suffices to prove that $\mathsf{L}(\Pi_X) \ge 2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(g)$. Now, Π_X is a protocol that solves the relaxed KW problem $R_g(t)$. By Proposition 2.19, which says that $R_g(t)$ is not much easier than R_g, it follows that

$$\mathsf{L}(\Pi_X) \ge \mathsf{L}(R_g(t)) \ge 2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(R_g) = 2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(g),$$

as required. ■

In the next subsection, we prove the following lemma, which says that there are not many bad matrices.

Lemma 4.11. *The probability that $\mathbf{X}$ is a bad matrix is at most 2^{-m}.*

We now show that Lemmas 4.10 and 4.11 imply Theorem 1.6. We first observe that $\log \mathsf{L}(g) \le m$, since for every function on m bits it is easy to construct a formula of size 2^m that computes it. Next,

$$\begin{aligned}
\mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{X}) + I(\Pi : \mathbf{a},\mathbf{b}|\mathbf{X})\\
\text{(Lemma 4.3)} \quad &\ge n + I(\Pi : \mathbf{a},\mathbf{b}|\mathbf{X})\\
&\ge n + \Pr\left[\mathbf{X} \text{ is good}\right] \cdot \mathbb{E}_{\text{good } X}\left[I(\Pi : \mathbf{a},\mathbf{b}|\mathbf{X} = X)\right]\\
\text{(Lemma 4.10)} \quad &\ge n + \Pr\left[\mathbf{X} \text{ is good}\right] \cdot (\log \mathsf{L}(g) - t\cdot(\log m + 2))\\
\text{(Lemma 4.11)} \quad &\ge n + (1 - 2^{-m}) \cdot (\log \mathsf{L}(g) - t\cdot(\log m + 2))\\
\text{(Since } \log \mathsf{L}(g) \le m) \quad &= n + \log \mathsf{L}(g) - O(1 + \frac{m}{n}) \cdot \log m,
\end{aligned}$$

as required.

4.2.4 Proof of Lemma 4.11

We prove that the probability that X is a bad matrix is at most 2^{-m}, or in other words, that there are at most $2^{-m} \cdot 2^{m \cdot n}$ bad matrices. The intuition for the proof is the following: Recall that Alice and Bob output $\perp$, and that this means that they have to be convinced that their matrices agree on some row j for which $\mathbf{a}_j \neq \mathbf{b}_j$. However, when X is bad, Alice and Bob do not know an index j such that $\mathbf{a}_j \neq \mathbf{b}_j$ at the end of the protocol. This means that they have to be convinced that they agree on many rows, as otherwise they run the risk of rejecting a legal pair of inputs. But verifying that they agree on many rows is very costly, and they can only do so for few matrices. Details follow.

First, recall that a matrix X is bad if and only if Π_X does not solve the relaxed KW problem $R_g(t)$. This implies that there exists some leaf ℓ' of Π_X, which is labeled with a rectangle $\mathcal{X}'_\ell \times \mathcal{Y}'_\ell$, and a string $a \in \mathcal{X}'_\ell$, such that the following holds:

- For every $\mathcal{J} \subseteq [m]$ such that $|\mathcal{J}| < t$, there exists $b \in \mathcal{Y}'_\ell$ such that $a|_\mathcal{J} = b|_\mathcal{J}$.

Going back from Π_X to Π, it follows that there exists some leaf ℓ of Π, which is labeled with a rectangle $\mathcal{X}_\ell \times \mathcal{Y}_\ell$, and a string $a \in g^{-1}(0)$, such that the following holds:

- $(X, a) \in \mathcal{X}_\ell$.
- For every $\mathcal{J} \subseteq [m]$ such that $|\mathcal{J}| < t$, there exists $b \in g^{-1}(1)$ such that $a|_\mathcal{J} = b|_\mathcal{J}$ and $(X, b) \in \mathcal{Y}_\ell$.

Now, without loss of generality, we may assume that

$$\mathsf{L}(\Pi) \leq \mathsf{L}(g) \cdot 2^n \leq 2^{m+n},$$

since otherwise Theorem 1.6 would follow immediately. Therefore, it suffices to prove that every pair of a leaf ℓ and a string a are "responsible" for at most $2^{-(3 \cdot m+n)} \cdot 2^{m \cdot n}$ bad matrices. This would imply that there are at most $2^{-m} \cdot 2^{m \cdot n}$ bad matrices, by taking union bound over all leaves of Π (at most 2^{m+n}) and all strings a (at most 2^m).

Fix a leaf ℓ of Π and a string $a \in g^{-1}(0)$. Let $\mathcal{T}$ be the set of bad matrices that are supported by ℓ and a. We prove that $|\mathcal{T}| \leq 2^{-(3 \cdot m+n)} \cdot 2^{m \cdot n}$. The key idea is that since Alice does not know a small set $\mathcal{J}$ such that $a|_\mathcal{J} \neq b|_\mathcal{J}$, Alice and Bob must be convinced that their matrices agree on at least t rows. This intuition is made rigorous in the following statement.

Claim 4.12. *Every two matrices $X, X' \in \mathcal{T}$ agree on at least t rows.*

Proof. Let $X, X' \in \mathcal{T}$, and let $\mathcal{J}$ be the set of rows on which they agree. By definition of $\mathcal{T}$, it holds that $(X, a), (X', a) \in \mathcal{T}$. Suppose that $|\mathcal{J}| < t$. Then, by the assumption on ℓ and a, there exists $b \in g^{-1}(1)$ such that $(X, b) \in \mathcal{Y}_\ell$ and $a|_\mathcal{J} = b|_\mathcal{J}$.

Next, observe that if we give the input (X', a) to Alice and (X, b) to Bob, the protocol will reach the leaf ℓ. Now, ℓ is a rejecting leaf, and therefore there must exist some index $j \in [m]$ such that $a_j \neq b_j$ but $X_j = X'_j$. However, we know that $a|_\mathcal{J} = b|_\mathcal{J}$, and therefore $j \notin \mathcal{J}$. It follows that X and Y agree on a row outside $\mathcal{J}$, contradicting the definition of $\mathcal{J}$. ■

Finally, we observe that Claim 4.12 is just another way of saying that $\mathcal{T}$ satisfies the t-agreement property (Definition 2.32), when viewed as a set of strings in $\mathbb{F}^m$ over the alphabet $\mathbb{F} = \{0,1\}^n$. Therefore, Lemma 2.34 implies that $|\mathcal{T}| \leq 2^{(m-t)\cdot n}$. Wrapping up, it follows that

$$\begin{aligned} |\mathcal{T}| &\leq 2^{(m-t)\cdot n} \\ &\leq 2^{(m-\frac{3m}{n}-1)\cdot n} \\ &= \frac{1}{2^{3\cdot m+n}} \cdot 2^{m\cdot n}, \end{aligned}$$

as required.

Remark 4.13. Note that Lemma 2.34 can only be applied if $m \leq 2^n$. However, this can be assumed without loss of generality, since for $m \geq 2^n$, the lower bound of Theorem 1.6 becomes less than $\log \mathsf{L}(g)$. However, it is easy to prove a lower bound of $\log \mathsf{L}(g)$ on $\log \mathsf{L}(R_{g\circ \mathrm{U}_n})$ by reducing R_g to $R_{g\circ \mathrm{U}_n}$.

5 A Candidate Hard Distribution for $R_{\oplus_m \circ f}$

In the introduction, we suggested proving the following conjecture as a step toward proving the KRW conjecture.

Conjecture (1.9, restated)**.** *For every function $f : \{0,1\}^n \to \{0,1\}$ and every $m \in \mathbb{N}$, it holds that*

$$\mathsf{L}(\oplus_m \circ f) = \tilde{\Omega}\left(m^2 \cdot \mathsf{L}(f)\right).$$

We could try to prove this conjecture using the information-complexity approach that is suggested in this paper. This would require us to show that for every protocol Π of $R_{\oplus_m \circ f}$, there exists a distribution μ such that

$$\mathsf{IC}_{\mu}(\Pi) \geq 2\log m + \log \mathsf{L}(f) - O(\log\log m + \log\log \mathsf{L}(f)). \tag{9}$$

In this section, we propose a way for constructing such a distribution μ for every protocol $R_{\oplus_m \circ f}$, which we believe to be a good candidate for facilitating such a proof. In particular, the distributions μ that we construct have a few properties that should be useful for such a proof, as discussed below. We note that our construction only works when f is balanced and hard on average (see Definition 5.2), but this can be assumed for our target applications.

The rest of this section is organized as follows. In Section 5.1, we provide motivation for our construction of μ, explain what are the useful properties that it has, and outline the construction. Then, in Section 5.2, we prove a general theorem about the construction of hardest distributions for average-case hard functions, which is used in the construction of μ. Finally, in Section 5.3, we construct μ for every protocol Π, and prove that it has the required properties.

5.1 Motivation and outline

We start by discussing a naive approach for constructing hard distributions for protocols for $R_{\oplus_m \circ f}$, which follows the lines of the construction of hard distributions for $R_{\vee_m \circ f}$ that was given in Section 3.2.1. We then discuss the shortcomings of this approach, and how we resolve one of them in our construction.

Recall that the relation $R_{\oplus_m \circ f}$ corresponds to the following communication problem: Alice and Bob get as inputs matrices $X, Y \in \{0,1\}^{m\times n}$, respectively, such that the strings $f(X), f(Y) \in \{0,1\}^m$ have even and odd weights, respectively. Their goal is to find an entry on which X and Y differ.

5.1.1 A naive construction

One straightforward way to construct a hard distribution for a protocol Π for $R_{\oplus_m \circ f}$ would be to try to combine a hardest distribution for R_f, and the hard distribution for $R_{\oplus_m}$ (which is to choose a random edge in the hypercube, see Section 3.1.1). To this end, we need the following notation:

- We denote the all-zeroes string and all-ones string by $\bar{0}$ and $\bar{1}$, respectively, and without loss of generality, we assume that $f(\bar{0}) = 0$ and $f(\bar{1}) = 1$.

Now, for a fixed protocol Π for $R_{\oplus_m \circ f}$, consider the distribution μ that is sampled as follows:

1. Let $\mathbf{a}$ and $\mathbf{b}$ be uniformly distributed strings of even and odd weights, respectively, which differ on a unique index $\mathbf{j} \in [m]$ that is chosen uniformly at random.

2. Let $\Pi_{\mathbf{a},\mathbf{b}}$ be the protocol for R_f obtained from Π by fixing each X_k and Y_k (for $k \in [m] - \{\mathbf{j}\}$) to $\bar{0}$ if $\mathbf{a}_k = 0$, and to $\bar{1}$ if $\mathbf{a}_k = 1$ (recall that $\mathbf{a}_k = \mathbf{b}_k$).

3. Sample a pair $(\mathbf{x_j}, \mathbf{y_j})$ of inputs for R_f from a hardest distribution for $\Pi_{\mathbf{a},\mathbf{b}}$, which exists by Corollary 3.2.

4. Output the pair $(\mathbf{X}, \mathbf{Y})$ of inputs for R_f, where

   (a) $\mathbf{X}_k = \mathbf{Y}_k = \bar{0}$ for every $k \in [m] - \{\mathbf{j}\}$ such that $\mathbf{a}_k = 0$;

   (b) $\mathbf{X}_k = \mathbf{Y}_k = \bar{1}$ for every $k \in [m] - \{\mathbf{j}\}$ such that $\mathbf{a}_k = 1$.

   (c) $\mathbf{X_j} = \mathbf{x_j}$, $\mathbf{Y_j} = \mathbf{y_j}$ if $\mathbf{a_j} = 0$, $\mathbf{b_j} = 1$; or $\mathbf{X_j} = \mathbf{y_j}$, $\mathbf{Y_j} = \mathbf{x_j}$ if $\mathbf{a_j} = 1$, $\mathbf{b_j} = 0$

Observe that in this distribution, it holds that $f(\mathbf{X}) = \mathbf{a}$ and $f(\mathbf{Y}) = \mathbf{b}$, and in particular $f(\mathbf{X})$ and $f(\mathbf{Y})$ indeed form a random edge in the boolean hypercube. One could hope to prove a lower bound of $2\log m + \log \mathsf{L}(f)$ on the information cost $\mathsf{IC}_\mu(\Pi)$ by an argument of the following form:

1. By the end of the protocol, both Alice and Bob must learn the index $\mathbf{j}$. Therefore, each of them must learn $\log m$ bits.

2. Even after the players know $\mathbf{j}$, they still have to solve Π_f on $\mathbf{X_j}$ and $\mathbf{Y_j}$, and therefore must transmit additional $\log \mathsf{L}(f)$ bits of information.

We do not know how to implement such an argument. As an example for how such an argument could be implemented in principle, consider the following (false) proof:

$$\begin{aligned}
\mathsf{IC}_\mu(\Pi) &= I(\Pi : \mathbf{X}, \mathbf{Y}) \\
\text{(Since } \mathbf{a} \text{ and } \mathbf{b} \text{ are determined by } \mathbf{X} \text{ and } \mathbf{Y}) &= I(\Pi : \mathbf{X}, \mathbf{Y}, \mathbf{a}, \mathbf{b}) \\
\text{(The chain rule)} &= I(\Pi : \mathbf{a}, \mathbf{b}) + I(\Pi : \mathbf{X}, \mathbf{Y} | \mathbf{a}, \mathbf{b}) \\
\text{(External information } \geq \text{internal information)} &\geq I(\Pi : \mathbf{a} | \mathbf{b}) + I(\Pi : \mathbf{b} | \mathbf{a}) \qquad (10) \\
&\quad + I(\Pi : \mathbf{X}, \mathbf{Y} | \mathbf{a}, \mathbf{b}) \\
&= 2\log m + I(\Pi : \mathbf{X}, \mathbf{Y} | \mathbf{a}, \mathbf{b}) \\
(\Pi \text{ behaves like } \Pi_{\mathbf{a},\mathbf{b}} \text{ conditioned on } \mathbf{a} \text{ and } \mathbf{b}) &= 2\log m + I(\Pi_{\mathbf{a},\mathbf{b}} : \mathbf{X_j}, \mathbf{Y_j}) \\
(\mathbf{X_j} \text{ and } \mathbf{Y_j} \text{ are a hardest distribution for } \Pi_{\mathbf{a},\mathbf{b}}) &\geq 2\log m + \log \mathsf{L}(f).
\end{aligned}$$

The error in the above derivation is in Inequality 10: Lemma 2.31, which says that the external information cost is at least the internal information cost, only holds when the information is measured with respect to the (whole) players' inputs. However, in Inequality 10, the strings $\mathbf{a}$ and $\mathbf{b}$ are not the players inputs[8], and therefore Lemma 2.31 cannot be applied.

In other words, we can apply Lemma 2.31 to the external information $I(\Pi : \mathbf{X}, \mathbf{Y}, \mathbf{a}, \mathbf{b})$, but not to the information $I(\Pi : \mathbf{a}, \mathbf{b})$. In fact, it is possible to construct a protocol Π in which $I(\Pi : \mathbf{a}, \mathbf{b})$ is smaller than $I(\Pi : \mathbf{a}|\mathbf{b}) + I(\Pi : \mathbf{b}|\mathbf{a})$, by making the players "encrypt" their messages about $\mathbf{a}$ and $\mathbf{b}$, using the bits of $\mathbf{X}$ and $\mathbf{Y}$ as "keys".

5.1.2 Motivation for our construction of μ

One obstacle toward implementing the above approach is the following: even if we could prove that Alice and Bob must learn the index $\mathbf{j}$ in order to solve the relation, it would not imply that each of them must learn $\log m$ bits of information. For example, in the way we constructed μ above, it is likely that Alice can deduce[9] $\mathbf{j}$ directly from her matrix $\mathbf{X}$, since the $\mathbf{j}$-th row is likely to be the unique row that is not $\overline{0}$ or $\overline{1}$. Therefore, she does not need to receive any bits from Bob about $\mathbf{j}$.

Our contribution in this paper is modifying the above construction of μ such that each of the matrices $\mathbf{X}$ and $\mathbf{Y}$ on its own does not reveal much about $\mathbf{j}$. Formally, we prove the following result.

Definition 5.1. Let $X \in \{0,1\}^{m\times n}$. Then, we denote by X_{-j} the $(m-1)\times n$ matrix that is obtained from X by removing the j-th row.

Definition 5.2 (Average-case hardness)**.** Let $f : \{0,1\}^n \to \{0,1\}$. We say that f is (s,ε)-hard if for every formula ϕ of size at most s it holds that $\Pr_{x\leftarrow\{0,1\}^n}[\phi(x) = f(x)] \le \frac{1}{2} + \varepsilon$ (where x is uniformly distributed over $\{0,1\}^n$).

Theorem 5.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a balanced and $\left(s, \frac{1}{4}\right)$-hard function, and let $m \in \mathbb{N}$. For every protocol Π there exists a distribution μ over inputs $(\mathbf{X}, \mathbf{Y})$ for $R_{\oplus_m \circ f}$ that satisfies the following properties:*

1. *The strings $f(\mathbf{X}), f(\mathbf{Y}) \in \{0,1\}^m$ are uniformly distributed strings of even and odd weights, respectively, which differ on a unique index $\mathbf{j} \in [m]$, which is uniformly distributed.*
2. *It holds that $H(\mathbf{j}|\mathbf{X}) \ge \log m - O(\log\log m)$ and $H(\mathbf{j}|\mathbf{Y}) \ge \log m - O(\log\log m)$.*
3. *It always holds that $\mathbf{X}_{-\mathbf{j}} = \mathbf{Y}_{-\mathbf{j}}$.*
4. *The distribution μ is hard for Π and R_f conditioned on $\mathbf{j}$ and on $\mathbf{X}_{-\mathbf{j}}, \mathbf{Y}_{-\mathbf{j}}$: for every $j \in [m]$, and $W \in \{0,1\}^{(m-1)\times n}$ it holds that*
$$\mathsf{IC}_{\mu_{j,W}}(\Pi : \mathbf{X}_{\mathbf{j}}, \mathbf{Y}_{\mathbf{j}} | \mathbf{j} = j, \mathbf{X}_{-\mathbf{j}} = \mathbf{Y}_{-\mathbf{j}} = W) \ge \log s.$$

Remark 5.4. The motivation for requiring that $\mathbf{X}_{-\mathbf{j}} = \mathbf{Y}_{-\mathbf{j}}$ in Theorem 5.3 is the following. Recall that the players are required to output an entry of $\mathbf{X}$ and $\mathbf{Y}$ on which they differ. The requirement $\mathbf{X}_{-\mathbf{j}} = \mathbf{Y}_{-\mathbf{j}}$ implies that this entry must belong to the $\mathbf{j}$-th row of $\mathbf{X}$ and $\mathbf{Y}$. Hopefully, this would force the players to solve the KW relation R_f on the $\mathbf{j}$-th rows of $\mathbf{X}$ and $\mathbf{Y}$, which are sampled from a hard distribution.

[8] The reader may wonder why this was not a problem when we applied Lemma 2.31 to $\mathbf{a}$ and $\mathbf{b}$ in Section 4.1.2, even though $\mathbf{a}$ and $\mathbf{b}$ were only parts of the players' inputs. The reason is that there the mutual information was conditioned on $\mathbf{X}$, and under this conditioning $\mathbf{a}$ and $\mathbf{b}$ could be viewed as the whole inputs of the players.

[9] Note that this was not a problem when we considered $R_{\vee_m \circ f}$ in Section 3.2.1, since there the argument relied on the fact that *an external observer* must learn $\mathbf{j}$, and the external observer does not know $\mathbf{X}$ and $\mathbf{Y}$. However, in the current context we wish to prove that Alice and Bob have to learn $\mathbf{j}$, since we wish to prove a lower bound that involves $2\log m$ rather than $\log m$.

5.1.3 Construction outline

We now sketch the proof for Theorem 5.3. Fix a balanced function $f : \{0,1\}^n \to \{0,1\}$ and a protocol Π for $R_{\oplus_m \circ f}$. As a warm-up, suppose we could make the following simplifying assumption:

- For every protocol Π_f for R_f, there exists a hardest distribution ν of Π_f, whose marginals over x and y are uniform over $f^{-1}(0)$ and $f^{-1}(1)$, respectively.

In this case, we could prove Theorem 5.3 rather easily. We would construct μ in the same way we did in Section 5.1.1, with the following modifications:

1. First, instead of setting $\mathbf{X}_k$ and $\mathbf{Y}_k$ to $\overline{0}$ or $\overline{1}$ (for $k \in [m] - \{\mathbf{j}\}$), we choose $\mathbf{X}_k = \mathbf{Y}_k$ to be a uniformly distributed string in $f^{-1}(\mathbf{a}_k)$. Let $\Pi_{\mathbf{j},\mathbf{X}_{-\mathbf{j}},\mathbf{Y}_{-\mathbf{j}}}$ denote the protocol for R_f obtained from Π after this fixing.

2. Second, we choose $\mathbf{X}_\mathbf{j}$ and $\mathbf{Y}_\mathbf{j}$ from the hardest distribution ν of $\Pi_{\mathbf{j},\mathbf{X}_{-\mathbf{j}},\mathbf{Y}_{-\mathbf{j}}}$ that is given by the simplifying assumption.

Clearly, this construction of μ satisfies the requirements of Theorem 5.3, and in particular:

- $\mathbf{X}$ does not give any information on $\mathbf{j}$. To see it, observe that the rows of $\mathbf{X}$ are distributed are distributed like uniform and independent strings in $\{0,1\}^n$, regardless of $\mathbf{j}$. The same goes for $\mathbf{Y}$. This implies that $H(\mathbf{j}|\mathbf{X}) = H(\mathbf{j}|\mathbf{Y}) = \log m$.

- For every $j \in [m]$, and $W \in \{0,1\}^{(m-1)\times n}$, it holds that

  $$\mathsf{IC}_{\mu_{j,W}}(\Pi : \mathbf{X}_\mathbf{j}, \mathbf{Y}_\mathbf{j} | \mathbf{j} = j, \mathbf{X}_{-\mathbf{j}} = \mathbf{Y}_{-\mathbf{j}} = W) \geq \log \mathsf{L}(f),$$

  since on the left-hand side, $\mathbf{X}_\mathbf{j}$ and $\mathbf{Y}_\mathbf{j}$ are distributed according to ν, which is a hardest distribution for R_f.

Our actual construction of μ is similar, and proceeds in two steps.

- First, we prove the following relaxed version of the above simplifying assumption: for every *average-case hard* function f, and a protocol Π_f for R_f, there exists a hard distribution ν, whose marginals over x and y are uniform over *large subsets* of $f^{-1}(0)$ and $f^{-1}(1)$, respectively. This is done in Section 5.2. Intuitively, this relaxed version is true because the average-case hardness of f implies that many x's and y's must occur in many distinct leaves of Π_f, and therefore it is possible to construct a hard distribution that is uniform over many x's and y's while simultaneously reaching many leaves of Π_f.

- Second, we use the same construction of μ as we did with the simplifying assumption, and prove that even under the relaxed version, the matrices $\mathbf{X}$ and $\mathbf{Y}$ still do not reveal much about $\mathbf{j}$. This is done in Section 5.3.

5.2 On hardest distributions for average-case hard functions

In this section, we prove the following result.

Theorem 5.5. *Let f be a balanced (s, ε)-hard function, and let Π be a KW protocol for R_f. Then, there exists a distribution ν over inputs (x, y) for R_f such that*

- *The marginals of ν over x and y are uniform over subsets of $f^{-1}(0)$ and $f^{-1}(1)$ of density exactly*[10] $\frac{1}{2}-\varepsilon$, *respectively.*
- *The distribution on leaves of* Π *induced by* ν *has (min-)entropy at least* $\log s$.

Remark 5.6. We note that the assumption that f is balanced is not crucial, and is made for convenience. We also note that the guarantee of the density being *exactly* $\frac{1}{2}-\varepsilon$ (rather than *at least*) is not crucial, but will make the use of this theorem in the next section more convenient.

Fix a balanced (s,ε)-hard function $f:\{0,1\}^n \to \{0,1\}$, and a protocol Π for R_f. We construct the corresponding distribution ν. We say that a leaf ℓ of Π contains a pair (x,y) if (x,y) belongs to the rectangle $\mathcal{X}_\ell \times \mathcal{Y}_\ell$ of ℓ.

The basic idea of our construction of ν is as follows: The distribution ν will be uniform over a set $\mathcal{P}$ of pairs (x,y). The set $\mathcal{P}$ will satisfy the following properties:

1. Every $x \in f^{-1}(0)$ and $y \in f^{-1}(1)$ belong to at most one pair of $\mathcal{P}$.
2. $\mathcal{P}$ is of size exactly $(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}$.
3. Every leaf ℓ contains at most $\frac{(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}}{s}$ pairs of $\mathcal{P}$.

The first two properties will guarantee that ν satisfies the first requirement of Theorem 5.5, and the third property will guarantee that it satisfies the second requirement. We construct the set $\mathcal{P}$ greedily. More specifically:

1. The greedy process works in multiple phases, where each phase consists of multiple iterations.
2. In each iteration, we pick a pair (x,y) such that x and y were never used before, and such that the pair (x,y) belongs to a leaf ℓ that was not used before in the current phase.
3. Once it is no longer possible to choose a pair (x,y) from an unused leaf, the current phase ends and a new phase starts, which allows us to use again the same leaves.

The fact that f is average-case hard will guarantee that this process can continue without reusing the same leaves too much, as long as $\mathcal{P}$ is smaller than $(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}$.

We turn to the formal proof. Our first step is to prove the following claim, which intuitively says that many x's and y's occur in many distinct leaves of Π. We will later use this claim to show that the greedy process can choose many pairs (x,y) without reusing the same leaves.

Claim 5.7. *For every two sets $\mathcal{S} \subseteq f^{-1}(0)$, $\mathcal{T} \subseteq f^{-1}(1)$ for which $\frac{|\mathcal{S}|+|\mathcal{T}|}{2^n} > \frac{1}{2}+\varepsilon$, it holds that the minimal set of leaves of* Π *that contains $\mathcal{S}\times\mathcal{T}$ is of size greater than s.*

Proof. We prove the claim by showing that if it was false, it would have been possible to construct a formula of size s that computes f with probability greater than $\frac{1}{2}+\varepsilon$. Let $\mathcal{S} \subseteq f^{-1}(0)$, $\mathcal{T} \subseteq f^{-1}(1)$ be such that $\frac{|\mathcal{S}|+|\mathcal{T}|}{2^n} > \frac{1}{2}+\varepsilon$, and let l be the number of leaves containing pairs from $\mathcal{S}\times\mathcal{T}$. We would like to prove that $l > s$. We construct a protocol Π' for the restriction of R_f to $\mathcal{S}\times\mathcal{T}$ as follows:

1. For each node v of Π, we replace the rectangle $\mathcal{X}_v \times \mathcal{Y}_v$ with the rectangle $(\mathcal{X}_v \cap \mathcal{S}) \times (\mathcal{Y}_v \cap \mathcal{T})$.
2. We remove from Π all the vertices v that are left with an empty rectangle.

[10] We assume here that $(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}$ is an integer. Otherwise, the size of those subsets is exactly $\lfloor(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}\rfloor$.

3. If a vertex v of Π and its parent u have the same rectangle, we merge them (note that after the previous step, v must be the only child of u).

It is not hard to see that Π' is a correct protocol for the restriction of R_f to $\mathcal{S} \times \mathcal{T}$, and that it has l leaves. Now, let ϕ' be the formula obtained from Π' using the KW connection (Theorem 2.12). Then, ϕ' separates $\mathcal{S}$ and $\mathcal{T}$, and thus, computes f correctly on $\mathcal{S}$ and $\mathcal{T}$. But, this implies that

$$\Pr\left[\phi(U_n) = f(U_n)\right] \geq \Pr\left[U_n \in \mathcal{S} \cup \mathcal{T}\right] > \frac{1}{2} + \varepsilon.$$

It follows that ϕ' is a formula of size at most l that computes f correctly with probability greater than $\frac{1}{2} + \varepsilon$, and hence $l > s$ (since by assumption f is (s, ε)-hard). ■

We turn to constructing the distribution ν. As discussed above, the distribution ν will be uniform over a set $\mathcal{P}$ of pairs (x, y), which is constructed by the following process:

1. Set $\mathcal{S} = f^{-1}(0)$, $\mathcal{T} = f^{-1}(1)$.
2. **Phases loop:** While $\mathcal{S}$ and $\mathcal{T}$ have density greater than $\frac{1}{2} + \varepsilon$ in $f^{-1}(0)$ and $f^{-1}(1)$, respectively:
   (a) Set $\mathcal{L}$ to be the set of leaves of Π.
   (b) **Iterations loop:** While there exists $(x, y) \in \mathcal{S} \times \mathcal{T}$ that is contained in a leaf $\ell \in \mathcal{L}$, and the condition of Step 2 holds:
      i. Add (x, y) to the support of ν.
      ii. Remove x from $\mathcal{S}$ and y from $\mathcal{T}$.
      iii. Remove ℓ from $\mathcal{L}$.

We now use Claim 5.7 to show that each phase, except for the last one, has at least s iterations. We will later use it to upper-bound the number of phases, and this will imply an upper bound on the number of pairs in $\mathcal{P}$ that a single leaf ℓ may contain.

Claim 5.8. *Every phase, probably except for the last one, has at least s iterations.*

Proof. Fix a phase that is not the last phase, and let $\mathcal{L}'$ be the set of leaves removed in this phase. We would like to to prove that $|L'| \geq s$. Since the phase we fixed is not the last phase, we know that at the end of this phase, the density of both $\mathcal{S}$ and $\mathcal{T}$ is greater than $\frac{1}{2} + \varepsilon$. By Claim 5.7, this implies that $\mathcal{S} \times \mathcal{T}$ is not contained in any set of leaves of size less than s. On the other hand, we know that $\mathcal{S} \times \mathcal{T}$ is contained in $\mathcal{L}'$, since this is the meaning of the stopping condition of the iterations loop. It follows that $|L'| \geq s$, as required. ■

We conclude by proving that ν has the required properties. First, we note that the above process must halt, since every iteration removes an element from $\mathcal{S}$ and an element from $\mathcal{T}$, and $\mathcal{S}$ and $\mathcal{T}$ are finite. Furthermore, due to the stopping conditions, the set $\mathcal{P}$ must be of size $\left(\frac{1}{2} - \varepsilon\right) \cdot 2^{n-1}$ at the end of the process. Moreover, it should be clear that every x and every y participates in at most one pair in $\mathcal{P}$. Together, those assertions imply that the marginals of ν over x and y are uniform over subsets of $f^{-1}(0)$ and $f^{-1}(1)$ of density exactly $\frac{1}{2} - \varepsilon$, respectively, as required by the theorem.

We turn to prove that ν satisfies the second requirement of the theorem. To this end, we upper bound the probability that every leaf has under the distribution that is induced by ν on the leaves of Π. The key point is that, by Claim 5.8, every phase lasts for at least s iterations, and therefore there are at most $\frac{\left(\frac{1}{2}-\varepsilon\right)\cdot 2^{n-1}}{s}$ phases (since the total number of iterations is $\left(\frac{1}{2}-\varepsilon\right)\cdot 2^{n-1}$). This means that every leaf contains at most $\frac{\left(\frac{1}{2}-\varepsilon\right)\cdot 2^{n-1}}{s}$ pairs of $\mathcal{P}$. Since the total number of pairs in $\mathcal{P}$ is $(\frac{1}{2}-\varepsilon)\cdot 2^{n-1}$, it follows that the probability of each leaf is at most

$$\frac{\frac{\left(\frac{1}{2}-\varepsilon\right)\cdot 2^{n-1}}{s}}{\left(\frac{1}{2}-\varepsilon\right)\cdot 2^{n-1}}=\frac{1}{s}.$$

This implies that the (min-)entropy of the distribution over the leaves is at least $\log s$, as required.

5.3 Construction of hard distributions for $R_{\oplus_m \circ f}$

We turn to prove Theorem 5.3, restated next.

Theorem 5.3. *Let $f:\{0,1\}^n\to\{0,1\}$ be a balanced and $\left(s,\frac{1}{4}\right)$-hard function, and let $m\in\mathbb{N}$. For every protocol Π there exists a distribution μ over inputs $(\mathbf{X},\mathbf{Y})$ for $R_{\oplus_m\circ f}$ that satisfies the following properties:*

1. *The strings $f(\mathbf{X}),f(\mathbf{Y})\in\{0,1\}^m$ are uniformly distributed strings of even and odd weights, respectively, which differ on a unique index $\mathbf{j}\in[m]$, which is uniformly distributed.*
2. *It holds that $H(\mathbf{j}|\mathbf{X})\geq\log m-O(\log\log m)$ and $H(\mathbf{j}|\mathbf{Y})\geq\log m-O(\log\log m)$.*
3. *It always holds that $\mathbf{X}_{-\mathbf{j}}=\mathbf{Y}_{-\mathbf{j}}$.*
4. *The distribution μ is hard for Π and R_f conditioned on $\mathbf{j}$ and on $\mathbf{X}_{-\mathbf{j}},\mathbf{Y}_{-\mathbf{j}}$: for every $j\in[m]$, and $W\in\{0,1\}^{(m-1)\times n}$ it holds that*

$$\mathsf{IC}_{\mu_{j,W}}(\Pi:\mathbf{X}_\mathbf{j},\mathbf{Y}_\mathbf{j}|\mathbf{j}=j,\mathbf{X}_{-\mathbf{j}}=\mathbf{Y}_{-\mathbf{j}}=W)\geq\log s.$$

Let $m\in\mathbb{N}$. Fix a balanced and $(s,\frac{1}{4})$-hard function $f:\{0,1\}^n\to\{0,1\}$, and a protocol Π for $R_{\oplus_m\circ f}$. The distribution μ over inputs $(\mathbf{X},\mathbf{Y})$ is sampled as follows:

1. Let $\mathbf{a}$ and $\mathbf{b}$ be uniformly distributed strings of even and odd weights, respectively, which differ on a unique index $\mathbf{j}\in[m]$ that is chosen uniformly at random.
2. For every $k\in[m]-\{\mathbf{j}\}$, let $\mathbf{w}_k$ be a uniformly distributed string in $f^{-1}(\mathbf{a}_k)=f^{-1}(\mathbf{b}_k)$ (recall that $\mathbf{a}_k=\mathbf{b}_k$). Let $\mathbf{W}$ be the matrix whose rows are the strings $\mathbf{w}_k$.
3. Let $\Pi_{\mathbf{j},\mathbf{W}}$ be the protocol for R_f obtained from Π by fixing $X_{-\mathbf{j}}=Y_{-\mathbf{j}}=W$ in Alice and Bob's inputs.
4. Sample a pair $(\mathbf{x}_\mathbf{j},\mathbf{y}_\mathbf{j})$ of inputs for R_f from the distribution $\nu_{\mathbf{j},\mathbf{W}}$ obtained for $\Pi_{\mathbf{j},\mathbf{W}}$ using Theorem 5.5.
5. Output the pair $(\mathbf{X},\mathbf{Y})$ of inputs for R_f , where $\mathbf{X}_{-\mathbf{j}}=\mathbf{Y}_{-\mathbf{j}}=\mathbf{W}$, and where
   (a) $\mathbf{X}_\mathbf{j}=\mathbf{x}_\mathbf{j}$, $\mathbf{Y}_\mathbf{j}=\mathbf{y}_\mathbf{j}$ if $\mathbf{a}_\mathbf{j}=0$ and $\mathbf{b}_\mathbf{j}=1$.
   (b) $\mathbf{X}_\mathbf{j}=\mathbf{y}_\mathbf{j}$, $\mathbf{Y}_\mathbf{j}=\mathbf{x}_\mathbf{j}$ if $\mathbf{a}_\mathbf{j}=1$ and $\mathbf{b}_\mathbf{j}=0$.

Observe that the distribution μ satisfies the first requirement of Theorem 5.3 since $f(\mathbf{X}) = \mathbf{a}$, $f(\mathbf{Y}) = \mathbf{b}$, and $(\mathbf{a}, \mathbf{b})$ were chosen as required by the theorem. Moreover, observe that the third requirement holds by definition, and the fourth requirement holds since $\nu_{j,W}$ is the hardest distribution obtained for $\Pi_{j,W}$ using Theorem 5.5.

In the rest of this section, we show that μ satisfies the second requirement, namely, that $H(\mathbf{j}|\mathbf{X})$ and $H(\mathbf{j}|\mathbf{Y})$ are at least $\log m - O(\log\log m)$. We only prove the lower bound for $H(\mathbf{j}|\mathbf{X})$, and a lower bound for $H(\mathbf{j}|\mathbf{Y})$ can be proved in a similar way. We actually prove something stronger, that is, that for every $a \in \{0,1\}^m$ of even weight, it holds that

$$H(\mathbf{j}|\mathbf{X}, \mathbf{a} = a) \geq \log m - O(\log\log m). \tag{11}$$

This implies that

$$H(\mathbf{j}|\mathbf{X}) \geq H(\mathbf{j}|\mathbf{X}, \mathbf{a}) \geq \log m - O(\log\log m),$$

as required. Without loss of generality, we prove Inequality 11 for the case where a is the all-zeroes string $\bar{0}$. For the rest of this section, all the events are conditioned on $\mathbf{a} = \bar{0}$, but we omit this conditioning for the sake of brevity.

5.3.1 High level idea

In order to explain the basic idea of the proof, we use the following notion of a "good index". Intuitively, an index j is good for a matrix X if conditioned on $\mathbf{X} = X$, it holds that j is a possible value for $\mathbf{j}$. Formally,

Definition 5.9. Let X be in the support of $\mathbf{X}$, and let $j \in [m]$. We say that j is a good index for X if X_j is in the support of $\mathbf{x}_j$, where $\mathbf{x}_j$ is drawn from the distribution $\nu_{j,X_{-j}}$ of Step 4 above.

The starting point for the proof is the observation that, conditioned on $\mathbf{X} = X$, the index $\mathbf{j}$ is distributed uniformly over all the good indices for X (see Claim 5.11 below). This means that in order to show that $H(\mathbf{j}|\mathbf{X})$ is large, it suffices to prove that with high probability over $\mathbf{X}$, there are many good indices. The bulk of the proof will focus on showing the latter claim.

In order to get intuition for how the random variable $\mathbf{X}$ behaves, it is useful to consider a second random variable, which we denote $\mathbf{X}'$. The random variable $\mathbf{X}'$ is a matrix whose rows are chosen uniformly and independently from $f^{-1}(0)$. The variables $\mathbf{X}$ and $\mathbf{X}'$ are tightly related: specifically, $\mathbf{X}$ is distributed like $\mathbf{X}'$ conditioned on a random index $\mathbf{j}$ being good (see Claim 5.10 below), So, let us start by trying to understand how many good indices $\mathbf{X}'$ usually has.

First, observe that for every fixed index $j \in [m]$, the probability that j is good for $\mathbf{X}'$ is exactly $\frac{1}{4}$: the reason is that, for every fixed X_{-j}, the support of $\mathbf{x}_j$ is of density exactly $\frac{1}{4}$ (where $\mathbf{x}_j$ is drawn from the distribution $\nu_{j,X_{-j}}$).Thus, in expectation, the matrix $\mathbf{X}'$ has $\frac{1}{4} \cdot m$ good indices.

Now, we would have liked to use a concentration bound to argue that $\mathbf{X}'$ has $\frac{1}{4} \cdot m$ good indices with high probability. Unfortunately, the events of the form "j is a good index" for $j \in [m]$ are not independent. For example, it could be the case that with probability $\frac{1}{4}$, all the indices are good for $\mathbf{X}'$, and otherwise, none of them are good. More generally, the events can be positively correlated, in which case a concentration bound cannot be applied. Therefore, it is not true in general that $\mathbf{X}'$ usually has many good indices.

The crux of our argument is that while positive correlation is an issue for $\mathbf{X}'$, it is not an issue for $\mathbf{X}$. The reason is that $\mathbf{X}$ is distributed like $\mathbf{X}'$ *conditioned on a random index being good*, and this interacts well with the positive correlation. For example, consider again the above case $\mathbf{X}'$ has m good indices with probability $\frac{1}{4}$, and none otherwise. In this case, the matrix $\mathbf{X}$ will have m good indices *with probability* 1, since the conditioning that a random index is good mean that $\mathbf{X}$

is always sampled from the "good event". This argument is implemented formally in Claims 5.12 and 5.13 below.

5.3.2 The formal proof

We begin with some notation. We let $\mathbf{X}'$ denote the random matrix in which the rows are independent and uniformly distributed over $f^{-1}(0)$. For each $j \in [m]$, we denote by G_j the event that the index j is a good index for $\mathbf{X}'$. We denote by M and M' the events that there are at least $\frac{m}{\log^2 m}$ good indices for $\mathbf{X}$ and $\mathbf{X}'$, respectively. We will prove that $\Pr[M] \geq 1 - \frac{5}{\log m}$, and show that this implies Inequality 11. We start by proving the following claim, which relates the distribution of $\mathbf{X}$ to the distribution of $\mathbf{X}'$.

Claim 5.10. *For every $j \in [m]$, the distribution $\mathbf{X}|\mathbf{j} = j$ is identical to the distribution $\mathbf{X}'|G_j$.*

Proof. Fix $j \in [m]$. The proof amounts to observing that both distributions are the uniform distribution over the set $\mathcal{S}$, defined as follows. The set $\mathcal{S}$ is the set of matrices X such that

1. All the the rows of X belong to $f^{-1}(0)$.

2. The row X_j belongs to the support of the distribution $\nu_{j,X_{-j}}$ on $\mathbf{x}_j$.

Note that $|\mathcal{S}| = \left(2^{n-1}\right)^{m-1} \cdot \left(\frac{1}{4} \cdot 2^{n-1}\right)$: the reason is that to choose a matrix $X \in \mathcal{S}$, there are $\left(2^{n-1}\right)^{m-1}$ possibilities for choosing the rows outside the j-th row, and then there are $\frac{1}{4} \cdot 2^{n-1}$ possibilities to choose the j-th row, since by Theorem 5.5 the marginal of $\nu_{j,X_{-j}}$ on $\mathbf{x}_j$ is of density exactly $\frac{1}{4}$.

On the one hand, it is easy to see that $\mathbf{X}'|G_j$ is the uniform distribution over $\mathcal{S}$: to see it, observe that $\mathbf{X}'$ is the uniform distribution over all matrices that satisfy the first condition above, and that G_j is just the event that $\mathbf{X}'$ belongs to $\mathcal{S}$.

On the other hand, showing that $\mathbf{X}|\mathbf{j} = j$ is uniform over $\mathcal{S}$ is a straightforward calculation. We would like to show that $\mathbf{X}|\mathbf{j} = j$ picks each matrix in $\mathcal{S}$ with probability $4/\left(2^{n-1}\right)^m$. To this end, observe that $\mathbf{X}|\mathbf{j} = j$ is sampled as follows: first, one chooses a random matrix $\mathbf{W} \in \{0,1\}^{(m-1)\times n}$ whose rows are all in $f^{-1}(0)$. Then, one sets $\mathbf{X}_{-j} = \mathbf{W}$ and sets $\mathbf{X}_j$ according to the distribution $\nu_{j,\mathbf{W}}$. Hence, for every matrix $X \in \mathcal{S}$ it holds that

$$\begin{aligned}
\Pr[\mathbf{X} = X|\mathbf{j} = j] &= \Pr[\mathbf{W} = X_{-j}] \cdot \Pr\left[\nu_{j,X_{-j}} = X_j\right] \\
&= \frac{1}{\left(2^{n-1}\right)^{m-1}} \cdot \Pr\left[\nu_{j,X_{-j}} = X_j\right] \\
&= \frac{1}{\left(2^{n-1}\right)^{m-1}} \cdot \frac{1}{\frac{1}{4} \cdot 2^{n-1}},
\end{aligned}$$

where the third inequality is since by Theorem 5.5, the marginal of $\nu_{j,X_{-j}}$ over $\mathbf{x}_j$ is uniform over its support, and this support is of density exactly $\frac{1}{4}$. ■

Next, we prove that given $\mathbf{X}$, the index $\mathbf{j}$ is uniformly distributed over all the good indices for $\mathbf{X}$. This means that in order to lower-bound the entropy $H(\mathbf{j}|\mathbf{X})$, it suffices to show that with high probability, $\mathbf{X}$ has many good indices. In other words, it will suffice to lower-bound the probability $\Pr[M]$.

Claim 5.11. *For every matrix X in the support of $\mathbf{X}$, the random variable $\mathbf{j}|\mathbf{X} = X$ is uniformly distributed over the good indices of X.*

Proof. Fix a matrix X and an index $j \in [m]$, and let us compute the probability $\Pr[\mathbf{j} = j|\mathbf{X} = X]$. We first observe that if $\mathbf{j}$ is not a good index, then $\Pr[\mathbf{j} = j|\mathbf{X} = X] = 0$, since $\mathbf{X}$ is chosen such that $\mathbf{j}$ is a good index of $\mathbf{X}$. Suppose now that j is a good index. Then,

$$\begin{aligned}
\Pr[\mathbf{j} = j|\mathbf{X} = X] &= \frac{\Pr[\mathbf{X} = X|\mathbf{j} = j] \cdot \Pr[\mathbf{j} = j]}{\Pr[\mathbf{X} = X]} \\
(\mathbf{j} \text{ is uniformly distributed}) &= \frac{\Pr[\mathbf{X} = X|\mathbf{j} = j]}{\frac{1}{m} \cdot \Pr[\mathbf{X} = X]} \\
(\text{see proof of Claim 5.10}) &= \frac{4/\left(2^{n-1}\right)^m}{\frac{1}{m} \cdot \Pr[\mathbf{X} = X]}.
\end{aligned}$$

Now, the last expression is independent of j, and therefore the probability of $\mathbf{j}$ to take the value j is independent of j as long as j is a good index. The required result follows. ■

The following claim is the crux of our argument. It says that for almost all the indices $j \in [m]$, if the index j is good for $\mathbf{X}'$, then many indices are good for $\mathbf{X}'$ with high probability. This is what allows us to show that $\mathbf{X}$ usually has many good indices even though $\mathbf{X}'$ does not.

Claim 5.12. *For at least* $1 - \frac{1}{\log m}$ *fraction of the indices* $j \in [m]$*, it holds that* $\Pr[M'|G_j] \geq 1 - \frac{4}{\log m}$.

Proof. We first note that it suffices to prove that for at least $1 - \frac{1}{\log m}$ fraction of the indices $j \in [m]$, it holds that

$$\Pr\left[G_j|\neg M'\right] \leq \frac{1}{\log m}, \tag{12}$$

since this would imply by the Bayes' rule that

$$\begin{aligned}
\Pr\left[\neg M'|G_j\right] &= \frac{\Pr[G_j|\neg M'] \cdot \Pr[\neg M']}{\Pr[G_j]} \\
&\leq \frac{\Pr[G_j = 1|\neg M']}{\frac{1}{4}} \\
&\leq \frac{4}{\log m}.
\end{aligned}$$

Intuitively, Inequality 12 holds since when there are not many good indices (i.e., when $\neg M'$ occurs), the probability of each particular index to be good is small. Formally, we prove it by Markov's inequality: we first prove an upper bound on the expectation $\mathbb{E}_{\mathbf{j}'}\left[\Pr\left[G_{\mathbf{j}'}|\neg M'\right]\right]$, where $\mathbf{j}'$ is uniformly distributed over $[m]$, and then use Markov's inequality to deduce that $\Pr[G_j|\neg M']$ is small for most j's. In the following equations, we denote by $\mathbf{1}_{G_j}$ the indicator random variable of G_j, and let $\mathbf{j}'$ be a random variable that is uniformly distributed over $[m]$. It holds that

$$\begin{aligned}
\mathbb{E}_{\mathbf{j}'}\left[\Pr_{\mathbf{X}'}\left[G_{\mathbf{j}'}|\neg M'\right]\right] &= \mathbb{E}_{\mathbf{j}'}\left[\mathbb{E}_{\mathbf{X}'}\left[\mathbf{1}_{G_{\mathbf{j}'}}|\neg M'\right]\right] \\
&= \mathbb{E}_{\mathbf{X}'}\left[\mathbb{E}_{\mathbf{j}'}\left[\mathbf{1}_{G_{\mathbf{j}'}}\right]\middle|\neg M'\right] \\
(\text{Definition of } M') &\leq \mathbb{E}_{\mathbf{X}'}\left[\frac{1}{\log^2 m}\middle|\neg M'\right] \\
&= \frac{1}{\log^2 m}.
\end{aligned}$$

Thus, by Markov inequality, the probability for a uniformly distributed $j \in [m]$ it holds that $\Pr[G_j|\neg M'] \geq \frac{1}{\log m}$ is at most $\frac{1}{\log m}$, as required. ■

We now show that $\mathbf{X}$ has many good indices with high probability by combining Claims 5.10 and 5.12.

Claim 5.13. $\Pr[M] \geq 1 - \frac{5}{\log m}$.

Proof. It holds that

$$\Pr[M] = \sum_{j \in [m]} \Pr[M \wedge \mathbf{j} = j] = \frac{1}{m} \cdot \sum_{j \in [m]} \Pr[M | \mathbf{j} = j].$$

Now, by Claim 5.10, it holds that $\mathbf{X}|\mathbf{j} = j$ is distributed like $\mathbf{X}'|G_j$, and therefore

$$\Pr[M] = \frac{1}{m} \sum_{j \in [m]} \Pr\left[M' | G_j\right].$$

Next, by Claim 5.12, for at least $1 - \frac{1}{\log m}$ fraction of the indices j it holds that $\Pr[M'|G_j]$ is at least $1 - \frac{4}{\log m}$, and hence

$$\begin{aligned} \Pr[M] &\geq \left(1 - \frac{1}{\log m}\right) \cdot \left(1 - \frac{4}{\log m}\right) \\ &\geq 1 - \frac{5}{\log m}, \end{aligned}$$

as required. ■

Finally, we combine Claims 5.11 and 5.13 to lower-bound the entropy $H(\mathbf{j}|\mathbf{X})$. Let $\mathbf{1}_M$ be the indicator random variable of the event M. It holds that

$$\begin{aligned} H(\mathbf{j}|\mathbf{X}) &\geq H(\mathbf{j}|\mathbf{X}, \mathbf{1}_M) \\ &\geq \Pr[M] \cdot H(\mathbf{j}|\mathbf{X}, M) \\ \text{(Claim 5.13)} &\geq \left(1 - \frac{5}{\log m}\right) \cdot H(\mathbf{j}|\mathbf{X}, M) \\ \text{(Claim 5.11)} &= \left(1 - \frac{5}{\log m}\right) \cdot \mathbb{E}[\log(\#\text{ good indices for }\mathbf{X})|M] \\ \text{(Definition of } M) &\geq \left(1 - \frac{5}{\log m}\right) \cdot \mathbb{E}\left[\log(\frac{m}{\log^2 m})|M\right] \\ &\geq \log m - 2\log\log m - 5, \end{aligned}$$

as required.

6 Combinatorial Proof of the Main Result

In this section, we provide a self-contained alternative proof of our main result. This proof is essentially the same as the one given in Section 4, but is formulated as a combinatorial double-counting argument, without any reference to information theory. This proof has the advantages of being more direct and of not requiring background in information theory, but on the other hand, it has a more ad-hoc flavor.

Let $g : \{0,1\}^m \to \{0,1\}$. We consider the relation $R_{g \circ \mathrm{U}_n}$, which corresponds to the following communication problem: Alice gets as an input a matrix $X \in \{0,1\}^{m \times n}$ and a string $a \in g^{-1}(0)$. Bob gets a matrix $Y \in \{0,1\}^{m \times n}$ and a vector $b \in g^{-1}(1)$. Their goal is to find an entry (j,i) on which X and Y differ, but they are allowed to reject if there exists an index $j \in [m]$ such that $a_j \neq b_j$ but $X_j = Y_j$. Formally,

Definition 4.1. Let $g : \{0,1\}^m \to \{0,1\}$, and $n \in \mathbb{N}$. The relation $R_{g\circ \mathrm{U}_n}$ is defined by

$$\begin{aligned} R_{g\circ \mathrm{U}_n} \overset{\text{def}}{=} & \left\{((X,a),(Y,b),(j,i)) : X,Y \in \{0,1\}^{m\times n}, a \in g^{-1}(0), b \in g^{-1}(1), X_{j,i} \neq Y_{j,i}\right\} \\ & \cup \left\{((X,a),(Y,b),\perp) : X,Y \in \{0,1\}^{m\times n}, a \in g^{-1}(0), b \in g^{-1}(1), \exists j : a_j \neq b_j, X_j = Y_j\right\}. \end{aligned}$$

Theorem 1.6 (main theorem, restated). *Let $m,n \in \mathbb{N}$, and let $g : \{0,1\}^m \to \{0,1\}$ be a non-constant function. Then,*

$$\mathsf{C}(R_{g\circ U}) \geq \log \mathsf{L}(R_{g\circ \mathrm{U}_n}) \geq \log \mathsf{L}(g) + n - O(1 + \frac{m}{n}) \cdot \log m.$$

In the rest of this section, we prove Theorem 1.6. We note that only the second inequality requires a proof, whereas the first inequality is trivial since a binary tree of depth c has at most 2^c leaves. Let $m,n \in \mathbb{N}$, let $g : \{0,1\}^m \to \{0,1\}$, and let Π be a protocol for $R_{g\circ \mathrm{U}_n}$. We would like to prove that Π has at least $\mathsf{L}(g) \cdot 2^{n-O(1+\frac{m}{n})\cdot \log m.}$ leaves.

The basic idea for the proof, as in Section 4, is the following. We lower-bound the number of leaves that output the rejection symbol $\perp$. For each such leaf ℓ, Alice and Bob must be convinced that there exists some $j \in [m]$ such that $a_j \neq b_j$ but $X_j = Y_j$. In particular:

1. They must be convinced that X and Y agree on at least one row. This is where we gain the factor of 2^n in the number of leaves.
2. They either find an index $j \in [m]$ such that $a_j \neq b_j$, or they do not:
   (a) If they find such a j, they must solve R_g. This gains a factor of $\mathsf{L}(g)$ in the number of leaves.
   (b) If they do not find such a specific index j, they must be convinced that X and Y agree on many rows. However, this forces them to reveal a lot of information about the matrices X and Y, and they cannot afford to do it for most matrices.

We turn to the formal proof. We use the following definition from Section 4.

Definition 4.2. Let ℓ be a leaf of Π and let $\mathcal{X}_\ell \times \mathcal{Y}_\ell$ be its corresponding rectangle.

- We say that the leaf ℓ supports a matrix $X \in \{0,1\}^{m\times n}$ if X can be given as an input to both players at ℓ. Formally, ℓ supports X if there exist $a,b \in \{0,1\}^m$ such that $(X,a) \in \mathcal{X}_\ell$ and $(X,b) \in \mathcal{Y}_\ell$. We also say that X is supported by ℓ and a, or by ℓ and b. Note that the leaf ℓ must be a leaf that outputs $\perp$.
- We say that the leaf ℓ supports $a \in g^{-1}(0)$ if a can be given as input to Alice at ℓ. Formally, ℓ supports a if there exists a matrix $X \in \{0,1\}^{m\times n}$ such that $(X,a) \in \mathcal{X}_\ell$. A similar definition applies to strings $b \in g^{-1}(1)$.

In order to prove a lower bound on $\mathsf{L}(\Pi)$, we double count the number of pairs (ℓ, X), where ℓ is a leaf of Π that outputs $\perp$, and X is a matrix that is supported by L. Specifically, in the next two subsections, we prove the following lemmas, which together imply Theorem 1.6.

Lemma 6.1. *The number of pairs (ℓ, X) is at most $\mathsf{L}(\Pi) \cdot 2^{(m-1)\cdot n}$.*

Lemma 6.2. *The number of pairs (ℓ, X) is at least $2^{mn-O(1+\frac{m}{n})\cdot \log m} \cdot \mathsf{L}(g)$.*

6.1 Proof of Lemma 6.1

We would like to prove that the number of pairs (ℓ, X) is at most $\mathsf{L}(\Pi) \cdot 2^{(m-1)\cdot n}$. To this end, we prove that every leaf can support at most $2^{(m-1)\cdot n}$ matrices. Fix a leaf ℓ, and let $\mathcal{T}$ be the set of matrices supported by ℓ. We prove that $|\mathcal{T}| \leq 2^{(m-1)\cdot n}$.

Intuitively, the reason for this upper bound is that at ℓ, Alice and Bob must be convinced that their matrices agree on at least one row. This intuition is formalized as follows.

Claim 4.5. *Every two matrices X, X' in $\mathcal{T}$ agree on at least one row.*

Proof. We use a standard "fooling set" argument. Let $\mathcal{X}_\ell \times \mathcal{Y}_\ell$ denote the rectangle that corresponds to ℓ. Suppose, for the sake of contradiction, that there exist $X, X' \in \mathcal{T}$ that do not agree on any row. By definition of $\mathcal{T}$, it follows that there exist $a \in g^{-1}(0)$ and $b \in g^{-1}(1)$ such that $(X, a) \in \mathcal{X}_\ell$ and $(X', b) \in \mathcal{Y}_\ell$. In particular, this means that if we give to Alice and Bob the inputs (X, a) and (X', b), respectively, the protocol will reach the leaf ℓ.

However, this is a contradiction: on the one hand, ℓ is a leaf on which the protocol outputs $\perp$. On the other hand, the players are not allowed to output $\perp$ on inputs (X, a), (X', b), since X and X' differ on all their rows, and in particular differ on the all the rows j for which $a_j \neq b_j$. The claim follows. ■

Finally, we observe that Claim 4.5 is just another way of saying that $\mathcal{T}$ satisfies the 1-agreement property (Definition 2.32), when viewed as a set of strings in $\mathbb{F}^m$ over the alphabet $\Sigma = \mathbb{F}^n$. Therefore, Lemma 2.34 implies that $|\mathcal{T}| \leq 2^{(m-1)\cdot n}$, as required.

6.2 Proof of Lemma 6.2

We would like to prove that the number of pairs (ℓ, X) is at least $2^{mn-1} \cdot 2^{-O(\frac{m \log m}{n})} \cdot \mathsf{L}(g)$. We start with the following auxilary definition of the protocol Π_X, which can be thought of as the protocol obtained from Π by fixing the players' matrices to be X.

Definition 4.8. Let $X \in \{0,1\}^{m \times n}$. Let Π_X be the protocol that is obtained from Π as follows: in the protocol tree of Π, we we replace each rectangle $\mathcal{X}_v \times \mathcal{Y}_v$ with the rectangle $\mathcal{X}'_v \times \mathcal{Y}'_v$ defined by

$$\begin{aligned} \mathcal{X}'_v &\stackrel{\text{def}}{=} \{a : (X, a) \in \mathcal{X}_v\} \\ \mathcal{Y}'_v &\stackrel{\text{def}}{=} \{b : (X, b) \in \mathcal{Y}_v\}. \end{aligned}$$

Then, we remove all vertices whose rectangles are empty, and merge all pairs of vertices that have identical rectangles.

In order to prove the lower bound, we partition the matrices X into "good matrices" and "bad matrices". Intuitively, a "good matrix" is a matrix X for which Π_X solves R_g. We will derive the lower bound by showing that that for each good matrix X, there are about $\mathsf{L}(g)$ pairs (ℓ, X), and that there are many good matrices. We define good and bad matrices as follows.

Definition 4.9. Let $t \stackrel{\text{def}}{=} \lceil \frac{6m}{n} \rceil + 2$. A matrix $X \in \{0,1\}^{m \times n}$ is good if Π_X is a protocol that solves the relaxed KW problem $R_g(t)$ (see Definition 2.15). Otherwise, we say that X is bad.

The following lemma says that good matrices have many pairs (ℓ, X), and it is an immediate corollary of Proposition 2.19 (which says that $R_g(t)$ is not much easier than R_g).

Lemma 4.10. *For every good matrix X, the protocol Π_X has at least $2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(g)$ leaves. In other words, there are at least $2^{-t\cdot(\log m+2)} \cdot \mathsf{L}(g)$ pairs (ℓ, X).*

In the next subsection, we will prove the following lemma, which says that there are not many bad matrices, and therefore there are many good matrices.

Lemma 4.11. *The number of bad matrices is at most $2^{-m} \cdot 2^{m\cdot n}$. Thus, the number of good matrices is at least $(1 - 2^{-m}) \cdot 2^{mn} \geq 2^{m\cdot n-1}$.*

Together, Lemmas 4.10 and 4.11 imply Lemma 6.2, as required.

6.2.1 Proof of Lemma 4.11

The intuition for the proof is the following: Recall that Alice and Bob output $\perp$, and this means that they have to be convinced that their matrices agree on some row j for which $a_j \neq b_j$. However, when X is bad, Alice and Bob do not know an index j such that $a_j \neq b_j$ at the end of the protocol. This means that they have to be convinced that they agree on many rows, as otherwise they run the risk of rejecting a legal pair of inputs. But verifying that they agree on many rows is very costly, and they can only do so for few matrices. Details follow.

First, recall that a matrix X is bad if and only if Π_X does not solve the relaxed KW problem $R_g(t)$. This implies that there exists some leaf ℓ' of Π_X, which is labeled with a rectangle $\mathcal{X}'_\ell \times \mathcal{Y}'_\ell$, and a string $a \in \mathcal{X}'_\ell$, such that the following holds:

- For every $\mathcal{J} \subseteq [m]$ such that $|\mathcal{J}| < t$, there exists $b \in \mathcal{Y}'_\ell$ such that $a|_{\mathcal{J}} = b|_{\mathcal{J}}$.

Going back from Π_X to Π, it follows that there exists some leaf ℓ of Π, which is labeled with a rectangle $\mathcal{X}_\ell \times \mathcal{Y}_\ell$, and a string $a \in g^{-1}(0)$, such that the following holds:

- $(X, a) \in \mathcal{X}_\ell$.
- For every $\mathcal{J} \subseteq [m]$ such that $|\mathcal{J}| < t$, there exists $b \in g^{-1}(1)$ such that $a|_{\mathcal{J}} = b|_{\mathcal{J}}$ and $(X, b) \in \mathcal{Y}_\ell$.

Now, without loss of generality, we may assume that

$$\mathsf{L}(\Pi) \leq \mathsf{L}(g) \cdot 2^n \leq 2^{m+n},$$

since otherwise Theorem 1.6 would follow immediately. Therefore, it suffices to prove that every pair of a leaf ℓ and a string a is "responsible" for at most $2^{-(3\cdot m+n)} \cdot 2^{m\cdot n}$ bad matrices. This would imply that there are at most $2^{-m} \cdot 2^{m\cdot n}$ bad matrices, by summing over all leaves of Π (at most 2^{m+n}) and all strings a (at most 2^m).

Fix a leaf ℓ of Π and a string $a \in g^{-1}(0)$. Let $\mathcal{T}$ be the set of bad matrices that are supported by ℓ and a. We prove that $|\mathcal{T}| \leq 2^{-(3\cdot m+n)} \cdot 2^{m\cdot n}$. The key idea is that since Alice does not know a small set $\mathcal{J}$ such that $a|_{\mathcal{J}} \neq b|_{\mathcal{J}}$, Alice and Bob must be convinced that their matrices agree on at least t rows. This intuition is made rigorous in the following statement.

Claim 4.12. *Every two matrices $X, X' \in \mathcal{T}$ agree on at least t rows.*

Proof. Let $X, X' \in \mathcal{T}$, and let $\mathcal{J}$ be the set of rows on which they agree. By definition of $\mathcal{T}$, it holds that $(X, a), (X', a) \in \mathcal{T}$. Suppose that $|\mathcal{J}| < t$. Then, by the assumption on ℓ and a, there exists $b \in g^{-1}(1)$ such that $(X, b) \in \mathcal{Y}_\ell$ and $a|_{\mathcal{J}} = b|_{\mathcal{J}}$.

Next, observe that if we give the input (X', a) to Alice and (X, b) to Bob, the protocol will reach the leaf ℓ. Now, ℓ is a rejecting leaf, and therefore there must exist some index $j \in [m]$ such that $a_j \neq b_j$ but $X_j = X'_j$. However, we know that $a|_{\mathcal{J}} = b|_{\mathcal{J}}$, and therefore $j \notin \mathcal{J}$. It follows that X and Y agree on a row outside $\mathcal{J}$, contradicting the definition of $\mathcal{J}$. ■

Finally, we observe that Claim 4.12 is just another way of saying that $\mathcal{T}$ satisfies the t-agreement property (Definition 2.32), when viewed as a set of strings in $\mathbb{F}^m$ over the alphabet $\mathbb{F} = \{0,1\}^n$. Therefore, Lemma 2.34 implies that $|\mathcal{T}| \leq 2^{(m-t)\cdot n}$. Wrapping up, it follows that

$$\begin{aligned} |\mathcal{T}| &\leq 2^{(m-t)\cdot n} \\ &\leq 2^{(m-\frac{3m}{n}-1)\cdot n} \\ &= \frac{1}{2^{3\cdot m+n}} \cdot 2^{m\cdot n}, \end{aligned}$$

as required.

Remark 6.3. Note that Lemma 2.34 can only be applied if $m \leq 2^n$. However, this can be assumed without loss of generality, since for $m \geq 2^n$, the lower bound of Theorem 1.6 becomes less than $\log \mathsf{L}(g)$. However, it is easy to prove a lower bound of $\log \mathsf{L}(g)$ on $\log \mathsf{L}(R_{g\circ \mathrm{U}_n})$ by reducing R_g to $R_{g\circ \mathrm{U}_n}$.

Acknowledgement. We would like to thank Noga Alon, Gillat Kol, Omer Reingold, Luca Trevisan, and Ryan Williams for helpful discussions.

References