Solving One of Many Instances Is Almost as Hard as Solving One

Nikolai Chukhin * Alexander S. Kulikov † Ivan Mihajlin ‡ Alexander V. Smal §

Abstract

In the 1-out-of-k problem choose_F^k for a function $F\colon A \times B \to \{0,1\}$, Alice gets k distinct inputs $x_1, \ldots, x_k \in A$, Bob gets k distinct inputs $y_1, \ldots, y_k \in B$, and they must output some index $i \in \{1, \ldots, k\}$ together with $F(x_i, y_i)$. Clearly, the players can always compute $F(x_1, y_1)$, so the question is whether the freedom to choose the instance they solve can make the problem substantially easier. This question has been studied in communication complexity and in formula complexity, where it is related to the KRW conjecture of Karchmer, Raz, and Wigderson (1995), which would imply $\mathsf{P} \not\subseteq \mathsf{NC}^1$.

We show that choosing among polynomially many instances saves at most logarithmically many bits, precisely: for every Boolean function F on n-bit inputs and every $k \le 2^n$,

$$\mathrm{CC}(\text{choose}_F^k) \ge \mathrm{CC}(F) - O(\log k + \log n),$$

and the analogous bound holds for formula depth, which proves the conjectures of Dinur and Meir (2016). Previously, only a square-root lower bound was known (Feder, Kushilevitz, Naor, and Nisan, 1995).

For the composition $U_m \diamond \mathrm{KW}_g$ of the universal relation with the KW game of $g\colon \{0,1\}^n \to \{0,1\}$, we prove that every g has a core $S \subseteq \{0,1\}^n$ such that the game restricted to any half of S^m still requires $\mathrm{D}(g) - O(\log n)$ bits. For almost every g, this holds even for $S = \{0,1\}^n$ and any 2^{1-m} fraction of the inputs when $m \le n^2$, which implies

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \ge m + n - O(\log n),$$

improving the bound $m + n - O(\sqrt{m})$ of Wu (2023).

*JetBrains Research. Email: buyolitsez1951@gmail.com
†JetBrains Research. Email: alexander.s.kulikov@gmail.com
‡JetBrains Research. Email: ivmihajlin@gmail.com
§JetBrains Research. Email: avsmal@gmail.com

1 Solving One of Many Instances

Suppose that Alice and Bob must compute a function $F\colon A \times B \to \{0,1\}$, but instead of one input pair they receive k pairs $(x_1, y_1), \ldots, (x_k, y_k)$, with pairwise-distinct inputs on each side, and must output $(i, F(x_i, y_i))$ for some $i \in [k]$. We denote this problem by choose_F^k. Answering the first pair shows that this problem is never harder than computing F, and we ask whether it can be substantially easier. It should be noted that the distinctness condition is essential: if repeated inputs were permitted, then providing each player with k copies of a single input would demonstrate that the problem is at least as hard as F.

Variants of this question go back to selection problems in computability and complexity theory [Joc68, Sel79, Ko83, BKS95] and in communication complexity [ABG+01, CDK+17]. The closest to our problem is the choice problem $\mathrm{choose}_{F_1,\ldots,F_k}$ of Beimel, Ben Daniel, Kushilevitz, and Weinreb [BBKW10, BBKW14], in which the players receive instances of k possibly different functions $F_1, \ldots, F_k$ and must output $(i, F_i(x_i, y_i))$ for some i. Since the functions may differ, repeated inputs do not make this problem trivial, and they are allowed, so choose_F^k is the restriction of $\mathrm{choose}_{F,\ldots,F}$ to distinct inputs. Their motivation was the direct-sum question, which asks whether computing all k values $F_1(x_1, y_1), \ldots, F_k(x_k, y_k)$ requires about $\sum_{j=1}^{k} \mathrm{CC}(F_j)$ bits, that is, about as many as solving the instances separately [FKNN95, KRW95]. Indeed, the players can compute all k values by repeatedly solving the choice problem on the remaining instances and removing the solved one, hence even a weak direct-sum theorem $\mathrm{CC}(F^k) \geq c \cdot k \cdot \mathrm{CC}(F)$ for distinct inputs would imply $\mathrm{CC}(\mathrm{choose}_F^k) \geq (c/2)\,\mathrm{CC}(F)$.

The direct-sum question itself was studied extensively in different models of computation. The direct sum is known to hold for decision trees [JKS10] and for formula size [Weg87], and up to an additive $O(k \log n)$ term for nondeterministic communication [KKN95]. It fails for Boolean circuits, since by a theorem of Uhlig [Uhl74, Weg87], many copies of a function of near-maximal complexity can be computed by a circuit of about the size of one copy, and for tensor rank, Shitov [Shi19] refuted Strassen's direct-sum conjecture. For randomized communication, the exact direct sum fails already for equality with private coins [FKNN95], the distributional direct sum fails [BR14, GKR16], and the general lower bound of Barak, Braverman, Chen, and Rao loses a factor of about $\sqrt{k}$ [BBCR13]. For deterministic communication, the strong form of direct sum fails [MS25], and whether a weaker form $\mathrm{CC}(F^k) = \Omega(k \cdot \mathrm{CC}(F))$ holds remains open. For the universal relation, in which the players must find a coordinate where two different strings differ, the direct sum holds almost exactly: a rank argument of Kozachinsky shows that k instances on n bits require $k(n-1) - 1$ bits [Mei18, Remark 6].

One motivation for both the direct-sum and the 1-out-of-k questions comes from formula lower bounds. Proving that some function in P requires De Morgan formulas of depth $\omega(\log n)$, or equivalently of superpolynomial size, would show that $\mathsf{P} \not\subseteq \mathsf{NC}^1$, yet the best known size lower bounds for explicit functions are only about n^3 [Hås98, Tal14]. Karchmer, Raz, and Wigderson [KRW95] proposed to approach this separation through composition of Boolean functions. For $f\colon \{0,1\}^m \to \{0,1\}$ and $g\colon \{0,1\}^n \to \{0,1\}$, the function $f \diamond g\colon \{0,1\}^{m \times n} \to \{0,1\}$ applies g to each row of its input matrix and then f to the m resulting bits, and the Karchmer–Raz–Wigderson (KRW) conjecture asserts that the formula depth of $f \diamond g$ is about the sum of the formula depths of f and g. Iterating such compositions would give a function in P that requires depth $\omega(\log n)$, and hence $\mathsf{P} \not\subseteq \mathsf{NC}^1$. Karchmer and Wigderson [KW90] showed that the formula depth of f equals the communication complexity of the game KW_f, in which Alice receives $x \in f^{-1}(0)$, Bob receives $y \in f^{-1}(1)$, and they must find a coordinate where x and y differ. In these terms, the KRW conjecture says that the natural protocol for $\mathrm{KW}_{f \diamond g}$ is essentially optimal: the players first find a row i with $g(X_i) \neq g(Y_i)$ by solving KW_f on the vectors of row values, and then solve KW_g on this row. Karchmer, Raz,

and Wigderson related this conjecture to the direct-sum question: the m rows are m instances of KW_g, and the conjecture says that working on all of them at once does not help. The difficulty is that a protocol need not commit to a row in advance: it may solve KW_g on whichever row with $g(X_i) \neq g(Y_i)$ turns out to be convenient. The 1-out-of-k problem captures this freedom: if Alice gets k distinct inputs in $g^{-1}(0)$ and Bob gets k distinct inputs in $g^{-1}(1)$, then every pair is a valid instance of KW_g and the players may answer any of them, so a composition lower bound of this kind requires that choosing does not make KW_g easier.

Dinur and Meir [DM16] formulated the problem with pairwise-distinct inputs in their program toward the KRW conjecture. They conjectured that choosing an instance saves at most polylogarithmically many bits, both in communication complexity and in formula depth [DM16, Conjectures 9.6 and 9.7]. Yet for general F the best known bound was only $\mathrm{CC}(F) \leq O((\mathrm{CC}(\mathrm{choose}_F^k) + \log n)^2)$, which follows from bounds of Feder, Kushilevitz, Naor, and Nisan [FKNN95], as observed by Dinur and Meir [DM16], and stronger bounds were known only in special cases, such as random functions. For a random function $f\colon \{0,1\}^n \to \{0,1\}$ and up to $\varepsilon 2^n$ distinct instances, Meir and Tal [MT18] proved that choice and agreement are almost as hard as f for circuits, formulas, and decision trees. In randomized query complexity, Bassilakis, Drucker, Göös, Hu, Ma, and Tan [BDG+20] proved that, for every function, analogues of agreement and choice on independent samples from a suitable distribution require a constant fraction of its complexity. We prove that $\mathrm{CC}(\mathrm{choose}_F^k) \geq \mathrm{CC}(F) - O(\log k + \log n)$, hence resolving the conjectures of Dinur and Meir for all k up to $2^{\mathrm{polylog}(n)}$.

1.1 Choosing Is Almost as Hard as Computing

Deterministic communication complexity. For a total Boolean function $F\colon A \times B \to \{0,1\}$, the *1-out-of-k problem* choose_F^k gives Alice pairwise-distinct inputs $x_1, \ldots, x_k \in A$ and Bob pairwise-distinct inputs $y_1, \ldots, y_k \in B$. They must output an index $i \in [k]$ and the value $F(x_i, y_i)$.

Theorem 1.1. *For every total $F\colon A \times B \to \{0,1\}$ with $|A|, |B| \leq 2^n$, and every $1 \leq k \leq \min\{|A|, |B|\}$,*

$$\mathrm{CC}(F) \leq \mathrm{CC}(\mathrm{choose}_F^k) + O(\log k + \log n).$$

Previously, the best bound for general F was $\mathrm{CC}(F) \leq O((\mathrm{CC}(\mathrm{choose}_F^k) + \log n)^2)$ [FKNN95]. Theorem 1.1 replaces this quadratic bound by an additive loss of $O(\log n)$ for polynomial k and of $\mathrm{polylog}(n)$ for all $k \leq 2^{\mathrm{polylog}(n)}$, resolving the conjecture of [DM16].

We obtain Theorem 1.1 from a more general theorem for total *search* problems. Let $\mathcal{R} \subseteq A \times B \times O$ be a total relation, so every pair $(x, y) \in A \times B$ has at least one valid output $o \in O$. In $\mathrm{choose}_{\mathcal{R}}^k$, each player receives k pairwise-distinct inputs and the players must find an index $i \in [k]$ and an output $o \in O$ with $(x_i, y_i, o) \in \mathcal{R}$.

Theorem 1.2. *Let $\mathcal{R} \subseteq A \times B \times O$ be a total relation, with $1 \leq |A|, |B| \leq 2^n$. For every $1 \leq k \leq \min\{|A|, |B|\}$,*

$$\mathrm{CC}(\mathcal{R}) \leq \max\left\{\mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^k), \lceil \log_2 |O| \rceil\right\} + O(\log k + \log n).$$

Formula depth. For a partial function $f\colon \{0,1\}^n \to \{0,1,*\}$, let $\mathrm{D}(f)$ and $\mathrm{L}(f)$ denote the minimum depth and leaf count of a fan-in-two De Morgan formula computing f. As recalled above, $\mathrm{D}(f) = \mathrm{CC}(\mathrm{KW}_f)$ [KW90].

The agreement function agree_f^k takes k distinct inputs promised to lie all in $f^{-1}(0)$ or all in $f^{-1}(1)$ and returns their common value under f. It is the analogue of choice for formulas: the

KW game of agree_f^k is $\mathrm{choose}^k_{\mathrm{KW}_f}$, so $\mathrm{D}(\mathrm{agree}_f^k) = \mathrm{CC}(\mathrm{choose}^k_{\mathrm{KW}_f})$. Applying a formula for f to the first input shows that $\mathrm{D}(\mathrm{agree}_f^k) \le \mathrm{D}(f)$. Agreement has also been studied before in [BBKW14, DM16, MT18, Wu23], but to our knowledge no nontrivial lower bound on $\mathrm{D}(\mathrm{agree}_f^k)$ was known for an arbitrary f, even for $k = 2$. Since $\mathrm{D}(\mathrm{agree}_f^k) = \mathrm{CC}(\mathrm{choose}^k_{\mathrm{KW}_f})$, Theorem 1.2 with $O = [n]$ immediately gives the following bound.

Corollary 1.3. *For every $k \ge 1$ and every partial $f\colon \{0,1\}^n \to \{0,1,*\}$ with $|f^{-1}(0)|, |f^{-1}(1)| \ge k$,*

$$\mathrm{D}(f) \le \mathrm{D}(\mathrm{agree}_f^k) + O(\log k + \log n).$$

This proves the formula-depth conjecture of Dinur and Meir [DM16] for any $k \le 2^{\mathrm{polylog}(n)}$, even for partial functions.

Exponentially many instances. For exponentially large k, the additive $O(\log k)$ loss in Theorem 1.1 and Corollary 1.3 can exceed the complexity of a single instance. Nevertheless, Theorem 1.2 implies a linear lower bound for every k.

Corollary 1.4. *For F, k as in Theorem 1.1 and f, k as in Corollary 1.3,*

$$\mathrm{CC}(\mathrm{choose}_F^k) \ge \Omega(\mathrm{CC}(F)) - O(\log n),$$
$$\mathrm{D}(\mathrm{agree}_f^k) \ge \Omega(\mathrm{D}(f)) - O(\log n).$$

This follows because a depth-d protocol has at most 2^d leaves and therefore can have at most 2^d coordinates of its input tuples as outputs, so Theorem 1.2 applies with 2^d in place of k. Once $\mathrm{CC}(F)$ exceeds a sufficiently large multiple of $\log n$, this improves the previous $\Omega(\sqrt{\mathrm{CC}(F)}) - O(\log n)$ bound to a linear bound for every k.

Two instances. For $k = 2$, Corollary 1.3 already gives an $O(\log n)$ depth loss. We also give a short alternative proof for this case by adapting a covering argument of Beimel, Ben Daniel, Kushilevitz, and Weinreb [BBKW10] to formulas; it gives an explicit additive loss in depth and also bounds the formula size.

Theorem 1.5. *For every partial $f\colon \{0,1\}^n \to \{0,1,*\}$ with $|f^{-1}(0)|, |f^{-1}(1)| \ge 2$,*

$$\mathrm{D}(f) \le \mathrm{D}(\mathrm{agree}_f^2) + \lceil \log_2 n \rceil + 1,$$
$$\mathrm{L}(f) \le 2n \cdot \mathrm{L}(\mathrm{agree}_f^2).$$

Note that these losses are necessary up to constants. For $n \ge 3$, the function $f_n(x) = \neg x_1 \wedge (x_2 \vee \cdots \vee x_n)$ depends on all n variables, so it needs at least n leaves and depth $\lceil \log_2 n \rceil$. Yet $\mathrm{agree}^2_{f_n}$ is computed by $\neg x_1^{(1)} \wedge \neg x_1^{(2)}$, with two leaves and depth one. Thus, for two instances, the formula depth can decrease by at most $\log n + O(1)$ and the formula size by at most a factor of $O(n)$, and both bounds are attained up to constants.

1.2 Universal Composition and Dense Restrictions

The freedom to choose a row is only one way in which a protocol for $\mathrm{KW}_{f \diamond g}$ could beat the natural protocol. More generally, every bit of communication tells the players something about each other's inputs: after t bits, following the larger half at each step, they reach a node whose rectangle contains at least a 2^{-t} fraction of the inputs, and the rest of the protocol must work only on this

rectangle. To show that the players still have to solve KW_g there, one needs a communication problem to remain hard when its inputs are restricted to a dense subset of the domain. This dense-restriction question generalizes the 1-out-of-k problem, which restricts the rows of Alice's matrix to $g^{-1}(0)$ and the rows of Bob's matrix to $g^{-1}(1)$.

A standard approach toward the KRW conjecture replaces KW_f or KW_g by the universal relation U_m (or U_n), in which the players receive two strings that are promised to be different and must find a coordinate where they differ. Analogues of the KRW conjecture are known for $U_m \diamond U_n$ [HW93, EIRS01] and for $\mathrm{KW}_f \diamond U_n$ [GMWW17, KM18], as well as when the inner function is parity [DM16], for classes of monotone functions [dRMN+24], and for stronger composition relations [Mei25, CKM25]. Here we study the composition $U_m \diamond \mathrm{KW}_g$ of the outer universal relation U_m with the KW game of a Boolean inner function g, also considered by Gavinsky, Meir, Weinstein, and Wigderson [GMWW17].

In $U_m \diamond \mathrm{KW}_g$, Alice and Bob receive matrices $X, Y \in \{0,1\}^{m\times n}$ and must output a pair (i,j) with $X_{i,j} \neq Y_{i,j}$, or reject when their vectors of row values $g^m(X) = (g(X_1), \ldots, g(X_m))$ and $g^m(Y)$ are equal. Note that Alice can send $g^m(X)$ using m bits, after which the players can either reject or solve KW_g on a row where $g^m(X)$ and $g^m(Y)$ differ, using $m + \mathrm{D}(g) + O(1)$ bits in total. The goal is to prove a matching lower bound of roughly $m + \mathrm{D}(g)$.

Mihajlin and Smal [MS21] showed that there is an inner function $g\colon \{0,1\}^n \to \{0,1\}$ with $\mathrm{CC}(U_n \diamond \mathrm{KW}_g) \geq 1.5n - o(n)$. Wu [Wu23] improved it by showing that for almost every $g\colon \{0,1\}^n \to \{0,1\}$, when $\log^2 n = o(m)$ and $m = o(n^2)$,

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \geq m + n - O(\sqrt{m}).$$

He asked whether the loss could be made polylogarithmic. We reduce it to $O(\log n)$, for all $m \leq n^2$.

Theorem 1.6. *With probability $1 - o(1)$ as $n \to \infty$, a uniformly random $g\colon \{0,1\}^n \to \{0,1\}$ satisfies, for all $m \leq n^2$,*

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \geq m + n - O(\log n) \geq m + \mathrm{D}(g) - O(\log n).$$

Unlike [MS21, Wu23], who reduce the problem that remains after about m bits of communication to a nondeterministic lower bound for non-equality, we follow the dense-restriction approach of Dinur and Meir [DM16]. For a subset $\mathcal{X} \subseteq \{0,1\}^{m\times n}$, we write $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ for the restriction of $U_m \diamond \mathrm{KW}_g$ to $\mathcal{X} \times \mathcal{X}$. Following a protocol for $U_m \diamond \mathrm{KW}_g$ on diagonal inputs (X, X) for $m - 1$ steps we can reach a node whose rectangle contains $\mathcal{X} \times \mathcal{X}$ for a subset $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of density at least 2^{1-m}, where the remaining subprotocol must solve $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$. To prove Theorem 1.6, it therefore suffices to show that $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ remains hard on every such subset $\mathcal{X}$, which we do for almost every g in Theorem 6.6.

Arbitrary inner functions. Theorem 1.6 holds only for almost every g, while the KRW conjecture concerns every inner function. We therefore also study the dense-restriction approach for an arbitrary g: we show what it gives at high densities and that it cannot work uniformly at the low densities that the composition requires. An arbitrary function g need not remain hard on dense subsets of the full domain $\{0,1\}^{m\times n}$, because g may be hard only on a small part of $\{0,1\}^n$; for this reason, the question is whether every g has a hard *core* $\mathcal{X}_0 \subseteq \{0,1\}^{m\times n}$ such that $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ remains hard for every sufficiently dense subset $\mathcal{X} \subseteq \mathcal{X}_0$ [DM16, Conjecture 9.4]. As a first step, they asked whether $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq \mathrm{D}(g) - 0.99m$ for every subset $\mathcal{X} \subseteq \mathcal{X}_0$ of relative density $1/2$ [DM16, p. 3:48].

We prove that every nonconstant function g has such a core with only an $O(\log n)$ loss.

Theorem 1.7. *For every nonconstant $g\colon \{0,1\}^n \to \{0,1\}$, there is a nonempty $S \subseteq \{0,1\}^n$ such that, for every $m \geq 2$ and every $\mathcal{X} \subseteq S^m$ of relative density at least $(1+2^{1-m})/3$,*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq \mathrm{D}(g) - O(\log n).$$

We note that the $O(\log n)$ loss is tight up to a constant factor (see Section 6.5). In particular, the answer to the question of Dinur and Meir depends on m: their bound holds for every g when m is at least a large constant times $\log n$, but for $m \leq (\log_2 n)/2$ and all sufficiently large n, there is a function g for which it fails in every core.

Moreover, Dinur and Meir [DM16, Conjecture 9.4] conjectured that the approach also works at the low densities that the composition requires: every g has a core $\mathcal{X}_0$ such that $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ loses at most $\widetilde{O}(\sqrt{m})$ bits for every $\mathcal{X} \subseteq \mathcal{X}_0$ of relative density $2^{-m+\widetilde{O}(\sqrt{m})}$, which, by the argument above, would give $\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \geq m + \mathrm{D}(g) - \widetilde{O}(\sqrt{m})$ for every g. We show that this fails when both $\widetilde{O}(\sqrt{m})$ terms depend only on m: no such bound holds uniformly over n, even for an arbitrary core $\mathcal{X}_0 \subseteq \{0,1\}^{m\times n}$.

Theorem 1.8. *For every sufficiently large even m, there is a balanced Boolean function g_m on $n = 2^{\Theta(m)}$ bits such that, for every nonempty core $\mathcal{X}_0 \subseteq \{0,1\}^{m\times n}$ and every integer $0 \leq \sigma \leq m-1$, some $\mathcal{X} \subseteq \mathcal{X}_0$ satisfies $|\mathcal{X}| \geq 2^{-m+\sigma}|\mathcal{X}_0|$, but*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_{g_m})|_{\mathcal{X}\times\mathcal{X}}) \leq \mathrm{D}(g_m) - m/2 + \sigma + O(\log m).$$

Thus no dense-core bound for subsets of relative density $2^{-m+\sigma}$ with $\sigma = o(m)$ and with depth loss $o(m)$, where both $o(m)$ terms depend only on m, can hold for all n and g. This refutes the uniform formulation of [DM16, Conjecture 9.4]. In our construction, $n = 2^{\Theta(m)}$, or equivalently $m = \Theta(\log n)$, so allowing an additional $O(\log n)$ depth loss or restricting to $n = m^{O(1)}$ remains open (see Section 7).

2 Proof Ideas

In this section, we describe the main ideas of our proofs.

2.1 1-out-of-2 Agreement

For $k = 2$, we adapt the covering argument of Beimel, Ben Daniel, Kushilevitz, and Weinreb [BBKW10]. Let $H(u,v)$ be a formula for agree^2_f, and consider its values on the pairs in $f^{-1}(0) \times f^{-1}(1)$. A greedy halving argument on this matrix, finds at most n inputs $v^{(1)}, \ldots, v^{(t)} \in f^{-1}(1)$ such that every $x \in f^{-1}(0)$ has $H(x, v^{(j)}) = 0$ for some j, or a symmetric family in $f^{-1}(0)$. In the first case,

$$f(x) = \bigwedge_{j=1}^{t} \Big(H(x, v^{(j)}) \vee \mathrm{EQ}_{v^{(j)}}(x) \Big)$$

on the promise: for $x \in f^{-1}(0)$, some term is zero, and for $x \in f^{-1}(1)$, each term is one by the correctness of H, except at $x = v^{(j)}$, where the two inputs of H coincide and the equality test corrects the term. This formula has depth $\mathrm{D}(\mathrm{agree}^2_f) + O(\log n)$. This proves Theorem 1.5, see Section 4 for more details.

2.2 The 1-out-of-k Problem

A natural first attempt is to reduce the problem to two instances. Let k be a power of two, and split the k instances into two halves. Each half is an instance of $\mathrm{choose}_{\mathcal{R}}^{k/2}$, and a protocol for $\mathrm{choose}_{\mathcal{R}}^{k}$ solves one of these two instances, so it is a protocol for $\mathrm{choose}_{\mathcal{S}}^{2}$ with $\mathcal{S} = \mathrm{choose}_{\mathcal{R}}^{k/2}$, restricted to halves whose entries are all distinct. Suppose that every relation $\mathcal{S}$ on N-bit inputs satisfied $\mathrm{CC}(\mathcal{S}) \leq \mathrm{CC}(\mathrm{choose}_{\mathcal{S}}^{2}) + O(\log N)$, even on this restriction. Applying this bound to $\mathrm{choose}_{\mathcal{R}}^{2^i}$, whose inputs have $2^i n$ bits, for $i = \log_2 k - 1, \ldots, 0$ would give

$$\mathrm{CC}(\mathcal{R}) \leq \mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^{k}) + \sum_{i=0}^{\log_2 k - 1} O(i + \log n) = \mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^{k}) + O(\log k \cdot (\log k + \log n)).$$

This loses a factor of $\log k$ compared with Theorem 1.2; for instance, for $k = \mathrm{poly}(n)$, it gives a loss of $O(\log^2 n)$ instead of $O(\log n)$. We therefore handle all k instances at once.

Let Π be an optimal protocol for $\mathrm{choose}_{\mathcal{R}}^{k}$, and let $x \in A$ and $y \in B$ be inputs for $\mathcal{R}$. The natural way to use Π is to put x and y at one position of two tuples whose other entries are fixed, and to run Π on these tuples. This fails because Π may answer at one of the fixed positions, and then its output says nothing about (x, y).

Our approach replaces a single run of Π with a game on its tree. Every leaf w of Π names a position i_w and an output, and this output is valid for every pair of inputs that can stand at position i_w in tuples reaching w. On single inputs x and y, the players walk down the tree of Π and choose the children at their own nodes freely: Alice steers the walk to leaves at which x can stand at the named position, and Bob steers it to leaves at which y can. If both succeed, the leaf they reach gives a valid output for (x, y). This raises two questions: for which pairs (x, y) do such strategies exist, and how do the players find out that their pair is one of them?

Call a rectangle $A' \times B'$ solvable if some protocol of depth $\max\{\mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^{k}), \lceil \log_2 |O| \rceil\}$ solves $\mathcal{R}$ on it. We show that for all distributions μ on A and ν on B, some solvable rectangle satisfies

$$\mu(A') + \nu(B') \geq 1 + k^{-2}.$$

This inequality says that $A' \times B'$ is slightly larger than the trivial solvable rectangles $A \times \varnothing$ and $\varnothing \times B$, which give exactly one. If μ or ν gives some input probability at least k^{-2}, such a rectangle is easy to find: for instance, if $\mu(x_0) \geq k^{-2}$ for some $x_0 \in A$, then $\{x_0\} \times B$ is solvable, since Bob can send a valid output for (x_0, y), and $\mu(\{x_0\}) + \nu(B) \geq 1 + k^{-2}$. Otherwise, random tuples $X \sim \mu^k$ and $Y \sim \nu^k$ have distinct entries with probability close to one. Running Π on X, Alice can steer the walk to leaves at which her own entry stands at the named position, and on Y, Bob can do the same. We judge the leaves at positions up to j by Alice's tuple and the other leaves by Bob's tuple, and move j from 0 to k. By Zermelo's theorem [Zer13], in every such game one of the players can force the outcome, so at some position i the probability that Alice wins increases by at least k^{-2}. Fixing all entries except the ith ones gives the rectangle (Lemma 5.2).

The players find a solvable rectangle that contains (x, y) with $O(\log k + \log n)$ bits of communication. By the minimax theorem, there is a random solvable rectangle that contains each input of either player with probability at least $1/2 + k^{-2}/4$; probability $1/2$ is trivial, by choosing $A \times \varnothing$ or $\varnothing \times B$. By minimax, it suffices that for every distribution on $A \sqcup B$, some solvable rectangle contains a random input with this probability; if the distribution puts noticeably more weight on one player, a trivial rectangle does, and otherwise the inequality above does. Majority amplification turns this random rectangle into a formula of depth $O(\log k + \log n)$ whose leaves are solvable rectangles. Then, the players walk down this formula to a rectangle that contains (x, y), and then solve $\mathcal{R}$ on it.

2.3 Dense Restrictions and Composition

Recall that in $U_m \diamond \mathrm{KW}_g$ the players receive matrices $X, Y \in \{0,1\}^{m\times n}$ and must find a cell where they differ or reject when $g^m(X) = g^m(Y)$, and that $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ is the restriction of this problem to $\mathcal{X}\times\mathcal{X}$, where $\mathcal{X} \subseteq \{0,1\}^{m\times n}$. We first show that lower bounds on $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}})$ for all dense sets $\mathcal{X}$ give lower bounds on $\mathrm{CC}(U_m \diamond \mathrm{KW}_g)$. Take a protocol for $U_m \diamond \mathrm{KW}_g$ and walk down its tree, at each step moving to a child that keeps at least half of the remaining diagonal inputs (X, X), on which both players have the same matrix $X \in \{0,1\}^{m\times n}$. Let $\mathcal{X}$ be the set of all matrices X such that (X, X) reaches the current node after $m-1$ steps; its density is at least 2^{1-m}. Hence a lower bound on $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}})$ for all sets $\mathcal{X}$ of density at least 2^{1-m} adds to the $m-1$ bits already sent (see Lemma 6.7).

To prove such lower bounds, we turn a protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ into a formula. Consider a leaf which rejects, i.e. all the inputs reaching such a leaf form a rectangle, and every pair (X, Y) in it satisfies $g^m(X) = g^m(Y)$, so all these matrices have the same value $c \in \{0,1\}^m$ of g^m, and we label the leaf with a new variable p_c. This gives a formula $\phi(X, p)$ in the bits of a matrix X and the variables $p = (p_c)_{c\in\{0,1\}^m}$ whose depth is at most that of the protocol, and we claim that $\phi(X, p) = p_{g^m(X)}$ for every $X \in \mathcal{X}$ and every p (Lemma 6.1). Indeed, if $\phi(X, p) = 1$, then we can walk from the root to a leaf through nodes of value one, following Alice's moves on X at her nodes and choosing a child of value one at Bob's nodes. This leaf can be reached with Alice's input X, so it is not a literal, since the literal at a leaf is zero on all Alice's inputs that reach this leaf; hence it is $p_{g^m(X)}$, and $p_{g^m(X)} = 1$. The case $\phi(X, p) = 0$ is symmetric, with Bob in place of Alice. Thus, ϕ determines $g^m(X)$, that is, the values of g on all rows of X, for every $X \in \mathcal{X}$.

For an arbitrary function $g\colon \{0,1\}^n \to \{0,1\}$ and a nonempty set $S \subseteq \{0,1\}^n$, if $\mathcal{X} \subseteq S^m$ has relative density at least $(1 + 2^{1-m})/3$, then substituting suitable values for p and a hybrid argument over the rows turn ϕ into a formula of depth $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) + 1$ that computes g on at least two thirds of S (Lemma 6.2). It remains to find a set S on which no formula of depth $\mathrm{D}(g) - O(\log n)$ computes g on two thirds of the inputs (Lemma 6.3). The construction of such a set resembles Nisan's proof of Impagliazzo's hard-core lemma [Imp95].

For a random g, a fixed formula ϕ defines a map from matrices to $\{0,1\}^m \cup \{\perp\}$ that does not depend on g and that equals g^m on $\mathcal{X}$, that is, on at least a 2^{1-m} fraction of all matrices. For a fixed map, this happens with very small probability (Lemma 6.4), so we take a union bound over all formulas of depth $n - O(\log n)$.

Finally, to prove Theorem 1.8, we construct a balanced function g that is hard for formulas but easy for random formulas in the following sense: g has depth about $3r$, whereas a random formula of depth about $2r$ computes $g(x)$ correctly at every input x with probability close to one (Lemma 6.9). In every core $\mathcal{X}_0$, some formula ψ from this distribution is correct on all m rows of at least half of the matrices. Keeping only the matrices with the most frequent values of g^m, we obtain a large subset $\mathcal{X}$ on which Alice sends $g^m(X)$ with few bits and the players then solve KW_ψ on a row where the values differ (Lemma 6.10).

3 Preliminaries

A relation $\mathcal{R} \subseteq A \times B \times O$ is *total* if every pair $(x, y) \in A \times B$ has an output $o \in O$ with $(x, y, o) \in \mathcal{R}$. A partial function $f\colon \{0,1\}^n \to \{0, 1, *\}$ is *nonconstant* if both $f^{-1}(0)$ and $f^{-1}(1)$ are nonempty.

3.1 Formulas and KW Games

A De Morgan formula is a binary tree whose inner nodes are AND and OR gates and whose leaves are literals x_j and $\neg x_j$ or constants. Its depth is the maximum number of gates on a path from the root to a leaf. A formula computes a partial function $f\colon \{0,1\}^n \to \{0,1,*\}$ if it is correct on the promise $f^{-1}(0) \cup f^{-1}(1)$, and we write $\mathrm{D}(f)$ and $\mathrm{L}(f)$ for the minimum depth and the minimum number of leaves of such a formula.

For fixed $z \in \{0,1\}^n$, the equality test

$$\mathrm{EQ}_z(x) = \bigwedge_{j=1}^{n} [x_j = z_j]$$

has depth $\lceil \log_2 n \rceil$ and exactly n leaves.

Let $\mathcal{R} \subseteq A \times B \times O$ be a total relation. A deterministic protocol for $\mathcal{R}$ is a binary tree in which every inner node belongs to Alice or to Bob and every leaf is labeled by an output $o \in O$. Each node of Alice is labeled by a function $A \to \{0,1\}$, and each node of Bob by a function $B \to \{0,1\}$, which tells which child to go to. On inputs $x \in A$ and $y \in B$, the players start at the root and repeatedly go to the child chosen by the owner of the current node on its own input, until they reach a leaf, whose label is the output. The protocol solves $\mathcal{R}$ if its output o satisfies $(x,y,o) \in \mathcal{R}$ for all $x \in A$ and $y \in B$. The depth of a protocol is the maximum length of a path from the root to a leaf, that is, the number of bits sent in the worst case. We write $\mathrm{CC}(\mathcal{R})$ for the minimum depth of a protocol that solves $\mathcal{R}$.

Karchmer and Wigderson [KW90] showed that formula depth can be expressed as the communication complexity of a game associated with the function.

Definition 3.1. Let $f\colon \{0,1\}^n \to \{0,1,*\}$ be a nonconstant partial function. The Karchmer–Wigderson game of f is the following relation:

$$\mathrm{KW}_f = \{(x,y,j) \in f^{-1}(0) \times f^{-1}(1) \times [n] : x_j \neq y_j\}.$$

Thus Alice gets $x \in f^{-1}(0)$, Bob gets $y \in f^{-1}(1)$, and they must output a coordinate j with $x_j \neq y_j$.

Theorem 3.2 ([KW90])**.** *For every nonconstant partial function f,*

$$\mathrm{D}(f) = \mathrm{CC}(\mathrm{KW}_f).$$

3.2 Agreement and 1-out-of-k Problems

We first define the agreement function, which asks for the common value of a partial function on k distinct inputs. For a set A and $k \geq 1$, let $\mathrm{Inj}_k(A)$ be the set of k-tuples of pairwise distinct elements of A:

$$\mathrm{Inj}_k(A) = \{(a_1, \ldots, a_k) \in A^k : a_i \neq a_j \text{ for } i \neq j\}.$$

Definition 3.3. Let $f\colon \{0,1\}^n \to \{0,1,*\}$ and let $k \geq 1$. If $|f^{-1}(0)|, |f^{-1}(1)| \geq k$, the agreement function agree_f^k is the partial function on kn bits defined by

$$(\mathrm{agree}_f^k)^{-1}(b) = \mathrm{Inj}_k(f^{-1}(b)) \quad \text{for } b \in \{0,1\},$$

and $*$ elsewhere.

That is, the input consists of k distinct strings $x^{(1)}, \ldots, x^{(k)} \in \{0,1\}^n$ with the promise that $f(x^{(1)}) = \cdots = f(x^{(k)}) \in \{0,1\}$, and the output is this common value.

Now, we define the 1-out-of-k problem for every total relation.

Definition 3.4. Let $\mathcal{R} \subseteq A \times B \times O$ be a total relation. For $1 \le k \le \min\{|A|, |B|\}$, the 1-out-of-$k$ problem $\text{choose}^k_{\mathcal{R}}$ gives Alice $X = (x_1, \dots, x_k) \in \text{Inj}_k(A)$ and Bob $Y = (y_1, \dots, y_k) \in \text{Inj}_k(B)$ and asks for a pair $(i, o) \in [k] \times O$ such that $(x_i, y_i, o) \in \mathcal{R}$.

We identify a function $F\colon A \times B \to \{0,1\}$ with the total relation $\{(x, y, F(x,y)) : x \in A,\, y \in B\}$.

Remark 3.5. If $|f^{-1}(0)|, |f^{-1}(1)| \ge k$, the KW game of agree^k_f is $\text{choose}^k_{\text{KW}_f}$, where an output (i, j) names the jth bit of the ith input. Indeed, Alice gets distinct $x_1, \dots, x_k \in f^{-1}(0)$, Bob gets distinct $y_1, \dots, y_k \in f^{-1}(1)$, and the jth bits of x_i and y_i differ exactly when $(x_i, y_i, j) \in \text{KW}_f$. Hence, by Theorem 3.2,

$$\text{D}(\text{agree}^k_f) = \text{CC}(\text{choose}^k_{\text{KW}_f}).$$

3.3 Minimax and Majority Amplification

We use von Neumann's minimax theorem [vN28] for zero-sum games.

Lemma 3.6 ([vN28])**.** *Let $\mathcal{H}$ and $\mathcal{Z}$ be finite nonempty sets, let $C \subseteq \mathcal{H} \times \mathcal{Z}$, and let $\delta \in \mathbb{R}$. If for every distribution π on $\mathcal{Z}$ some $H \in \mathcal{H}$ satisfies $\Pr_{z\sim\pi}[(H, z) \in C] \ge \delta$, then some distribution Λ on $\mathcal{H}$ satisfies $\Pr_{H\sim\Lambda}[(H, z) \in C] \ge \delta$ for every $z \in \mathcal{Z}$.*

The lemma below turns a distribution over formulas that is correct with probability $1/2 + \varepsilon$ at every point into a single formula that is correct at every point.

Lemma 3.7 ([Val84])**.** *Let $\mathcal{Z}$ be a finite set, let $f\colon \mathcal{Z} \to \{0,1\}$, let $\mathcal{F}$ be a set of functions $\mathcal{Z} \to \{0,1\}$, and let $0 < \varepsilon \le 1/2$. If a distribution Λ over formulas of depth at most d with AND and OR gates and leaves from $\mathcal{F}$ satisfies $\Pr_{H\sim\Lambda}[H(z) = f(z)] \ge 1/2 + \varepsilon$ for every $z \in \mathcal{Z}$, then some formula H^* of depth $d + O(\log(1/\varepsilon) + \log\log|\mathcal{Z}|)$ with leaves from $\mathcal{F}$ satisfies $H^*(z) = f(z)$ for every $z \in \mathcal{Z}$.*

4 A Simple Proof for 1-out-of-2 Agreement

In this section, we give a simple proof of Theorem 1.5, which takes its main idea from the proof of [BBKW10, Theorem 4].

Theorem 1.5. *For every partial $f\colon \{0,1\}^n \to \{0,1,*\}$ with $|f^{-1}(0)|, |f^{-1}(1)| \ge 2$,*

$$\text{D}(f) \le \text{D}(\text{agree}^2_f) + \lceil \log_2 n \rceil + 1,$$
$$\text{L}(f) \le 2n \cdot \text{L}(\text{agree}^2_f).$$

Proof. Let H be any formula for agree^2_f with depth d and ℓ leaves. We prove that

$$\text{D}(f) \le d + \lceil \log_2 n \rceil + 1, \qquad \text{L}(f) \le \min\{n, \ell\}(\ell + n).$$

Write the input to H as two binary strings (u, v), where $u, v \in \{0,1\}^n$. The variable in each leaf of H belongs either to the first string u or to the second string v. Let n_u and n_v be the numbers of distinct variables of u and of v, respectively, that appear in the leaves, so $n_u + n_v \le \ell$. If $n_u = 0$ or $n_v = 0$, then H depends on only one string, and it already computes f on that string. So, from now on, we assume that $n_u, n_v \ge 1$.

Consider the Boolean matrix on $F_0 \times F_1$, where $F_0 = f^{-1}(0)$ and $F_1 = f^{-1}(1)$, with entries $H(u, v)$. Since H reads at most n_u variables from u and at most n_v variables from v, and since

$|F_0|, |F_1| < 2^n$, keeping one representative for each distinct row and column gives an $r \times s$ matrix M with $r \le \min\{2^{n_u}, 2^n - 1\}$ and $s \le \min\{2^{n_v}, 2^n - 1\}$. We claim that for some $t \le \min\{n, \ell\}$,

$$\exists\, v^{(1)}, \ldots, v^{(t)} \in F_1 \colon \forall x \in F_0\ \exists j, H(x, v^{(j)}) = 0,$$

or

$$\exists\, u^{(1)}, \ldots, u^{(t)} \in F_0 \colon \forall x \in F_1\ \exists j, H(u^{(j)}, x) = 1.$$

To prove this, we keep a set R of uncovered rows and a set C of uncovered columns of M, initially all r rows and all s columns. If at least half of the entries of M in $R \times C$ are zeros, then some column in C has zeros in at least half of the rows in R, and we pick it and remove these rows from R. Otherwise, symmetrically, some row in R has ones in more than half of the columns in C, and we pick it and remove these columns from C. Each picked column halves R and each picked row halves C, so before R or C becomes empty, we pick at most $\lceil \log_2(r+1) \rceil$ columns and at most $\lceil \log_2(s+1) \rceil$ rows. If R becomes empty, the picked columns give the first alternative, and if C becomes empty, the picked rows give the second. Since $n_u + n_v \le \ell$ with $n_u, n_v \ge 1$, we have $\lceil \log_2(r+1) \rceil \le \min\{n_u + 1, n\} \le \min\{n, \ell\}$ and similarly $\lceil \log_2(s+1) \rceil \le \min\{n, \ell\}$, and hence $t \le \min\{n, \ell\}$.

In the first case, set

$$G(x) = \bigwedge_{j=1}^{t} \left(H(x, v^{(j)}) \vee \mathrm{EQ}_{v^{(j)}}(x) \right).$$

For $x \in F_0$, we have $\mathrm{EQ}_{v^{(j)}}(x) = 0$ for every j because $v^{(j)} \in F_1$, and $H(x, v^{(j)}) = 0$ for at least one j by the covering property, so $G(x) = 0$. For $x \in F_1$, each factor is one: if $x = v^{(j)}$, then $\mathrm{EQ}_{v^{(j)}}(x) = 1$, and if $x \neq v^{(j)}$, then $(x, v^{(j)}) \in \mathrm{Inj}_2(F_1)$ and hence $H(x, v^{(j)}) = 1$. Thus G computes f. The second case is symmetric.

Each term has depth at most $\max\{d, \lceil \log_2 n \rceil\} + 1$ and at most $\ell + n$ leaves, and a balanced tree over the t terms adds depth $\lceil \log_2 t \rceil \le \min\{d, \lceil \log_2 n \rceil\}$, since $t \le \min\{n, \ell\}$ and $\ell \le 2^d$. Hence G has depth at most $\max\{d, \lceil \log_2 n \rceil\} + \min\{d, \lceil \log_2 n \rceil\} + 1 = d + \lceil \log_2 n \rceil + 1$ and at most $t(\ell + n) \le \min\{n, \ell\} \cdot (\ell + n) \le 2n\ell$ leaves. □

5 The 1-out-of-k Problem

In this section, we prove our main result, Theorem 1.2.

Theorem 1.2. *Let $\mathcal{R} \subseteq A \times B \times O$ be a total relation, with $1 \le |A|, |B| \le 2^n$. For every $1 \le k \le \min\{|A|, |B|\}$,*

$$\mathrm{CC}(\mathcal{R}) \le \max\left\{ \mathrm{CC}(\mathrm{choose}^k_{\mathcal{R}}), \lceil \log_2 |O| \rceil \right\} + O(\log k + \log n).$$

Before proving Theorem 1.2, we derive the Boolean-function and agreement consequences stated in the introduction.

Theorem 1.1. *For every total $F \colon A \times B \to \{0,1\}$ with $|A|, |B| \le 2^n$, and every $1 \le k \le \min\{|A|, |B|\}$,*

$$\mathrm{CC}(F) \le \mathrm{CC}(\mathrm{choose}^k_F) + O(\log k + \log n).$$

Proof. Let $D = \mathrm{CC}(\mathrm{choose}^k_F)$. By Theorem 1.2 with the output set $\{0,1\}$,

$$\mathrm{CC}(F) \le \max\{D, 1\} + O(\log k + \log n) \le D + O(\log k + \log n).$$ □

Corollary 1.3. *For every $k \geq 1$ and every partial $f\colon \{0,1\}^n \to \{0,1,*\}$ with $|f^{-1}(0)|, |f^{-1}(1)| \geq k$,*

$$\mathrm{D}(f) \leq \mathrm{D}(\mathrm{agree}_f^k) + O(\log k + \log n).$$

Proof. By Theorem 3.2 and Remark 3.5, we have $\mathrm{D}(f) = \mathrm{CC}(\mathrm{KW}_f)$ and $\mathrm{D}(\mathrm{agree}_f^k) = \mathrm{CC}(\mathrm{choose}_{\mathrm{KW}_f}^k)$. The relation $\mathrm{KW}_f \subseteq f^{-1}(0) \times f^{-1}(1) \times [n]$ is total because $f^{-1}(0)$ and $f^{-1}(1)$ are disjoint. Since moreover $|f^{-1}(0)|, |f^{-1}(1)| \leq 2^n$ and $k \leq \min\{|f^{-1}(0)|, |f^{-1}(1)|\}$, Theorem 1.2 with the output set $[n]$ gives

$$\begin{aligned}\mathrm{D}(f) &\leq \max\{\mathrm{D}(\mathrm{agree}_f^k), \lceil \log_2 n \rceil\} + O(\log k + \log n)\\ &\leq \mathrm{D}(\mathrm{agree}_f^k) + O(\log k + \log n).\end{aligned}$$

□

The proof of Theorem 1.2 combines protocols that are correct only on part of the inputs. Throughout, $\mathcal{R} \subseteq A \times B \times O$ is the total relation of Theorem 1.2: Alice's inputs lie in A, Bob's inputs lie in B, and O is the set of outputs. For $A' \subseteq A$ and $B' \subseteq B$, a protocol *solves* $\mathcal{R}$ *on* $A' \times B'$ if, whenever Alice receives $x \in A'$ and Bob receives $y \in B'$, the output o at the leaf they reach satisfies $(x, y, o) \in \mathcal{R}$. We call the rectangle $A' \times B'$ *solvable* if some protocol of depth at most $\max\{\mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^k), \lceil \log_2 |O| \rceil\}$ solves $\mathcal{R}$ on it. Also, we regard a rectangle as the pair of its sides, so that $A \times \varnothing$ and $\varnothing \times B$ are different solvable rectangles.

Fix an optimal protocol Π for $\mathrm{choose}_{\mathcal{R}}^k$, of depth $\mathrm{CC}(\mathrm{choose}_{\mathcal{R}}^k)$. We follow the plan of Section 2.2: Lemma 5.2 below gives a large solvable rectangle for each pair of distributions, and the proof of Theorem 1.2 combines these rectangles by minimax and majority into a protocol for $\mathcal{R}$.

For each node v of Π, let $\mathcal{A}_v \subseteq \mathrm{Inj}_k(A)$ be the set of tuples on which Alice's messages agree with the edges leaving her nodes on the path from the root to v, and define $\mathcal{B}_v \subseteq \mathrm{Inj}_k(B)$ in the same way for Bob. For each leaf w with output (i_w, o_w), let $S_w = \{x_{i_w} : (x_1, \ldots, x_k) \in \mathcal{A}_w\}$ and $T_w = \{y_{i_w} : (y_1, \ldots, y_k) \in \mathcal{B}_w\}$ be the projections of $\mathcal{A}_w$ and $\mathcal{B}_w$ onto coordinate i_w. Since every pair of tuples in $\mathcal{A}_w \times \mathcal{B}_w$ reaches w and receives the output (i_w, o_w), we have

$$(x, y, o_w) \in \mathcal{R} \qquad \text{for all } x \in S_w \text{ and } y \in T_w. \tag{1}$$

In the end, Alice and Bob hold single inputs x and y instead of tuples, so they cannot simply run Π; instead, they walk down its tree and choose the children at their nodes freely. For a set L of leaves of Π, Alice *can force* L if she can choose one child at each of her nodes so that the walk ends in L whatever children Bob chooses at his nodes; Bob can force L in the symmetric sense.

Lemma 5.1 ([Zer13])**.** *For every set L of leaves of Π, either Alice can force L or Bob can force the complement of L, but not both.*

The following lemma is the main part of our argument.

Lemma 5.2. *For every distribution μ on A and distribution ν on B, some solvable rectangle $A' \times B'$ satisfies*

$$\mu(A') + \nu(B') \geq 1 + k^{-2}.$$

Proof. Suppose first that $\mu(x_0) \geq k^{-2}$ for some $x_0 \in A$. The rectangle $\{x_0\} \times B$ is solvable: Bob uses $\lceil \log_2 |O| \rceil$ bits to announce some $o \in O$ with $(x_0, y, o) \in \mathcal{R}$, which exists by totality, and the players output o. Since $\mu(x_0) + \nu(B) \geq 1 + k^{-2}$, this rectangle satisfies the claim. Symmetrically, if $\nu(y_0) \geq k^{-2}$ for some $y_0 \in B$, then Alice announces an output that is valid for (x, y_0), and the rectangle $A \times \{y_0\}$ satisfies the claim.

Otherwise, draw $X = (x_1, \dots, x_k) \sim \mu^k$ and $Y = (y_1, \dots, y_k) \sim \nu^k$ independently. For any two distinct indices, the collision probability under μ is $\sum_{x \in A} \mu(x)^2 \le \max_{x \in A} \mu(x) < k^{-2}$, so a union bound over the $\binom{k}{2}$ pairs gives $\Pr[X \notin \mathrm{Inj}_k(A)] \le \binom{k}{2} k^{-2}$, and similarly $\Pr[Y \notin \mathrm{Inj}_k(B)] \le \binom{k}{2} k^{-2}$. For $0 \le j \le k$, let p_j be the probability that Alice can force the set

$$L_j = \{w : i_w \le j,\ x_{i_w} \in S_w\} \cup \{w : i_w > j,\ y_{i_w} \notin T_w\},$$

in which leaves at positions up to j are judged by Alice's tuple and the other leaves by Bob's tuple. If $X \in \mathrm{Inj}_k(A)$, Alice can force L_k by following the messages of Π on X, since every leaf w reached in this way satisfies $X \in \mathcal{A}_w$ and hence $x_{i_w} \in S_w$. Similarly, if $Y \in \mathrm{Inj}_k(B)$, Bob can force the complement of L_0, and then Alice cannot force L_0 by Lemma 5.1. Hence $p_k - p_0 \ge 1 - 2\binom{k}{2} k^{-2} = 1/k$, so some index $i \in [k]$ satisfies $p_i - p_{i-1} \ge k^{-2}$. By averaging over all coordinates x_j and y_j with $j \ne i$, we can fix their values so that the difference $p_i - p_{i-1}$ remains at least k^{-2}; the coordinates $x_i \sim \mu$ and $y_i \sim \nu$ remain independent.

After this fixing, L_{i-1} and L_i contain the same set L' of leaves w with $i_w \ne i$; let L'' be the set of the remaining leaves w with $i_w \ne i$. Let A' be the set of $x \in A$ for which Alice can force $L' \cup \{w : i_w = i,\ x \in S_w\}$, and let B' be the set of $y \in B$ for which Bob can force $L'' \cup \{w : i_w = i,\ y \in T_w\}$. The rectangle $A' \times B'$ is solvable: on inputs $x \in A'$ and $y \in B'$, the players walk down the tree of Π, which has depth $\mathrm{CC}(\mathrm{choose}^k_{\mathcal{R}})$; at each node, its owner sends one bit naming the child prescribed by a forcing strategy for the set in the definition of A' or B'. The leaf w they reach lies in both forced sets; since L' and L'' are disjoint, $i_w = i$, so $x \in S_w$ and $y \in T_w$, and the output o_w is valid by (1).

Finally, Alice can force $L_i = L' \cup \{w : i_w = i,\ x_i \in S_w\}$ exactly when $x_i \in A'$. Since $L_{i-1} = L' \cup \{w : i_w = i,\ y_i \notin T_w\}$ is the complement of $L'' \cup \{w : i_w = i,\ y_i \in T_w\}$, Lemma 5.1 shows that Alice can force L_{i-1} exactly when $y_i \notin B'$. Therefore,

$$k^{-2} \le p_i - p_{i-1} = \Pr[x_i \in A'] - \Pr[y_i \notin B'] = \mu(A') + \nu(B') - 1. \qquad \square$$

Proof of Theorem 1.2. Lemma 5.2 gives a separate rectangle for each pair of distributions; we first turn these rectangles into one random rectangle that works for all inputs at once. Let $\mathcal{H}$ be the set of all solvable rectangles; it is finite and contains $A \times \varnothing$ and $\varnothing \times B$. Recall that $A' \times B'$ covers $x \in A$ if $x \in A'$, and covers $y \in B$ if $y \in B'$, and put $\delta = 1/2 + k^{-2}/4$. We want a distribution Λ on $\mathcal{H}$ such that

$$\Pr_{H \sim \Lambda}[H \text{ covers } z] \ge \delta \quad \text{for every } z \in A \sqcup B.$$

By Lemma 3.6, applied to the set $\mathcal{H}$, to $\mathcal{Z} = A \sqcup B$, and to the relation $C = \{(H, z) : H \text{ covers } z\}$, it suffices to show that for every distribution π on $A \sqcup B$, some $H \in \mathcal{H}$ satisfies

$$\Pr_{z \sim \pi}[H \text{ covers } z] \ge \delta.$$

Fix π, and let $\lambda = \Pr_{z \sim \pi}[z \in A]$. If $\lambda \ge \delta$, then $H = A \times \varnothing$ works, since it covers z exactly when $z \in A$; similarly, if $1 - \lambda \ge \delta$, then $H = \varnothing \times B$ works. Otherwise, $|\lambda - 1/2| < k^{-2}/4$. Let μ and ν be the conditional distributions of π on A and on B, respectively, and let $A' \times B'$ be the rectangle of Lemma 5.2, so that $\mu(A') + \nu(B') \ge 1 + k^{-2}$. Then $H = A' \times B'$ works: since $|\mu(A') - \nu(B')| \le 1$,

we have

$$\begin{aligned}\Pr_{z\sim\pi}[H \text{ covers } z] &= \lambda\mu(A') + (1-\lambda)\nu(B') \\ &= \frac{\mu(A')+\nu(B')}{2} + \left(\lambda - \frac{1}{2}\right)(\mu(A') - \nu(B')) \\ &\geq \frac{1+k^{-2}}{2} - \frac{k^{-2}}{4} \\ &= \delta.\end{aligned}$$

Hence there is a distribution Λ such that for a random rectangle $A' \times B'$ from Λ,

$$\Pr[x \in A'] \geq \delta \text{ for every } x \in A, \qquad \Pr[y \in B'] \geq \delta \text{ for every } y \in B.$$

We now amplify the probability by a majority argument. Regard every solvable rectangle $A' \times B'$ as the function on $A \sqcup B$ that equals $\mathbf{1}[x \notin A']$ at each $x \in A$ and $\mathbf{1}[y \in B']$ at each $y \in B$, and let $\mathcal{F}$ be the set of these functions. Then Λ is a distribution over formulas of depth zero with leaves from $\mathcal{F}$, and by the display above, a leaf drawn from Λ is zero at each $x \in A$ with probability at least δ and one at each $y \in B$ with probability at least δ. Since $|A \sqcup B| \leq 2^{n+1}$, Lemma 3.7 with $\mathcal{Z} = A \sqcup B$ and $\varepsilon = k^{-2}/4$ gives a formula Ψ of depth $O(\log k + \log n)$ with leaves from $\mathcal{F}$ that is zero at every $x \in A$ and one at every $y \in B$. On inputs $x \in A$ and $y \in B$, the players walk down Ψ from the root while the current subformula is zero at x and one at y, as in the proof of Theorem 3.2 [KW90]: at an AND gate, both children are one at y, and Alice names a child that is zero at x; at an OR gate, both children are zero at x, and Bob names a child that is one at y. At the leaf $A' \times B'$ they reach, $x \in A'$ and $y \in B'$, so they finish by running a protocol of depth at most $\max\{\mathrm{CC}(\mathrm{choose}^k_{\mathcal{R}}), \lceil \log_2 |O| \rceil\}$ that solves $\mathcal{R}$ on $A' \times B'$. The resulting protocol for $\mathcal{R}$ has depth

$$\max\{\mathrm{CC}(\mathrm{choose}^k_{\mathcal{R}}), \lceil \log_2 |O| \rceil\} + O(\log k + \log n). \qquad \square$$

Finally, we derive the bound that holds even for exponentially many instances.

Corollary 1.4. *For F, k as in Theorem 1.1 and f, k as in Corollary 1.3,*

$$\begin{aligned}\mathrm{CC}(\mathrm{choose}^k_F) &\geq \Omega(\mathrm{CC}(F)) - O(\log n), \\ \mathrm{D}(\mathrm{agree}^k_f) &\geq \Omega(\mathrm{D}(f)) - O(\log n).\end{aligned}$$

Proof. Let $\mathcal{R}$ be either F or KW_f (with $A = f^{-1}(0)$, $B = f^{-1}(1)$, and $O = [n]$), and let Π be an optimal protocol for $\mathrm{choose}^k_{\mathcal{R}}$, of depth d. The leaves of Π name a set $I \subseteq [k]$ of at most 2^d positions; put $r = |I|$. Given tuples in $\mathrm{Inj}_r(A)$ and $\mathrm{Inj}_r(B)$, the players place them at the positions in I, and each player fills the other $k - r$ positions with distinct inputs outside its own tuple by a fixed rule, which is possible since $k \leq \min\{|A|, |B|\}$. Since Π always answers at a position in I, running Π on these tuples solves $\mathrm{choose}^r_{\mathcal{R}}$ in depth d. Hence Theorem 1.2 with $r \leq 2^d$ instances gives

$$\mathrm{CC}(\mathcal{R}) \leq \max\{d, \lceil \log_2 |O| \rceil\} + O(d + \log n),$$

where $\lceil \log_2 |O| \rceil \leq 1 + \log_2 n$ in both cases. For $\mathcal{R} = F$, this is the first bound; for $\mathcal{R} = \mathrm{KW}_f$, it is the second, since $\mathrm{CC}(\mathrm{KW}_f) = \mathrm{D}(f)$ by Theorem 3.2 and $d = \mathrm{D}(\mathrm{agree}^k_f)$ by Remark 3.5. $\square$

6 Dense Restrictions and Universal Composition

This section proves our results on dense restrictions. For a nonconstant total $g\colon \{0,1\}^n \to \{0,1\}$ and a matrix $X \in \{0,1\}^{m\times n}$ with rows $X_1, \ldots, X_m$, write $g^m(X) = (g(X_1), \ldots, g(X_m))$. We identify a matrix with the tuple of its rows, so that $S^m \subseteq \{0,1\}^{m\times n}$ for every $S \subseteq \{0,1\}^n$. Recall that for a nonempty subset $\mathcal{X} \subseteq \{0,1\}^{m\times n}$, in the restriction $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ of $U_m \diamond \mathrm{KW}_g$, the players receive $X, Y \in \mathcal{X}$ and must output a cell (i,j) with $X_{i,j} \neq Y_{i,j}$, or to reject when $g^m(X) = g^m(Y)$. We call any nonempty set $\mathcal{X}_0 \subseteq \{0,1\}^{m\times n}$ a *core*, and the *relative density* of a subset $\mathcal{X} \subseteq \mathcal{X}_0$ is $|\mathcal{X}|/|\mathcal{X}_0|$. Unless stated otherwise, the core is $\{0,1\}^{m\times n}$ itself.

Section 6.1 encodes protocols for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ as formulas with auxiliary variables and turns them into predictors for g. Section 6.2 uses these predictors to prove Theorem 1.7, and Section 6.3 establishes a bound of the same form for random functions at density 2^{1-m} in $\{0,1\}^{m\times n}$ in Theorem 6.6. Section 6.4 derives the composition bounds, and Section 6.5 proves Theorem 1.8.

6.1 From Protocols to Formulas

Recall that the KW correspondence [KW90] turns a protocol for KW_f into a formula for f by labeling every leaf with a literal on which the two inputs differ. We apply a similar idea to protocols for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$. To get a De Morgan formula, we need to decide how to handle the reject leaves. Consider a reject leaf of a protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$. Every matrix in a pair (X,Y) that reaches this leaf has the same value $c \in \{0,1\}^m$ of g^m, that is, $g^m(X) = g^m(Y) = c$. We label such a leaf with an auxiliary variable p_c.

Lemma 6.1. *For every protocol of depth d for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ there is a De Morgan formula $\phi(X,p)$ of depth at most d in the matrix variables $X = (X_{i,j}) \in \{0,1\}^{m\times n}$ and the auxiliary variables $p = (p_c)_{c\in\{0,1\}^m}$, such that no auxiliary variable appears negated and*

$$\phi(X,p) = p_{g^m(X)} \qquad \text{for every } X \in \mathcal{X} \text{ and every assignment } p \text{ to the auxiliary variables.}$$

Proof. We first turn the protocol into a formula. Remove all nodes v of the protocol whose reaching rectangle $\mathcal{X}_v \times \mathcal{Y}_v \subseteq \mathcal{X}\times\mathcal{X}$ is empty, and contract every node that has only one remaining child. At a leaf v that outputs a cell (i,j), every pair $(X,Y) \in \mathcal{X}_v \times \mathcal{Y}_v$ satisfies $X_{i,j} \neq Y_{i,j}$, so $X_{i,j}$ is constant on $\mathcal{X}_v$ and takes the opposite value on $\mathcal{Y}_v$. We label this leaf with the literal $X_{i,j}$ or $\neg X_{i,j}$ that evaluates to zero on $\mathcal{X}_v$ and to one on $\mathcal{Y}_v$. At a reject leaf v, every pair $(X,Y) \in \mathcal{X}_v \times \mathcal{Y}_v$ satisfies $g^m(X) = g^m(Y)$, so all matrices in $\mathcal{X}_v \cup \mathcal{Y}_v$ share a single value $c \in \{0,1\}^m$ of g^m; we label this leaf with the auxiliary variable p_c. All reject leaves with the same value c share the same auxiliary variable p_c. Replace Alice's nodes by AND gates and Bob's nodes by OR gates.

We now show that this formula satisfies the claimed identity. For each node v, let ϕ_v be the resulting subformula and $\mathcal{X}_v \times \mathcal{Y}_v$ its reaching rectangle. For every assignment p to the auxiliary variables, induction gives

$$\phi_v(X,p) \le p_{g^m(X)} \quad \text{for } X \in \mathcal{X}_v \qquad \text{and} \qquad \phi_v(Y,p) \ge p_{g^m(Y)} \quad \text{for } Y \in \mathcal{Y}_v.$$

Indeed, these inequalities hold at every leaf: a literal leaf evaluates to $0 \le p_{g^m(X)}$ on $\mathcal{X}_v$ and to $1 \ge p_{g^m(Y)}$ on $\mathcal{Y}_v$, while a leaf labeled by an auxiliary variable p_c equals $p_{g^m(X)}$ on $\mathcal{X}_v$ and $p_{g^m(Y)}$ on $\mathcal{Y}_v$. At an Alice node v, any $X \in \mathcal{X}_v$ belongs to $\mathcal{X}_u$ for one child u, and any $Y \in \mathcal{Y}_v$ belongs to $\mathcal{Y}_u$ for both children, so taking the AND of the children's subformulas preserves both inequalities. The case of a Bob node is symmetric. At the root r, we have $\mathcal{X}_r = \mathcal{Y}_r = \mathcal{X}$, so both inequalities hold for every $X \in \mathcal{X}$, and hence $\phi(X,p) = p_{g^m(X)}$. □

The next lemma turns this formula into a formula on $\{0,1\}^n$ that predicts g under a distribution ν better than a random guess whenever $\mathcal{X}$ has density above 2^{-m} under ν^m.

Lemma 6.2. *Let $m, n \geq 1$, let $g\colon \{0,1\}^n \to \{0,1\}$, let $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ be nonempty, let ν be any distribution on $\{0,1\}^n$, and let $\rho = \Pr_{X\sim\nu^m}[X \in \mathcal{X}]$ be the density of $\mathcal{X}$ under ν^m. If $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ has a depth-d protocol and $\rho \geq 2^{-m}$, then there is a formula h of depth at most $d+1$ that satisfies*

$$\Pr_{x\sim\nu}[h(x) = g(x)] \geq \frac{1}{2} + \frac{2^m\rho - 1}{2(2^m - 1)}.$$

Proof. First, for every set $U \subseteq \{0,1\}^m$ we build a formula Φ_U such that $\Phi_U(X) = \mathbf{1}[g^m(X) \in U]$ for every $X \in \mathcal{X}$. Take the formula ϕ from Lemma 6.1. Since no auxiliary variable appears negated in ϕ, it is monotone in its auxiliary variables. For a set $U \subseteq \{0,1\}^m$, let $\mathbf{1}_U \in \{0,1\}^{2^m}$ be the indicator of U. Consider the formula

$$\Phi_U(X) = \phi(X, \mathbf{1}_U) \wedge \neg\phi(X, \mathbf{1}_{\{0,1\}^m\setminus U}).$$

On $\mathcal{X}$ the second conjunct is redundant: for $X \in \mathcal{X}$ we have $\phi(X, \mathbf{1}_{\{0,1\}^m\setminus U}) = 1 - \phi(X, \mathbf{1}_U)$, so $\Phi_U(X) = \phi(X, \mathbf{1}_U)$. Outside $\mathcal{X}$, Lemma 6.1 says nothing about $\phi(X, \cdot)$, and it may happen that $\phi(X, \mathbf{1}_U) = \phi(X, \mathbf{1}_V) = 1$ for disjoint sets U and V. The second conjunct rules out $\Phi_U(X) = \Phi_V(X) = 1$. The formula Φ_U has depth at most $d+1$, since negation can be pushed to the input literals and constants. For disjoint U, V, monotonicity gives the pointwise inequality

$$\Phi_U(X) + \Phi_V(X) \leq \Phi_{U\sqcup V}(X).$$

Indeed, if either $\Phi_U(X) = 1$ or $\Phi_V(X) = 1$, monotonicity in p implies $\Phi_{U\sqcup V}(X) = 1$, and both cannot equal one simultaneously because $U \subseteq \{0,1\}^m \setminus V$ implies $\phi(X, \mathbf{1}_U) \leq \phi(X, \mathbf{1}_{\{0,1\}^m\setminus V})$.

Next, we measure how often Φ_U is one when U is the set of vectors $c \in \{0,1\}^m$ that agree with g on the first i rows. For $X \sim \nu^m$ and $0 \leq i \leq m$, define

$$C_i(X) = \{c \in \{0,1\}^m : c_j = g(X_j) \text{ for } 1 \leq j \leq i\}, \qquad M_i = \mathbb{E}_X[\Phi_{C_i(X)}(X)].$$

Here $C_0(X) = \{0,1\}^m$ and $M_0 \leq 1$. Moreover, $C_m(X) = \{g^m(X)\}$, and $\Phi_{\{g^m(X)\}}(X) = 1$ for every $X \in \mathcal{X}$ since $\phi(X, p) = p_{g^m(X)}$. Hence $M_m \geq \rho$.

Let α be the largest advantage over $1/2$ with which a formula of depth at most $d+1$ predicts g under ν, that is,

$$\alpha = \max_h \Pr_{x\sim\nu}[h(x) = g(x)] - \frac{1}{2},$$

where h ranges over all formulas of depth at most $d+1$ on $\{0,1\}^n$. We now show that $M_i \leq \frac{1}{2}M_{i-1} + \alpha$ for every $i \in [m]$. Informally, $\Phi_U(X) = 1$ means that at X the formula ϕ claims $g^m(X) \in U$, so M_i is the probability that ϕ claims the correct values of g on the first i rows. Fix $i \in [m]$ and fix all rows of X except the ith row X_i. The set $C_{i-1}(X)$ depends only on the first $i-1$ rows, so it is fixed. Partition it into $U_0 = \{c \in C_{i-1}(X) : c_i = 0\}$ and $U_1 = \{c \in C_{i-1}(X) : c_i = 1\}$. For $b \in \{0,1\}$, let $h_b(x)$ be the formula in $x \in \{0,1\}^n$ obtained from $\Phi_{U_b}(X)$ by substituting x for the ith row X_i. Both h_0 and h_1 have depth at most $d+1$. Thus $h_b(x) = 1$ means that, at the matrix with ith row x, the formula ϕ claims the correct values on the first $i-1$ rows and the value b on the ith row; we then say for short that ϕ claims b at x. Since U_0 and U_1 are disjoint, the pointwise inequality for Φ above shows that $h_0(x)$ and $h_1(x)$ cannot both equal one, so ϕ claims at most one value at x. Hence $h_{g(x)}(x) = 1$ exactly when ϕ claims the correct value $g(x)$, and $h_0(x) + h_1(x) = 1$ exactly when ϕ claims some value. To predict $g(x)$, consider a predictor that outputs the value claimed

by ϕ at x, or a uniformly random bit if ϕ claims no value. This predictor is the uniform mixture of the formulas h_1 and $\neg h_0$: if ϕ claims a value, both formulas output this value, and otherwise h_1 outputs zero and $\neg h_0$ outputs one. Hence, for every $x \in \{0,1\}^n$, this predictor outputs $g(x)$ with probability

$$\frac{\mathbf{1}[h_1(x) = g(x)] + \mathbf{1}[\neg h_0(x) = g(x)]}{2} = \frac{1}{2} + h_{g(x)}(x) - \frac{h_0(x) + h_1(x)}{2}.$$

Indeed, both sides equal one if ϕ claims $g(x)$, zero if ϕ claims $1 - g(x)$, and $1/2$ if ϕ claims no value. Both h_1 and $\neg h_0$ are formulas of depth at most $d+1$, since negation can be pushed to the input literals and constants; hence each agrees with g under ν with probability at most $1/2+\alpha$, and so does their mixture. This holds for every fixing of the rows other than the ith. Since $C_i(X) = U_{g(X_i)}$, we have $h_{g(X_i)}(X_i) = \Phi_{C_i(X)}(X)$, whose expectation over X is M_i. Taking the expectation over $X_i \sim \nu$ and then over the remaining rows of X, and writing $T_i = \mathbb{E}_X[h_0(X_i) + h_1(X_i)]$ for the probability that ϕ claims a value on row i, we obtain $M_i - T_i/2 \le \alpha$. Since U_0 and U_1 are disjoint with $U_0 \sqcup U_1 = C_{i-1}(X)$, the pointwise inequality above gives $h_0(X_i) + h_1(X_i) \le \Phi_{C_{i-1}(X)}(X)$, so $T_i \le M_{i-1}$ and hence

$$M_i \le \frac{1}{2} M_{i-1} + \alpha.$$

Finally, iterating this inequality for $i = 1, \dots, m$ yields

$$\rho \le M_m \le 2^{-m} M_0 + \alpha \sum_{j=0}^{m-1} 2^{-j} \le 2^{-m} + 2\alpha(1 - 2^{-m}).$$

Rearranging gives $\alpha \ge (2^m\rho - 1)/(2(2^m - 1))$. Hence a formula h attaining the maximum in the definition of α satisfies the lemma. □

6.2 Hard Cores for Every Function

The core S^m of Theorem 1.7 is built from a set $S \subseteq \{0,1\}^n$ on which shallow formulas cannot predict g well. The next lemma shows that such a set exists whenever g is hard for formulas.

Lemma 6.3. *Let $g\colon \{0,1\}^n \to \{0,1\}$, and let $t \ge 0$ be an integer. Suppose that every nonempty $S \subseteq \{0,1\}^n$ admits a formula of depth at most t that agrees with g on at least $2|S|/3$ points of S. Then $\mathrm{D}(g) \le t + O(\log n)$.*

Proof. We show that if $\mathrm{D}(g)$ is larger, then rounding a hard distribution produces a set S that violates the assumption. Let $K = \Theta(2^t \log n)$ be an integer large enough for the union bound at the end of this proof, and let $d = \lceil \log_2 n \rceil + \lceil \log_2 K \rceil$, so that $t \le d \le t + O(\log n)$. A formula of depth d can memorize g on any K points and output a fixed constant elsewhere: take a balanced OR of the equality tests for the points where g differs from the constant, and complement it if necessary.

We first find a distribution under which no formula of depth at most d predicts g well. Suppose that $\mathrm{D}(g)$ exceeds d by more than the depth that Lemma 3.7 adds for $\varepsilon = 1/16$ and $\mathcal{Z} = \{0,1\}^n$, which is $O(\log n)$. Then some distribution μ on $\{0,1\}^n$ makes every formula h of depth at most d inaccurate: its accuracy under μ satisfies

$$\Pr_{x \sim \mu}[h(x) = g(x)] < 9/16.$$

Indeed, otherwise Lemma 3.6, applied to the set $\mathcal{H}$ of formulas of depth at most d, to $\mathcal{Z} = \{0,1\}^n$, and to the relation $C = \{(h, x) : h(x) = g(x)\}$, gives a distribution Λ over these formulas such that

$$\Pr_{h \sim \Lambda}[h(x) = g(x)] \ge 9/16 \quad \text{for every } x \in \{0,1\}^n,$$

and Lemma 3.7, with $\mathcal{F}$ the set of literals and constants, turns Λ into a formula for g of depth $d + O(\log n)$, a contradiction.

Next, we remove the inputs of largest probability, since a formula of depth d can memorize g on them. The support of μ has more than K points, since otherwise memorization would be correct everywhere. Let T be the set of the K inputs of largest probability under μ. Memorizing g on T and using the better constant outside T gives accuracy at least $\mu(T) + (1 - \mu(T))/2$, so $\mu(T) < 1/8$. Let ν be the conditional distribution of $x \sim \mu$ given $x \notin T$. Every input outside T has probability at most $\mu(T)/K < 1/(8K)$ under μ, so $\nu(x) = \mu(x)/(1 - \mu(T)) < 1/(7K)$ for every x. Moreover, every formula of depth at most $t \le d$ has accuracy less than $(9/16)/(7/8) = 9/14$ under ν.

Finally, we round ν to a set S that violates the assumption. Include each point x in a random set S independently with probability $K\nu(x) < 1$. For a fixed formula h of depth at most t, let

$$w_h = \sum_{x \in \{0,1\}^n} \mathbf{1}[x \in S](\mathbf{1}[h(x) = g(x)] - 2/3)$$

be the number of points of S on which h agrees with g, minus $2|S|/3$. It is a sum of independent variables bounded by one in absolute value, with mean less than $K(9/14 - 2/3) = -K/42$ and variance at most K, so Bernstein's inequality gives $\Pr[w_h \ge 0] \le 2^{-\Omega(K)}$. Padding to a complete binary tree shows that there are at most $2^{2^t}(2n+2)^{2^t} = 2^{2^t \log_2(4n+4)}$ formulas of depth at most t. Therefore, by the choice of K, a union bound over these formulas gives a set S with $w_h < 0$ for every such h. This set is nonempty, since $w_h = 0$ on the empty set, and it violates the assumption. Hence $\mathrm{D}(g) \le d + O(\log n) = t + O(\log n)$. □

Theorem 1.7. *For every nonconstant $g\colon \{0,1\}^n \to \{0,1\}$, there is a nonempty $S \subseteq \{0,1\}^n$ such that, for every $m \ge 2$ and every $\mathcal{X} \subseteq S^m$ of relative density at least $(1 + 2^{1-m})/3$,*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \ge \mathrm{D}(g) - O(\log n).$$

Proof. We first choose S as a set on which formulas of depth about $\mathrm{D}(g)$ cannot predict g well. For an integer $t \ge 0$, call a nonempty set $S \subseteq \{0,1\}^n$ *t-hard* if no formula of depth t agrees with g on at least $2|S|/3$ points of S. In these terms, Lemma 6.3 says that if no set is t-hard, then $\mathrm{D}(g) \le t + O(\log n)$. If no set is 0-hard, the lemma with $t = 0$ gives $\mathrm{D}(g) = O(\log n)$, and the theorem holds trivially for every nonempty S. Otherwise, let t be the largest integer such that some set is t-hard, and fix a t-hard set S. This t exists and is smaller than $\mathrm{D}(g)$, since a formula for g of depth $\mathrm{D}(g)$ agrees with g everywhere, so no set is t'-hard for $t' \ge \mathrm{D}(g)$. By the maximality of t, no set is $(t+1)$-hard, so Lemma 6.3 with $t+1$ in place of t gives $\mathrm{D}(g) \le t + 1 + O(\log n)$.

Now let $m \ge 2$, let $\mathcal{X} \subseteq S^m$ have relative density $\rho \ge (1 + 2^{1-m})/3$, and suppose that $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ has a protocol of depth d. Let ν be the uniform distribution on S. Then $\Pr_{X\sim\nu^m}[X \in \mathcal{X}] = |\mathcal{X}|/|S|^m = \rho$, and $\rho \ge 2^{-m}$, since $(1 + 2^{1-m})/3 \ge 2^{-m}$. Hence Lemma 6.2 gives a formula h of depth at most $d+1$ that agrees with g on a fraction at least

$$\frac{1}{2} + \frac{2^m\rho - 1}{2(2^m - 1)} \ge \frac{1}{2} + \frac{(2^m + 2)/3 - 1}{2(2^m - 1)} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

of the points of S. Since S is t-hard, $d + 1 > t$, that is, $d \ge t$. This holds for every protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$, so $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \ge t \ge \mathrm{D}(g) - 1 - O(\log n)$. □

6.3 Dense Restrictions of Random Functions

This subsection proves Theorem 6.6. We must show that, for almost every g, no $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of relative density at least 2^{1-m} admits a shallow protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$. A union bound

over the sets $\mathcal{X}$ is hopeless, since there are $2^{2^{mn}}$ of them. Instead, we use Lemma 6.1: a protocol of depth d for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ gives a formula of depth at most d that by itself determines $g^m(X)$ for every matrix $X \in \mathcal{X}$. Thus a shallow protocol for such a set $\mathcal{X}$ yields a shallow formula whose predictor of g^m is correct on at least a 2^{1-m} fraction of the matrices. For a fixed formula and a random g, this happens only with tiny probability (see Lemma 6.4 below), and there are few shallow formulas, so a union bound over formulas proves the theorem.

For a fixed formula $\phi(X, p)$ as in Lemma 6.1 and a matrix $X \in \{0,1\}^{m\times n}$, define $\tau_\phi(X) = c$ if $\phi(X, p) = p_c$ for every assignment p to the auxiliary variables, and put $\tau_\phi(X) = \perp$ otherwise. Since different auxiliary variables define distinct functions of p, this gives a well-defined map $\tau_\phi \colon \{0,1\}^{m\times n} \to \{0,1\}^m \cup \{\perp\}$ that depends only on ϕ. If ϕ is obtained by Lemma 6.1 from a protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$, then $\phi(X, p) = p_{g^m(X)}$, and hence $\tau_\phi(X) = g^m(X)$ for every $X \in \mathcal{X}$. However, τ_ϕ is defined for every De Morgan formula $\phi(X, p)$ in the matrix variables X and the auxiliary variables p, not necessarily obtained from a protocol for some $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$, so the union bound below can range over all such formulas.

Throughout this subsection, put $N = 2^n$. The next lemma shows that a fixed predictor is correct on at least a 2^{1-m} fraction of the matrices only with probability $2^{-\Omega(N/m^2)}$ over a uniformly random function g.

Lemma 6.4. *Let $m \geq 1$ satisfy $N \geq 4m^2$, and fix a map $\tau \colon \{0,1\}^{m\times n} \to \{0,1\}^m \cup \{\perp\}$. For a uniformly random function $g \colon \{0,1\}^n \to \{0,1\}$, put $q_\tau(g) = \Pr_{X\in\{0,1\}^{m\times n}}[\tau(X) = g^m(X)]$. Then for every integer $t \geq 1$,*

$$\mathop{\mathbb{E}}_{g}[q_\tau(g)^t] \leq 2^{-mt} \exp\left(\frac{m^2t^2}{2N}\right), \qquad \textit{and consequently} \qquad \Pr_{g}[q_\tau(g) \geq 2^{1-m}] \leq 2^{-\Omega(N/m^2)}.$$

Proof. We first bound the tth moment of $q_\tau(g)$ in terms of the number of distinct rows of t random matrices. Sample t independent uniform matrices in $\{0,1\}^{m\times n}$, and let J be the set of distinct rows among their mt rows. Conditioned on these t matrices, the event that $\tau(X) = g^m(X)$ for all t matrices requires that none of them has $\tau(X) = \perp$, that their predicted bits agree whenever two rows coincide, and that $g(x)$ matches the predicted bit at each of the $|J|$ distinct rows $x \in J$. Thus the conditional probability that all t predictions are correct is either zero or $2^{-|J|}$, which gives $\mathbb{E}_g[q_\tau(g)^t] \leq \mathbb{E}[2^{-|J|}]$.

Next, we bound $\mathbb{E}[2^{-|J|}]$. Reveal the mt sampled rows one by one, and let J_r be the set of distinct rows among the first r of them. Then $r - |J_r|$ is the number of the first r rows that repeat an earlier row. Conditional on the revealed rows, the next row repeats one of them with probability $|J_r|/N$. Hence

$$\mathbb{E}[2^{r+1-|J_{r+1}|} \mid \text{first } r \text{ rows}] = 2^{r-|J_r|}\left(\left(1 - \tfrac{|J_r|}{N}\right)\cdot 1 + \tfrac{|J_r|}{N}\cdot 2\right) \leq 2^{r-|J_r|}(1 + r/N).$$

Iterating from $J_0 = \varnothing$ and using $1 + x \leq e^x$ gives

$$\mathbb{E}[2^{-|J|}] \leq 2^{-mt} \prod_{r=0}^{mt-1} (1 + r/N) \leq 2^{-mt} \exp\left(\frac{m^2t^2}{2N}\right).$$

Finally, we derive the tail bound. Take $t = \lfloor N \ln 2/m^2 \rfloor$. Since $N \ln 2/m^2 \geq 2$, we have $N \ln 2/(2m^2) \leq t \leq N \ln 2/m^2$, so $m^2t^2/(2N) \leq (t \ln 2)/2$, and Markov's inequality gives

$$\Pr_{g}[q_\tau(g) \geq 2^{1-m}] \leq \frac{\mathbb{E}_g[q_\tau(g)^t]}{2^{(1-m)t}} \leq 2^{-t} \exp\left(\frac{m^2t^2}{2N}\right) \leq 2^{-t/2} = 2^{-\Omega(N/m^2)}. \qquad \square$$

To apply this bound to every protocol at once, we count formulas. For fixed m and n, every leaf of a De Morgan formula $\phi(X, p)$ is one of the $2mn$ input literals, the 2^{m+1} literals of the auxiliary variables, or two constants. Padding a formula of depth at most d to a complete binary tree of depth d shows that there are at most $2^{2^d}(2mn + 2^{m+1} + 2)^{2^d} = 2^{O(2^d(m+\log n))}$ formulas. Together, Lemma 6.4 and this count give the following bound for all numbers of rows up to about $2^{n/2}$.

Lemma 6.5. *Let $n \geq 4$ and $1 \leq M \leq \sqrt{2^n}/2$. With probability at least $1 - M2^{-\Omega(2^n/M^2)}$, a uniformly random function $g\colon \{0,1\}^n \to \{0,1\}$ satisfies*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq n - 3\log_2 m - O(\log\log n)$$

simultaneously for every $1 \leq m \leq M$ and every $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of relative density at least 2^{1-m}.

Proof. Fix $1 \leq m \leq M$ and an integer $d \geq 0$. Since $m \leq M \leq \sqrt{2^n}/2$, we have $N \geq 4m^2$, as Lemma 6.4 requires. If some $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of relative density at least 2^{1-m} has $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \leq d$, then by Lemma 6.1 some formula ϕ of depth at most d has $\tau_\phi(X) = g^m(X)$ for every $X \in \mathcal{X}$, and hence $q_{\tau_\phi}(g) \geq 2^{1-m}$. By Lemma 6.4 and the count above, this happens with probability at most $2^{O(2^d(m+\log n))-\Omega(N/m^2)}$. Now let d be the largest integer for which the first term in this exponent is at most half of the second, so that the probability above is at most $2^{-\Omega(N/m^2)}$. If $d < 0$, then $n - 2\log_2 m - \log_2(m + \log_2 n) - O(1) < 0$, so the claimed bound is negative and the lemma holds trivially for this m and every g. Otherwise,

$$d \geq n - 2\log_2 m - \log_2(m + \log_2 n) - O(1) \geq n - 3\log_2 m - O(\log\log n).$$

Outside this event, every such $\mathcal{X}$ has $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq d + 1$. A union bound over $m \leq M$ shows that the bound fails for some m with probability at most $\sum_{m\leq M} 2^{-\Omega(N/m^2)} \leq M2^{-\Omega(N/M^2)}$. □

Theorem 6.6 (Dense restrictions of random functions)**.** *With probability $1 - o(1)$, a uniformly random $g\colon \{0,1\}^n \to \{0,1\}$ satisfies, for every $1 \leq m \leq n^2$ and every $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of relative density at least 2^{1-m},*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq n - O(\log n).$$

Proof. Apply Lemma 6.5 with $M = n^2$, which satisfies $M \leq \sqrt{2^n}/2$ for sufficiently large n. The conclusion of the lemma fails with probability at most $n^2 2^{-\Omega(2^n/n^4)} = o(1)$, and for every $m \leq n^2$ the lower bound is $n - 6\log_2 n - O(\log\log n) = n - O(\log n)$. □

6.4 Universal Composition with Random Functions

This subsection derives the composition bounds from the bounds for dense restrictions by the strategy of Dinur and Meir described in Section 1: the protocol is followed on diagonal inputs for $m - 1$ bits, and a bound for dense restrictions is applied to the remaining square. The next lemma carries out this strategy. The walk in its proof is the averaging argument behind [DM16, Lemma 9.2].

Lemma 6.7. *Let $g\colon \{0,1\}^n \to \{0,1\}$ be nonconstant, let $m \geq 1$, and let $d > 0$. If every $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of density at least 2^{1-m} satisfies $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq d$, then*

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \geq m - 1 + d.$$

Proof. We find a node at depth $m-1$ whose subprotocol solves a dense restriction. Starting at the root with the set of all diagonal pairs (X, X) for $X \in \{0,1\}^{m\times n}$, walk down the protocol tree by always choosing a child that contains at least half of the remaining diagonal pairs. After t steps, the set $\mathcal{X}_t \subseteq \{0,1\}^{m\times n}$ of matrices X whose diagonal pair (X, X) reaches the current node has density at least 2^{-t}, and the rectangle of that node contains $\mathcal{X}_t \times \mathcal{X}_t$. For $t \le m-1$, the density of $\mathcal{X}_t$ is at least 2^{1-m}, so the current node cannot be a leaf because $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}_t\times\mathcal{X}_t}) \ge d > 0$. Hence the walk reaches a node at depth $m-1$, and the subprotocol rooted at this node solves $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}_{m-1}\times\mathcal{X}_{m-1}}$ and therefore has depth at least d, so the whole protocol has depth at least $m-1+d$. □

Theorem 1.6. *With probability $1-o(1)$ as $n\to\infty$, a uniformly random $g\colon\{0,1\}^n\to\{0,1\}$ satisfies, for all $m \le n^2$,*

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \ge m + n - O(\log n) \ge m + \mathrm{D}(g) - O(\log n).$$

Proof. By Theorem 6.6, with probability $1-o(1)$, simultaneously for every $1 \le m \le n^2$, every $\mathcal{X} \subseteq \{0,1\}^{m\times n}$ of density at least 2^{1-m} satisfies $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \ge n - O(\log n)$. For every such function, Lemma 6.7 gives $\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \ge m + n - O(\log n)$, and the second inequality holds since $\mathrm{D}(g) \le n + O(\log n)$. □

The bound of Theorem 1.6 is tight up to the logarithmic term, since for every g there is a natural protocol solving $U_m \diamond \mathrm{KW}_g$ using $m + \lceil \log_2(m+1) \rceil + \mathrm{D}(g)$ bits.

For more rows, Lemma 6.5 gives the following weaker bound.

Corollary 6.8. *Fix $0 < \varepsilon < 1/3$. With probability $1-o(1)$, a uniformly random $g\colon\{0,1\}^n\to\{0,1\}$ satisfies, simultaneously for every $1 \le m \le 2^{\varepsilon n}$,*

$$\mathrm{CC}(U_m \diamond \mathrm{KW}_g) \ge m + (1-3\varepsilon)n - O(\log\log n).$$

Proof. Take $M = \lfloor 2^{\varepsilon n} \rfloor$ in Lemma 6.5. For sufficiently large n, we have $M \le \sqrt{2^n}/2$, and the conclusion of the lemma fails with probability at most $M 2^{-\Omega(2^{(1-2\varepsilon)n})} = o(1)$. For every $m \le M$, the dense-subset lower bound of the lemma is at least $n - 3\varepsilon n - O(\log\log n)$, which is positive for sufficiently large n since $\varepsilon < 1/3$. Lemma 6.7 completes the proof. □

6.5 Limits of Hard Cores

This subsection proves Theorem 1.8. We must find a function g such that, for every integer $0 \le \sigma \le m-1$, every core $\mathcal{X}_0 \subseteq \{0,1\}^{m\times n}$ contains a subset $\mathcal{X}$ of relative density at least $2^{-m+\sigma}$ with $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \le \mathrm{D}(g) - m/2 + \sigma + O(\log m)$. Here σ measures how much denser than the trivial threshold 2^{-m} the subset $\mathcal{X}$ must be. Unlike in Theorem 1.7, the core $\mathcal{X}_0$ is arbitrary and need not have the form S^m for a set $S \subseteq \{0,1\}^n$. The function depends only on m, so it is fixed before $\mathcal{X}_0$ and σ are chosen. The idea is that a protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}$ does not need a formula for g: it suffices to have a shallow formula ψ that agrees with g on every row of every matrix in $\mathcal{X}$. If moreover only a few values of g^m occur in $\mathcal{X}$, then Alice can announce $g^m(X)$, Bob can name a row i with $g(X_i) \ne g(Y_i)$, and the players can solve the KW game for ψ on that row. Hence we need a function of large formula depth that has shallow predictors correct with high probability at every input. Lemma 6.9 constructs such a function, and Lemma 6.10 turns its predictors into a shallow protocol on a dense subset of any core.

Lemma 6.9. *For every integer $r \ge 2$, there is a balanced total function g on $n = 2^r + 2r + 1$ bits with $\mathrm{D}(g) \ge 3r - O(\log r)$. Moreover, for every integer $q \ge 1$, there is a distribution over formulas h of depth $2r + O(\log r + \log q)$ such that $\Pr_h[h(x) \ne g(x)] \le 2^{-q}$ for every $x \in \{0,1\}^n$.*

Proof. We first find a map T for which checking whether $y = T(u)$ requires large depth. For an arbitrary map $T\colon \{0,1\}^{2r} \to \{0,1\}^{2^r}$, let $f_T(u,y) = \mathbf{1}[y = T(u)]$ be the function of $2r + 2^r$ variables that checks whether y is the value of T at u. Different maps give different functions f_T, so there are $(2^{2^r})^{2^{2r}} = 2^{2^{3r}}$ such functions. Since $2r + 2^r \le 2^{r+1}$, padding to a complete binary tree shows that there are at most $2^{2^t}(2^{r+2}+2)^{2^t} \le 2^{2^{t(r+4)}}$ formulas of depth at most t on these variables. For $t = 3r - \lceil \log_2(r+4) \rceil - 1$, this is less than $2^{2^{3r}}$, so some f_T has depth greater than t, that is, at least $3r - O(\log r)$. Fix such a map T.

Next, we construct the predictors for f_T. Choose independent uniform $v_1, \dots, v_q \in \{0,1\}^{2^r}$ and define

$$H(u,y) = \bigwedge_{j=1}^{q} \mathbf{1}[v_j \cdot y = v_j \cdot T(u)],$$

with dot products over $\mathbb{F}_2$. If $y = T(u)$, then every conjunct is one, so $H(u,y) = 1 = f_T(u,y)$. If $y \ne T(u)$, then the jth conjunct is one exactly when $v_j \cdot (y + T(u)) = 0$. Since $y + T(u) \ne 0$, this happens with probability $1/2$ independently for each j, so $\Pr[H(u,y) \ne f_T(u,y)] = 2^{-q}$. Every function of $2r$ bits, in particular $u \mapsto v_j \cdot T(u)$, is computed by a DNF of depth $2r + \lceil \log_2(2r) \rceil$, and the parity $v_j \cdot y$ of at most 2^r bits is computed by a De Morgan formula of depth $2 \log_2 2^r = 2r$. Equality adds two levels, and the final conjunction adds $\lceil \log_2 q \rceil$, so H has depth $2r + O(\log r + \log q)$.

Finally, we add a balancing bit. Consider $g(z,u,y) = z \oplus f_T(u,y)$ on $1 + 2r + 2^r = n$ input variables. For each (u,y), exactly one value of z makes g equal to one, so g is balanced. Substituting $z = 0$ into a formula for g gives a formula for f_T of no larger depth, so $\mathrm{D}(g) \ge \mathrm{D}(f_T) \ge 3r - O(\log r)$. The formulas $z \oplus H(u,y)$ have the same error probability as H and depth larger by two, which proves the claim. □

The next lemma turns such predictors into a shallow protocol on a dense subset of any core.

Lemma 6.10. *Let $g\colon \{0,1\}^n \to \{0,1\}$, let $m \ge 1$, and let $d \ge 0$. Suppose that there is a distribution over formulas h of depth at most d such that $\Pr_h[h(x) \ne g(x)] \le 1/(2m)$ for every $x \in \{0,1\}^n$. For every core $\mathcal{X}_0 \subseteq \{0,1\}^{m \times n}$ and every integer $0 \le \sigma \le m-1$, some nonempty $\mathcal{X} \subseteq \mathcal{X}_0$ has relative density at least $2^{-m+\sigma}$ in $\mathcal{X}_0$ and*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X} \times \mathcal{X}}) \le d + \sigma + O(\log(m+1)).$$

Proof. We first find a formula that is correct on all rows of many matrices of the core. Sample a formula h from the distribution. For each fixed matrix $X \in \mathcal{X}_0$, a union bound over its m rows shows that $h(X_i) = g(X_i)$ for all $i \in [m]$ with probability at least $1 - m/(2m) = 1/2$ (this does not require the rows to be distinct or independent). Hence the expected number of such matrices is at least $|\mathcal{X}_0|/2$, so we can fix a formula ψ of depth at most d and a set $\mathcal{Y} \subseteq \mathcal{X}_0$ with $|\mathcal{Y}| \ge |\mathcal{X}_0|/2$ such that $\psi(X_i) = g(X_i)$ for every $X \in \mathcal{Y}$ and every $i \in [m]$.

Next, we keep only the most frequent values of g^m. Split $\mathcal{Y}$ according to the 2^m values of g^m, including empty parts. Let $\mathcal{X}$ be the union of the $K = 2^{\sigma+1} \le 2^m$ largest parts. These parts contain at least a $K/2^m$ fraction of $\mathcal{Y}$, so

$$\frac{|\mathcal{X}|}{|\mathcal{X}_0|} \ge \frac{K}{2^m} \cdot \frac{|\mathcal{Y}|}{|\mathcal{X}_0|} \ge \frac{2^{\sigma+1}}{2^m} \cdot \frac{1}{2} = 2^{-m+\sigma}.$$

In particular, $\mathcal{X}$ is nonempty.

Finally, we give a protocol for $(U_m \diamond \mathrm{KW}_g)|_{\mathcal{X} \times \mathcal{X}}$. At most K distinct values of g^m appear in $\mathcal{X}$, so Alice can send $g^m(X)$ using $\log_2 K = \sigma + 1$ bits. Bob then either rejects if $g^m(X) = g^m(Y)$ or sends the index of a row i with $g(X_i) \ne g(Y_i)$, using $\lceil \log_2(m+1) \rceil$ bits. Since $X, Y \in \mathcal{X} \subseteq \mathcal{Y}$,

we have $\psi(X_i) = g(X_i) \neq g(Y_i) = \psi(Y_i)$, and both players know which of these two values is zero, since they know $g^m(X)$ and i. Walking down ψ from the root as in the KW correspondence, the players find a coordinate j with $X_{i,j} \neq Y_{i,j}$ using at most d bits and output the cell (i, j). In total, the protocol uses at most $\sigma + 1 + \lceil \log_2(m+1) \rceil + d \leq d + \sigma + O(\log(m+1))$ bits. □

Theorem 1.8. *For every sufficiently large even m, there is a balanced Boolean function g_m on $n = 2^{\Theta(m)}$ bits such that, for every nonempty core $\mathcal{X}_0 \subseteq \{0,1\}^{m \times n}$ and every integer $0 \leq \sigma \leq m-1$, some $\mathcal{X} \subseteq \mathcal{X}_0$ satisfies $|\mathcal{X}| \geq 2^{-m+\sigma}|\mathcal{X}_0|$, but*

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_{g_m})|_{\mathcal{X}\times\mathcal{X}}) \leq \mathrm{D}(g_m) - m/2 + \sigma + O(\log m).$$

Proof. Set $r = m/2$ and consider the balanced function g_m from Lemma 6.9, so that $n = 2^{m/2} + m + 1$ and $\mathrm{D}(g_m) \geq 3m/2 - O(\log m)$. With $q = \lceil \log_2(2m) \rceil$, its predictors have error at most $1/(2m)$ and depth $m + O(\log m)$. Lemma 6.10 gives a set $\mathcal{X}$ of relative density at least $2^{-m+\sigma}$ in $\mathcal{X}_0$ with

$$\mathrm{CC}((U_m \diamond \mathrm{KW}_{g_m})|_{\mathcal{X}\times\mathcal{X}}) \leq m + \sigma + O(\log m) \leq \mathrm{D}(g_m) - m/2 + \sigma + O(\log m).$$ □

In particular, let $s(m), \ell(m) \geq 0$ be any functions that are $o(m)$. Taking $\sigma = \lceil s(m) \rceil$ in Theorem 1.8 shows that, for all sufficiently large even m, no core $\mathcal{X}_0$ guarantees $\mathrm{CC}((U_m \diamond \mathrm{KW}_{g_m})|_{\mathcal{X}\times\mathcal{X}}) \geq \mathrm{D}(g_m) - \ell(m)$ for all $\mathcal{X} \subseteq \mathcal{X}_0$ of relative density at least $2^{-m+s(m)}$, because $2^{-m+\lceil s(m) \rceil} \geq 2^{-m+s(m)}$ and $m/2 - \lceil s(m) \rceil - O(\log m) > \ell(m)$.

Tightness of Theorem 1.7. The same construction shows that the logarithmic loss in Theorem 1.7 is necessary. Let $r \geq 2$ and $m \geq 1$ be integers, take $q = \lceil \log_2(2m) \rceil$ in Lemma 6.9, and apply Lemma 6.10 with $\sigma = m - 1$. For the resulting function g on $n = 2^r + 2r + 1$ bits, every core $\mathcal{X}_0 \subseteq \{0,1\}^{m \times n}$ contains a subset $\mathcal{X}$ of relative density at least $1/2$ with

$$\mathrm{D}(g) - \mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq r - m - O(\log r + \log m).$$

Adding dummy variables changes neither $\mathrm{D}(g)$ nor the predictors. Hence, for every $n \geq 9$, we may take the largest r with $2^r + 2r + 1 \leq n$, which satisfies $n < 2^{r+1} + 2(r+1) + 1 \leq 2^{r+2}$, that is, $r > \log_2 n - 2$. This gives a function on n bits whose loss at density $1/2$ is at least $\log_2 n - m - O(\log\log n + \log m)$. For $m \leq (\log_2 n)/2$, this loss is $\Omega(\log n)$ and exceeds $0.99m$ for all sufficiently large n. Since $(1+2^{1-m})/3 \leq 1/2$ for $m \geq 2$, this applies in particular to the core S^m of Theorem 1.7, so its $O(\log n)$ loss is tight up to a constant factor.

This also answers the question of Dinur and Meir [DM16] whether every g has a core in which $\mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq \mathrm{D}(g) - 0.99m$ for every subset $\mathcal{X}$ of relative density $1/2$. When m is at least a large constant times $\log n$, the answer is positive by Theorem 1.7. When $m \leq (\log_2 n)/2$ and n is sufficiently large, the answer is negative, since the bound fails in every core for the function g above.

7 Open Problems

We conclude with the questions that we find most interesting.

Direct sum. For a total $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $k \geq 1$, let F^k be the function that receives k independent input pairs and outputs all k values of F.

Question 7.1. Is there an absolute constant $c > 0$ such that

$$\mathrm{CC}(F^k) \geq c\,k\ \mathrm{CC}(F)$$

for every total F and every k?

For arbitrary total F, the amortized bounds of Feder, Kushilevitz, Naor, and Nisan [FKNN95] lose a square root. Mackenzie and Saffidine [MS25] refuted the strongest form $\mathrm{CC}(F^k) \geq k(\mathrm{CC}(F) - O(1))$, but their counterexample does not rule out $\mathrm{CC}(F^k) \geq c\,k\ \mathrm{CC}(F)$.

Choice and agreement for exponentially many instances. Our bounds lose $O(\log k)$, so the question remains open for exponentially many instances.

Question 7.2. Is it true that

$$\mathrm{CC}(\mathrm{choose}_F^k) \geq \mathrm{CC}(F) - \mathrm{polylog}(n) \qquad \text{and} \qquad \mathrm{D}(\mathrm{agree}_f^k) \geq \mathrm{D}(f) - \mathrm{polylog}(n)$$

for every total $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with $1 \leq k \leq 2^n$ and every nonconstant total $f\colon \{0,1\}^n \to \{0,1\}$ with $1 \leq k \leq \min\{|f^{-1}(0)|, |f^{-1}(1)|\}$, respectively?

These are Conjectures 9.7 and 9.6 of Dinur and Meir [DM16]; the agreement question for partial functions is even stronger. Theorem 1.1 and Corollary 1.3 settle both conjectures for $k \leq 2^{\mathrm{polylog}(n)}$, and Corollary 1.4 preserves a constant fraction of the complexity even for exponentially many instances, up to an additive $O(\log n)$ term.

A corrected dense-core conjecture. Theorem 1.8 shows that the uniform formulation of Conjecture 9.4 is false, which raises the question of how to repair it.

Question 7.3. Does an explicit $O(\log n)$ loss suffice to repair the uniform formulation of Conjecture 9.4? More precisely, are there absolute constants $C, c > 0$ such that, for every m, n and nonconstant g, some nonempty core $\mathcal{X}_0$ satisfies

$$\frac{|\mathcal{X}|}{|\mathcal{X}_0|} \geq 2^{-m + C\sqrt{m}(\log(m+2))^c} \quad \Longrightarrow \quad \mathrm{CC}((U_m \diamond \mathrm{KW}_g)|_{\mathcal{X}\times\mathcal{X}}) \geq \mathrm{D}(g) - C\left(\sqrt{m}(\log(m+2))^c + \log n\right)$$

for every $\mathcal{X} \subseteq \mathcal{X}_0$ whenever the density threshold is at most one?

Theorem 6.6 establishes this statement for almost every g: it gives a stronger bound for $m \leq n^2$, and the requested lower bound is nonpositive for $m > n^2$. The unresolved case is an arbitrary inner function. Theorem 1.7 gives hard cores for every g at constant relative density, while Theorem 1.8 shows why the dependence on n cannot simply be omitted.

From choice to ordinary composition. Our composition results concern the universal outer relation, and they do not give the KRW conjecture for Boolean outer functions.

Question 7.4. How can lower bounds for the 1-out-of-k KW game or for dense restrictions be used to prove lower bounds for ordinary composition $f \diamond g$?

In $\mathrm{KW}_{f\diamond g}$, the players may output a differing cell in a row where the two inner values $g(X_i)$ and $g(Y_i)$ are equal, whereas in the 1-out-of-k game of KW_g every pair of corresponding rows has opposite g-values. Moreover, the proof of Theorem 1.6 for $U_m \diamond \mathrm{KW}_g$ tracks the protocol on diagonal pairs (X, X), which do not appear in $\mathrm{KW}_{f\diamond g}$ because Alice's and Bob's inputs have opposite outer values $f(g^m(A)) \neq f(g^m(B))$.

AI Methodology Statement

We follow the recommendations of the working group report [AIT26] and use its template for this section.

Activity	**AI used**
Asking the research question	No
Coming up with the approach	Yes
Proof development	Yes
Writing and exposition	Yes
Checking for bugs	Yes
Other supporting tasks	Yes

Details (elaborating on the points above). We used OpenAI ChatGPT and Codex to develop candidate proofs, check sources and drafting the text. We used Anthropic Claude Opus 5.5 and Google Antigravity to restructure the paper and rewrite its exposition, including the abstract and the introduction. The authors are responsible for all statements.

References

[ABG+01] Andris Ambainis, Harry Buhrman, William I. Gasarch, Bala Kalyanasundaram, and Leen Torenvliet. The Communication Complexity of Enumeration, Elimination, and Selection. *Journal of Computer and System Sciences*, 63(2):148–185, 2001.

[AIT26] AI and TCS: The Next Six Months. Working Group Report. Simons Institute for the Theory of Computing, https://simons.berkeley.edu/AITCSreport, September 2026.

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

[BBKW10] Amos Beimel, Sebastian Ben Daniel, Eyal Kushilevitz, and Enav Weinreb. Choosing, Agreeing, and Eliminating in Communication Complexity. In *37th International Colloquium on Automata, Languages and Programming (ICALP 2010)*, volume 6198 of *Lecture Notes in Computer Science*, pages 451–462, 2010.

[BBKW14] Amos Beimel, Sebastian Ben Daniel, Eyal Kushilevitz, and Enav Weinreb. Choosing, Agreeing, and Eliminating in Communication Complexity. *Computational Complexity*, 23(1):1–42, 2014.

[BDG+20] Andrew Bassilakis, Andrew Drucker, Mika Göös, Lunjia Hu, Weiyun Ma, and Li-Yang Tan. The Power of Many Samples in Query Complexity. In *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *LIPIcs*, pages 9:1–9:18, 2020.

[BKS95] Richard Beigel, Martin Kummer, and Frank Stephan. Approximable Sets. *Information and Computation*, 120(2):304–314, 1995.

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

[CDK+17] Arkadev Chattopadhyay, Pavel Dvořák, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Lower Bounds for Elimination via Weak Regularity. In *34th Symposium on Theoretical Aspects of Computer Science (STACS 2017)*, volume 66 of *LIPIcs*, pages 21:1–21:14, 2017.

[CKM25] Nikolai Chukhin, Alexander S. Kulikov, and Ivan Mihajlin. Toward Better Depth Lower Bounds: Strong Composition of XOR and a Random Function. In *42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025)*, volume 327 of *LIPIcs*, pages 26:1–26:15, 2025.

[DM16] Irit Dinur and Or Meir. Toward the KRW Composition Conjecture: Cubic Formula Lower Bounds via Communication Complexity. In *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *LIPIcs*, pages 3:1–3:51, 2016.

[dRMN+24] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, and Robert Robere. KRW Composition Theorems via Lifting. *Computational Complexity*, 33(1):4, 2024.

[EIRS01] Jeff Edmonds, Russell Impagliazzo, Steven Rudich, and Jirí Sgall. Communication Complexity Towards Lower Bounds on Circuit Depth. *Computational Complexity*, 10(3):210–246, 2001.

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized Communication Complexity. *SIAM Journal on Computing*, 24(4):736–750, 1995.

[GKR16] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of Information and Communication for Boolean Functions. *Journal of the ACM*, 63(5):46:1–46:31, 2016.

[GMWW17] Dmitry Gavinsky, Or Meir, Omri Weinstein, and Avi Wigderson. Toward Better Formula Lower Bounds: The Composition of a Function and a Universal Relation. *SIAM Journal on Computing*, 46(1):114–131, 2017.

[Hås98] Johan Håstad. The Shrinkage Exponent of de Morgan Formulas is 2. *SIAM Journal on Computing*, 27(1):48–64, 1998.

[HW93] Johan Håstad and Avi Wigderson. Composition of the Universal Relation. In *Advances in Computational Complexity Theory, Proceedings of a DIMACS Workshop*, volume 13 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 119–134, 1993.

[Imp95] Russell Impagliazzo. Hard-Core Distributions for Somewhat Hard Problems. In *36th Annual Symposium on Foundations of Computer Science (FOCS 1995)*, pages 538–545, 1995.

[JKS10] Rahul Jain, Hartmut Klauck, and Miklos Santha. Optimal Direct Sum Results for Deterministic and Randomized Decision Tree Complexity. *Information Processing Letters*, 110(20):893–897, 2010.

[Joc68] Carl G. Jockusch, Jr. Semirecursive Sets and Positive Reducibility. *Transactions of the American Mathematical Society*, 131(2):420–436, 1968.

[KKN95] Mauricio Karchmer, Eyal Kushilevitz, and Noam Nisan. Fractional Covers and Communication Complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.

[KM18] Sajin Koroth and Or Meir. Improved Composition Theorems for Functions and Relations. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *LIPIcs*, pages 48:1–48:18, 2018.

[Ko83] Ker-I Ko. On Self-Reducibility and Weak P-Selectivity. *Journal of Computer and System Sciences*, 26(2):209–221, 1983.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-Logarithmic Depth Lower Bounds Via the Direct Sum in Communication Complexity. *Computational Complexity*, 5(3/4):191–204, 1995.

[KW90] Mauricio Karchmer and Avi Wigderson. Monotone Circuits for Connectivity Require Super-Logarithmic Depth. *SIAM Journal on Discrete Mathematics*, 3(2):255–265, 1990.

[Mei18] Or Meir. The Direct Sum of Universal Relations. *Information Processing Letters*, 136:105–111, 2018.

[Mei25] Or Meir. Toward Better Depth Lower Bounds: A KRW-like Theorem For Strong Composition. *SIAM Journal on Computing*, 54(5):1193–1240, 2025.

[MS21] Ivan Mihajlin and Alexander Smal. Toward Better Depth Lower Bounds: The XOR-KRW Conjecture. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *LIPIcs*, pages 38:1–38:24, 2021.

[MS25] Simon Mackenzie and Abdallah Saffidine. Refuting the Direct Sum Conjecture for Total Functions in Deterministic Communication Complexity. In *57th Annual ACM Symposium on Theory of Computing (STOC 2025)*, pages 572–583, 2025. Full version: https://arxiv.org/abs/2411.19003v2.

[MT18] Or Meir and Avishay Tal. The Choice and Agreement Problems of a Random Function. *Information Processing Letters*, 133:16–20, 2018.

[Sel79] Alan L. Selman. P-Selective Sets, Tally Languages, and the Behavior of Polynomial Time Reducibilities on NP. *Mathematical Systems Theory*, 13:55–65, 1979.

[Shi19] Yaroslav Shitov. Counterexamples to Strassen's Direct Sum Conjecture. *Acta Mathematica*, 222(2):363–379, 2019.

[Tal14] Avishay Tal. Shrinkage of De Morgan Formulae by Spectral Techniques. In *55th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2014)*, pages 551–560, 2014.

[Uhl74] Dietmar Uhlig. On the Synthesis of Self-Correcting Schemes from Functional Elements with a Small Number of Reliable Elements. *Mathematical Notes*, 15(6):558–562, 1974.

[Val84] Leslie G. Valiant. Short Monotone Formulae for the Majority Function. *Journal of Algorithms*, 5(3):363–366, 1984.

[vN28] John von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

[Weg87] Ingo Wegener. *The Complexity of Boolean Functions*. Wiley-Teubner, 1987.

[Wu23] Hao Wu. An Improved Composition Theorem of a Universal Relation and Most Functions via Effective Restriction. *CoRR*, abs/2310.07422, 2023. Version 2, November 2023.

[Zer13] Ernst Zermelo. Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the Fifth International Congress of Mathematicians*, volume 2, pages 501–504. Cambridge University Press, 1913.