THEORY OF COMPUTING, Volume 19 (7), 2023, pp. 1–51
www.theoryofcomputing.org

Shrinkage under Random Projections, and Cubic Formula Lower Bounds for $\boldsymbol{AC}^0$

Yuval Filmus* Or Meir† Avishay Tal

Received February 19, 2021; Revised December 28, 2021; Published December 5, 2023

Abstract. Håstad showed that any De Morgan formula (composed of AND, OR and NOT gates) shrinks by a factor of $\tilde{O}(p^2)$ under a random restriction that leaves each variable alive independently with probability p [SICOMP, 1998]. Using this result, he gave an $\widetilde{\Omega}(n^3)$ formula size lower bound for the Andreev function, which, up to lower order improvements, remains the state-of-the-art lower bound for any explicit function.

In this paper, we extend the shrinkage result of Håstad to hold under a far wider family of random restrictions and their generalization — random projections. Based on our shrinkage results, we obtain an $\widetilde{\Omega}(n^3)$ formula size lower bound for an explicit function computable in $\boldsymbol{AC}^0$. This improves upon the best known formula size lower bounds for $\boldsymbol{AC}^0$, that were only quadratic prior to our work. In addition, we prove that the KRW conjecture [Karchmer et al., Computational Complexity 5(3/4), 1995] holds for inner functions for which the unweighted quantum adversary bound is tight. In particular, this holds for inner functions with a tight Khrapchenko bound.

Our random projections are tailor-made to the function's structure so that the function maintains structure even under projection — using such projections is necessary, as standard random restrictions simplify $\boldsymbol{AC}^0$ circuits. In contrast, we show that any De Morgan formula shrinks by a quadratic factor under our random projections, allowing us to prove the cubic lower bound.

An extended abstract of this article appeared in the Proceeding of the 12th Innovations in Theoretical Computer Science Conference (ITCS'21).

*Taub Fellow — supported by the Taub Foundations. The research was funded by ISF grant 1337/16.

†Partially supported by the Israel Science Foundation (grant No. 1445/16).

 DOI: 10.4086/toc.2023.v019a007

Our proof techniques build on Håstad's proof for the simpler case of balanced formulas. This allows for a significantly simpler proof at the cost of slightly worse parameters. As such, when specialized to the case of p-random restrictions, our proof can be used as an exposition of Håstad's result.

ACM Classification: F.1.1

AMS Classification: 68Q06

Key words and phrases: circuit complexity, lower bounds, shrinkage, formulas, formula complexity, KRW conjecture

1 Introduction

1.1 Background

Is there an efficiently solvable computational task that does not admit an efficient parallel algorithm? A formalization of this question asks, is $\mathbf{P} \not\subseteq \mathbf{NC}^1$? The answer is still unknown. The question can be rephrased as follows (ignoring issues of uniformity): is there a Boolean function in $\mathbf{P}$ that does not have a (De Morgan) formula of polynomial size? (We define De Morgan formulas formally in Section 2; informally, these are formulas composed of AND, OR and NOT gates.)

The history of formula lower bounds for functions in $\mathbf{P}$ goes back to the 1960s, with the seminal result of Subbotovskaya [55] that introduced the technique of random restrictions. Subbotovskaya showed that the Parity function on n variables requires formulas of size at least $\Omega(n^{1.5})$. Khrapchenko [33], using a different proof technique, showed that in fact the Parity function on n variables requires formulas of size $\Theta(n^2)$. Later, Andreev [4] came up with a new explicit function (now known as the Andreev function) for which he was able to obtain an $\Omega(n^{2.5})$ size lower bound. This lower bound was subsequently improved in [29, 44, 23, 56] to $n^{3-o(1)}$.

The line of work initiated by Subbotovskaya and Andreev relies on the *shrinkage* of formulas under p-random restrictions. A p-random restriction is a randomly chosen partial assignment to the inputs of a function. Set a parameter $p \in (0,1)$. We fix each variable independently with probability $1-p$ to a uniformly random bit, and we keep the variable alive with probability p. Under such a restriction, formulas shrink (in expectation) by a factor more significant than p. Subbotovskaya showed that formulas shrink to at most $p^{1.5}$ times their original size, whereas subsequent work [44, 29] improved the bound to $p^{1.55}$ and then to $p^{1.63}$. Finally, Håstad [23] showed that the shrinkage exponent of formulas is 2, or in other words, that formulas shrink by a factor of $p^{2-o(1)}$ under p-random restrictions. Tal [56] improved the shrinkage factor to $O(p^2)$ — obtaining a tight result, as exhibited by the Parity function.

In a nutshell, shrinkage results are useful for proving lower bounds as long as the explicit function being analyzed maintains structure under such restrictions and does not trivialize. For example, the Parity function does not become constant as long as at least one variable remains alive. Thus any formula F that computes Parity must be of at least quadratic size, or else the formula F under restriction, keeping each variable alive with probability $100/n$, would likely become a constant function, whereas Parity would not. Andreev's idea is similar, though he manages to construct a function such that under a random restriction keeping only $\Theta(\log n)$ of the variables, the formula size should be at least $\tilde{\Omega}(n)$ (in expectation). This ultimately gives the nearly cubic lower bound.

1.1.1 The KRW Conjecture

Despite much effort, proving $\mathbf{P} \not\subseteq \mathbf{NC}^1$, and even just breaking the cubic barrier in formula lower bounds, have remained a challenge for more than two decades. An approach to solve the $\mathbf{P}$ versus $\mathbf{NC}^1$ problem was suggested by Karchmer, Raz and Wigderson [31]. They conjectured that when composing two Boolean functions, f and g, the formula size of the resulting function, $f \diamond g$, is (roughly) the product of the formula sizes of f and g.[1] We will refer to this conjecture as the "KRW conjecture". Under the KRW conjecture (and even under weaker variants of it), [31] constructed a function in $\mathbf{P}$ with no polynomial-size formulas. It remains a major open challenge to settle the KRW conjecture.

A few special cases of the KRW conjecture are known to be true. The conjecture holds when either f or g is the AND or the OR function. Håstad's result [23] and its improvement [56] show that the conjecture holds when the inner function g is the Parity function and the outer function f is any function. This gives an alternative explanation to the $n^{3-o(1)}$ lower bound for the Andreev function. Indeed, the Andreev function is at least as hard as the composition of a maximally hard function f on $\log n$ bits and $g = \mathsf{Parity}_{n/\log n}$, where the formula size of f is $\tilde{\Omega}(n)$ and the formula size of $\mathsf{Parity}_{n/\log n}$ is $\Theta(n^2/\log^2 n)$. Since the KRW conjecture holds for this special case, the formula size of the Andreev function is at least $\tilde{\Omega}(n^3)$. In other words, the state-of-the-art formula size lower bounds for explicit functions follow from a special case of the KRW conjecture — the case in which g is the Parity function. Moreover, this special case follows from the shrinkage offormulas under p-random restrictions.

1.1.2 Bottom-up versus top-down techniques

Whereas random restrictions are a "bottom-up" proof technique [24], a different line of work suggested a "top-down" approach using the language of communication complexity. The connection between formula size and communication complexity was introduced in the seminal paper by Karchmer and Wigderson [32]. They defined for any Boolean function f a two-party communication problem KW_f: Alice gets an input x such that $f(x) = 1$, and Bob gets an input y such that $f(y) = 0$. Their goal is to identify a coordinate i on which $x_i \neq y_i$, while minimizing their communication. It turns out that there is a one-to-one correspondence between any protocol tree solving KW_f and any formula computing the function f. Since protocols naturally traverse the tree from root to leaf, proving lower bounds on their size or depth is done usually in a top-down fashion. This framework has proven to be very useful in proving formula lower

[1] More precisely, the original KRW conjecture [31] concerns depth complexity rather than formula complexity. The variant of the conjecture for formula complexity, which is discussed above, was posed in [19].

bounds in the monotone setting (see, e. g., [32, 20, 47, 31, 46, 21, 45]) and in studying the KRW conjecture (see, e. g., [31, 14, 26, 19, 13, 34, 41, 12, 42]). Moreover, in a recent paper [13], Dinur and Meir were able to reprove Håstad's cubic lower bound using the framework of Karchmer and Wigderson. As Dinur and Meir's proof showed that top-down techniques can replicate Håstad's cubic lower bound, a natural question (which motivated this project) arose:

Are top-down techniques superior to bottom-up techniques?

Towards that, we focused on a candidate problem: prove a cubic lower bound for an explicit function in $\boldsymbol{AC}^0$.[2] Based on the work of Dinur and Meir [13], we suspected that such a lower bound could be achieved using top-down techniques. We were also *certain* that the problem cannot be solved using the random restriction technique. Indeed, in order to prove a lower bound on a function f using random restrictions, one should argue that f remains hard under a random restriction, however, it is well-known that functions in $\boldsymbol{AC}^0$ trivialize under p-random restrictions [1, 16, 57, 22]. Based on this intuition, surely random restrictions cannot show that a function in $\boldsymbol{AC}^0$ requires cubic size. Our intuition turned out to be false.

1.2 Our results

In this article, we construct an explicit function in $\boldsymbol{AC}^0$ which requiresformulas of size $n^{3-o(1)}$. Surprisingly, our proof is conducted via the bottom-up technique of random projections, which is a generalization of random restrictions (more details below).

Theorem 1.1. *There exists a family of Boolean functions $h_n\colon \{0,1\}^n \to \{0,1\}$ for $n \in \mathbb{N}$ such that*

1. *h_n can be computed by uniform depth-4 unbounded fan-in formulas of size $O(n^3)$.*
2. *The formula size of h_n is at least $n^{3-o(1)}$.*

Prior to our work, the best formula size lower bounds for an explicit function in $\boldsymbol{AC}^0$ were only quadratic [43, 10, 30, 6].

Our hard function is a variant of the Andreev function. More specifically, recall that the Andreev function is based on the composition $f \diamond g$, where f is a maximally hard function and g is the Parity function. Since Parity is not in $\boldsymbol{AC}^0$, we cannot take g to be the Parity function in our construction. Instead, our hard function is obtained by replacing the Parity function with the Surjectivity function of Beame and Machmouchi [6].

As in the case of the Andreev function, we establish the hardness of our function by proving an appropriate special case of the KRW conjecture. To this end, we introduce a generalization of the unweighted adversary method [2], called the *soft-adversary bound*, which we denote $\mathrm{Adv}_s(g)$ (see Section 7.1). We prove the KRW conjecture for the special case in which the outer function f is any function, and g is a function whose formula complexity is bounded tightly by the soft-adversary bound. We then obtain Theorem 1.1 by applying this version of the KRW conjecture to the case where g is the Surjectivity function. We note that our KRW result holds in particular

[2] Recall that $\boldsymbol{AC}^0$ is the class of functions computed by constant depth polynomial size circuits composed of AND and OR gates of unbounded fan-in, with variables or their negation at the leaves.

for functions g with a large Khrapchenko bound, which in turn implies the known lower bounds in the cases where g is the Parity function [23] and the Majority function [17].

Our proof of the special case of the KRW conjecture follows the methodology of Håstad [23], who proved the special case in which g is Parity on m variables. Håstad proved that formulas shrink by a factor of (roughly) p^2 under p-random restrictions. Choosing $p = 1/m$ shrinks a formula for $f \diamond g$ by a factor of roughly m^2, which coincides with the formula complexity of g. On the other hand, on average each copy of g simplifies to a single input variable, and so $f \diamond g$ simplifies to f. This shows that $L(f \diamond g) \gtrsim L(f) \cdot L(g)$.

Our main technical contribution is a new shrinkage theorem that works in a far wider range of scenarios than just p-random restrictions: A *random projection* is a generalization of a random restriction in which each of the input variables $x_1, \ldots, x_n$ may either be fixed to a constant or be replaced with a literal from the set $y_1, \ldots, y_m, \overline{y_1}, \ldots, \overline{y_m}$ where $y_1, \ldots, y_m$ are new variables. Given a function g with soft-adversary bound $\mathrm{Adv}_s(g)$, we construct a random projection which, on the one hand, shrinks formulas by a factor of $\mathrm{Adv}_s(g)$, and on the other hand, simplifies $f \diamond g$ to f. We thus show that $L(f \diamond g) \gtrsim L(f) \cdot \mathrm{Adv}_s(g)$, and in particular, if $\mathrm{Adv}_s(g) \approx L(g)$, then $L(f \diamond g) \gtrsim L(f) \cdot L(g)$, just as in Håstad's proof. Our random projections are tailored specifically to the structure of the function $f \diamond g$, ensuring that $f \diamond g$ simplifies to f under projection. This enables us to overcome the aforementioned difficulty. In contrast, p-random restrictions that do not respect the structure of $f \diamond g$ would likely result in a restricted function that is much simpler than f and in fact would be a constant function with high probability.

Our shrinkage theorem applies more generally to two types of random projections, which we call *fixing projections* and *hiding projections*. Roughly speaking, fixing projections are random projections in which substituting a constant for one of the variables $y_1, \ldots, y_m$ results in a projection that is much more probable. Hiding projections are random projections in which substituting a constant for a variable y_j hides which of the input variables $x_1, \ldots, x_n$ were mapped to $\{y_j, \overline{y}_j\}$ before the substitution. We note that our shrinkage theorem for fixing projections captures Håstad's result for p-random restrictions as a special case.

The proof of our shrinkage theorem is based on Håstad's proof [23], but also simplifies it. In particular, we take the simpler argument that Håstad uses for the special case of completely balanced trees, and adapt it to the general case. As such, our proof avoids a complicated case analysis, at the cost of slightly worse bounds. Using our bounds, it is nevertheless easy to obtain the $n^{3-o(1)}$ lower bound for the Andreev function. Therefore, one can see the specialization of our shrinkage result to p-random restrictions as an exposition of Håstad's cubic lower bound.

1.2.1 An example: our techniques when specialized to $f \diamond \mathsf{Majority}_m$

To illustrate our choice of random projections, we present its instantiation to the special case of $f \diamond g$, where $f\colon \{0,1\}^k \to \{0,1\}$ is non-constant and $g = \mathsf{Majority}_m$ for some odd integer m. In this case, the input variables to $f \diamond g$ are composed of k disjoint blocks, $B_1, \ldots, B_k$, each containing m variables. We use the random projection that for each block $B_i = \{x_{m(i-1)+1}, \ldots, x_{mi}\}$, picks one variable in the block B_i uniformly at random, projects this variable to the new variable y_i, and fixes the rest of the variables in the block in a balanced way so that the number of zeros and ones in the block is equal (i. e., we have exactly $(m-1)/2$ zeros and $(m-1)/2$ ones). It is not hard to see that under this choice, $f \diamond g$ simplifies to f. On the other hand, we show that this choice of random projections shrinks the formula complexity by a factor of $\approx 1/m^2$. Combining the two together, we get that $L(f \diamond \mathsf{Majority}_m) \gtrsim L(f) \cdot m^2$. Note that in this distribution of random projections, the different coordinates are not independent of one another, and this feature allows us to maintain structure.

1.3 Related work

Our technique of using tailor-made random projections was inspired by the celebrated result of Rossman, Servedio, and Tan [25] that proved an average-case depth hierarchy. In fact, the idea to use tailor-made random restrictions goes back to Håstad's thesis [27, Chapter 6.2]. Similar to our case, in [27, 25], p-random restrictions are too crude to separate depth d from depth $d+1$ circuits. Given a circuit C of depth $d+1$, the main challenge is to construct a distribution of random restrictions or projections (tailored to the circuit C) that on the one hand maintains structure for C, but on the other hand simplify any depth d circuit C'.

1.4 Paper outline

The paper starts with brief preliminaries in Section 2. Then, in Section 3, we define the notions of fixing and hiding projections, state the corresponding shrinkage theorems, and sketch how they can be used to prove a cubic formula lower bound for $\boldsymbol{AC}^0$. We prove the shrinkage theorems for fixing and hiding projections in Sections 4 and 5, respectively. In Section 6 we provide a brief interlude on the join operation for projections. The quantum adversary bound, the Khrapchenko bound, and their relation to hiding projections are discussed in Section 7. Finally, Section 8 contains a proof of Theorem 1.1, as a corollary of a more general result which is a special case of the KRW conjecture. In the same section we also rederive the cubic lower bound on Andreev's function, and the cubic lower bound on the Majority-based variant considered in [17].

2 Preliminaries

Throughout the paper, we use bold letters to denote random variables. For any $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1, \ldots, n\}$. Given a bit $\sigma \in \{0, 1\}$, we denote its negation by $\overline{\sigma}$. We assume familiarity with the basic definitions of communication complexity (see, e. g., [37]). All logarithms in this paper are base 2.

Definition 2.1. A *(De Morgan) formula (with bounded fan-in)* is a binary tree, whose leaves are labeled with literals from the set $\{x_1, \overline{x}_1, \ldots, x_n, \overline{x}_n\}$, and whose internal vertices are labeled as AND ($\wedge$) or OR ($\vee$) gates. The *size* of a formula ϕ, denoted size(ϕ), is the number of leaves in the tree. The *depth* of the formula is the depth of the tree. A *formula with unbounded fan-in* is defined similarly, but every internal vertex in the tree can have any number of children. Unless stated explicitly otherwise, whenever we say "formula" we refer to a formula with bounded fan-in.

Definition 2.2. A formula ϕ computes a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ in the natural way. The *formula complexity* of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$, denoted $L(f)$, is the size of the smallest formula that computes f. The *depth complexity* of f, denoted $D(f)$, is the smallest depth of a formula that computes f. For convenience, we define the size and depth of the constant functions to be zero.

A basic property of formula complexity is that it is subadditive:

Fact 2.3. *For every two functions $f_1, f_2\colon \{0,1\}^n \to \{0,1\}$ we have $L(f_1 \wedge f_2) \le L(f_1) + L(f_2)$ and $L(f_1 \vee f_2) \le L(f_1) + L(f_2)$.*

The following theorem shows that every small formula can be "balanced" to obtain a shallow formula.

Theorem 2.4 (Formula balancing, [7], following [54, 8])**.** *For every $s \in \mathbb{N}$ and $\alpha > 0$, the following holds: For every formula ϕ of size s, there exists an equivalent formula ϕ' of depth at most $O(2^{\frac{1}{\alpha}} \cdot \log s)$ and size at most $s^{1+\alpha}$ (where the constant inside the big-O notation does not depend on the choice of α).*

Notation 2.5. With a slight abuse of notation, we will often identify a formula ϕ with the function it computes. In particular, the notation $L(\phi)$ denotes *the formula complexity of the function* computed by ϕ, and *not the size of* ϕ (which is denoted by $\operatorname{size}(\phi)$).

Notation 2.6. Given a Boolean variable z, we denote by z^0 and z^1 the literals z and $\overline{z}$, respectively. In other words, $z^b = z \oplus b$.

Notation 2.7. Given a literal ℓ, we define $\mathsf{var}(\ell)$ to be the underlying variable, that is, $\mathsf{var}(z) = \mathsf{var}(\overline{z}) = z$.

Notation 2.8. Let Π be a deterministic communication protocol that takes inputs from $\mathcal{A} \times \mathcal{B}$, and recall that the leaves of the protocol induce a partition of $\mathcal{A} \times \mathcal{B}$ to combinatorial rectangles. For every leaf ℓ of Π, we denote by $\mathcal{A}_\ell \times \mathcal{B}_\ell$ the combinatorial rectangle that is associated with ℓ.

We use the framework of Karchmer–Wigerson relations [32], which relates the complexity of f to the complexity of a related communication problem KW_f.

Definition 2.9 ([32])**.** Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function. The *Karchmer–Wigderson relation* of f, denoted KW_f, is the following communication problem: The inputs of Alice and Bob are strings $a \in f^{-1}(1)$ and $b \in f^{-1}(0)$, respectively, and their goal is to find a coordinate $i \in [n]$ such that $a_i \neq b_i$. Note that such a coordinate must exist since $f^{-1}(1) \cap f^{-1}(0) = \emptyset$ and hence $a \neq b$.

Theorem 2.10 ([32], see also [48])**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$. The communication complexity of KW_f is equal to $D(f)$, and the minimal number of leaves in a protocol that solves KW_f is $L(f)$.*

We use the following two standard inequalities.

Fact 2.11 (the AM–GM inequality)**.** *If $x, y \ge 0$ then $\sqrt{x \cdot y} \le \frac{x+y}{2}$.*

Fact 2.12 (special case of Cauchy-Schwarz inequality)**.** *For every t non-negative real numbers $x_1, \dots, x_t$ we have $\sqrt{x_1} + \ldots + \sqrt{x_t} \le \sqrt{t} \cdot \sqrt{x_1 + \ldots + x_t}$.*

3 Main results

In this section we define the notions of fixing and hiding projections and state their associated shrinkage theorems (see Sections 3.1 and 3.2, respectively). We then sketch the proof of our cubic formula lower bounds for $\boldsymbol{AC}^0$ assuming those theorems (see Section 3.3). We start by defining projections and the relevant notation.

Definition 3.1. Let $x_1, \dots, x_n$ and $y_1, \dots, y_m$ be Boolean variables. A *projection π from $x_1, \dots, x_n$ to $y_1, \dots, y_m$* is a function from the set $\{x_1, \dots, x_n\}$ to the set $\{0, 1, y_1, \overline{y_1}, \dots, y_m, \overline{y_m}\}$. Given such a projection π and a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ over the variables $x_1, \dots, x_n$, we denote by $f|_\pi\colon \{0,1\}^m \to \{0,1\}$ the function obtained from f by substituting $\pi(x_i)$ for x_i for each $i \in [n]$ in the natural way. Unless stated explicitly otherwise, all projections in this section are from $x_1, \dots, x_n$ to $y_1, \dots, y_m$, and all functions from $\{0,1\}^n$ to $\{0,1\}$ are over the variables $x_1, \dots, x_n$. A *random projection* is a distribution over projections.

Notation 3.2. Let π be a projection. For every $j \in [m]$ and bit $\sigma \in \{0,1\}$, we denote by $\pi_{y_j \leftarrow \sigma}$ the projection that is obtained from π by substituting σ for y_j.

Notation 3.3. With a slight abuse of notation, if a projection π maps all the variables $x_1, \dots, x_n$ to constants in $\{0,1\}$, we will sometimes treat it as a binary string in $\{0,1\}^n$.

3.1 Fixing projections

Intuitively, a q-fixing projection is a random projection $\boldsymbol{\pi}$ in which for every variable x_i, the probability that $\boldsymbol{\pi}$ maps a variable x_i to a literal y_j^σ is much smaller than the probability that $\boldsymbol{\pi}$ *fixes* y_j to a constant, *regardless of the values that π assigns to the other variables*. This property is essentially the minimal property that is required in order to carry out the argument of Håstad [23]. Formally, we define q-fixing projections as follows.

Definition 3.4. Let $0 \le q_0, q_1 \le 1$. We say that a random projection $\boldsymbol{\pi}$ is a *(q_0, q_1)-fixing projection* if for every projection π, every bit $\sigma \in \{0,1\}$, and every variable x_i, we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i)) \leftarrow \sigma} = \pi\right] \le q_\sigma \cdot \Pr[\boldsymbol{\pi} = \pi]\,. \tag{3.1}$$

For shorthand, we say that $\boldsymbol{\pi}$ is a q-fixing projection, for $q = \sqrt{q_0 q_1}$.

If needed, one can consider without loss of generality only variables x_i such that $\pi(x_i) \in \{0,1\}$, as otherwise Equation 3.1 holds trivially, with the left-hand side equaling zero.

Example 3.5. In order to get intuition for the definition of fixing projections, let us examine how this definition applies to random restrictions. In our terms, a restriction is a projection from $x_1, \dots, x_n$ to $x_1, \dots, x_n$ that maps every variable x_i either to itself or to $\{0,1\}$. Suppose that $\boldsymbol{\rho}$ is any distribution over restrictions, and that ρ is some fixed restriction. In this case, the condition of being q-fixing can be rewritten as follows:

$$\Pr[\boldsymbol{\rho}(x_i) = x_i \text{ and } \boldsymbol{\rho}_{x_i \leftarrow \sigma} = \rho] \le q_\sigma \cdot \Pr[\boldsymbol{\rho} = \rho]\,.$$

Denote by $\boldsymbol{\rho}', \rho'$ the restrictions obtained from $\boldsymbol{\rho}, \rho$ by truncating x_i (i. e., $\boldsymbol{\rho}' = \boldsymbol{\rho}|_{\{x_1,\dots,x_n\}-\{x_i\}}$). Using this notation, we can rewrite the foregoing equation as

$$\Pr[\boldsymbol{\rho}(x_i) = x_i \text{ and } \boldsymbol{\rho}' = \rho' \text{ and } \boldsymbol{\rho}_{x_i \leftarrow \sigma}(x_i) = \rho(x_i)] \leq q_\sigma \cdot \Pr[\boldsymbol{\rho}(x_i) = \rho(x_i) \text{ and } \boldsymbol{\rho}' = \rho'] \,.$$

Now, observe that it is always the case $\boldsymbol{\rho}_{x_i \leftarrow \sigma}(x_i) = \sigma$, and therefore the probability on the left-hand side is non-zero only if $\rho(x_i) = \sigma$. Hence, we can restrict ourselves to the latter case, and the foregoing equation can be rewritten again as

$$\Pr[\boldsymbol{\rho}(x_i) = x_i \text{ and } \boldsymbol{\rho}' = \rho'] \leq q_\sigma \cdot \Pr[\boldsymbol{\rho}(x_i) = \sigma \text{ and } \boldsymbol{\rho}' = \rho'] \,.$$

Finally, if we divide both sides by $\Pr[\boldsymbol{\rho}' = \rho']$, we obtain the following intuitive condition:

$$\Pr[\boldsymbol{\rho}(x_i) = x_i \mid \boldsymbol{\rho}' = \rho'] \leq q_\sigma \cdot \Pr[\boldsymbol{\rho}(x_i) = \sigma \mid \boldsymbol{\rho}' = \rho'] \,.$$

This condition informally says the following: $\boldsymbol{\rho}$ is a fixing projection if the probability of leaving x_i unfixed is at most q_σ times the probability of fixing it to σ, and this holds regardless of what the restriction assigns to the other variables.

In particular, it is now easy to see that the classic random restrictions are fixing projections. Recall that a *p-random restriction* fixes each variable independently with probability $1-p$ to a random bit. Due to the independence of the different variables, the foregoing condition simplifies to

$$\Pr[\boldsymbol{\rho}(x_i) = x_i] \leq q_\sigma \cdot \Pr[\boldsymbol{\rho}(x_i) = \sigma] \,,$$

and it is easy to see that this condition is satisfied for $q_0 = q_1 = \frac{2p}{1-p}$.

We prove the following shrinkage theorem for q-fixing projections, which is analogous to the shrinkage theorem of [23] for random restrictions in the case of *balanced* formulas.

Theorem 3.6 (Shrinkage under fixing projections)**.** *Let ϕ be a formula of size s and depth d, and let $\boldsymbol{\pi}$ be a q-fixing projection. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right] = O\!\left(q^2 \cdot d^2 \cdot s + q \cdot \sqrt{s}\right).$$

We prove Theorem 3.6 in Section 4. It is instructive to compare our shrinkage theorem to the shrinkage theorem by Håstad [23] for *unbalanced* formulas:

Theorem 3.7 ([23])**.** *Let ϕ be a formula of size s, and let $\boldsymbol{\rho}$ be a p-random restriction. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\rho}})\right] = O\!\left(p^2 \cdot \left(1 + \log^{3/2}\left(\min\left\{\frac{1}{p}, s\right\}\right)\right) \cdot s + p \cdot \sqrt{s}\right).$$

Our Theorem 3.6 has somewhat worse parameters compared to Theorem 3.7: specifically, the factor of d^2 does not appear in Theorem 3.7. The reason is that the proof of [23] uses a fairly complicated case analysis in order to avoid losing that factor, and we chose to skip this analysis in order to obtain a simpler proof. We did not check if the factor of d^2 in our result can be avoided by using a similar case analysis. By applying formula balancing (Theorem 2.4) to our shrinkage theorem, we can obtain the following result, which is independent of the depth of the formula.

Corollary 3.8. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function with formula complexity s, and let $\boldsymbol{\pi}$ be a q-fixing projection. Then*

$$\mathbb{E}\left[L(f|_{\boldsymbol{\pi}})\right] = q^2 \cdot s^{1+O\left(\frac{1}{\sqrt{\log s}}\right)} + q \cdot s^{1/2+O\left(\frac{1}{\sqrt{\log s}}\right)}.$$

Proof. By assumption, there exists a formula ϕ of size s that computes f. We balance the formula ϕ by applying Theorem 2.4 with $\alpha = \frac{1}{\sqrt{\log s}}$, and obtain a new formula ϕ' that computes f and has size $s^{1+\frac{1}{\sqrt{\log s}}}$ and depth $O(2^{\sqrt{\log s}} \cdot \log s) = s^{O(\frac{1}{\sqrt{\log s}})}$. The required result now follows by applying Theorem 3.6 to ϕ'. □

3.2 Hiding projections

Intuitively, a hiding projection is a random projection $\boldsymbol{\pi}$ in which, when given $\boldsymbol{\pi}_{y_j \leftarrow \sigma}$, it is hard to tell in which locations the variable y_j appears in $\boldsymbol{\pi}$. Formally, we define q-hiding projections as follows.

Definition 3.9. Let $0 \le q_0, q_1 \le 1$. We say that a random projection $\boldsymbol{\pi}$ is a (q_0, q_1)-*hiding projection* if for every projection π, every bit $\sigma \in \{0,1\}$, and all variables x_i, y_j, we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi\right] \le q_\sigma,$$

whenever the event conditioned on has positive probability. For shorthand, we say that $\boldsymbol{\pi}$ is a q-hiding projection, for $q = \sqrt{q_0 q_1}$.

To illustrate the definition, consider the following natural random restriction: given n variables $x_1, \dots, x_n$, the restriction chooses a set of m variables uniformly at random, and fixes all the other variables to random bits. This restriction is not captured by the notion of p-random restrictions or by fixing projections, but as we demonstrate next, it can be implemented by hiding projections. We start with the simple case of $m = 1$, and then consider the general case.

Example 3.10. In order to implement the case of $m = 1$, consider the random projection $\boldsymbol{\pi}$ from $x_1, \dots, x_n$ to y that is defined as follows: the projection $\boldsymbol{\pi}$ chooses an index $\boldsymbol{i} \in [n]$ and a bit $\boldsymbol{\tau} \in \{0,1\}$ uniformly at random, sets $\boldsymbol{\pi}(x_{\boldsymbol{i}}) = y^{\boldsymbol{\tau}}$, and sets $\boldsymbol{\pi}(x_{i'})$ to a random bit for all $i' \in [n] \setminus \{\boldsymbol{i}\}$. It is clear that $\boldsymbol{\pi}$ is essentially equivalent to the random restriction described above for $m = 1$. We claim that $\boldsymbol{\pi}$ is a $\frac{1}{n}$-hiding projection. To see this, observe that for every bit $\sigma \in \{0,1\}$, the projection $\boldsymbol{\pi}_{y \leftarrow \sigma}$ is a uniformly distributed string in $\{0,1\}^n$, and moreover, this is true conditioned on any possible value of $\boldsymbol{i}$. In particular, the random variable $\boldsymbol{i}$ is independent of $\boldsymbol{\pi}_{y \leftarrow \sigma}$. Therefore, for every projection $\pi \in \{0,1\}^n$ and index $i \in [n]$ we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \{y, \overline{y}\} \,\middle|\, \boldsymbol{\pi}_{y \leftarrow \sigma} = \pi\right] = \Pr\left[\boldsymbol{i} = i \,\middle|\, \boldsymbol{\pi}_{y \leftarrow \sigma} = \pi\right] = \Pr[\boldsymbol{i} = i] = \frac{1}{n},$$

so $\boldsymbol{\pi}$ satisfies the definition of a (q_0, q_1)-hiding projection with $q_0 = q_1 = \frac{1}{n}$. Intuitively, given $\boldsymbol{\pi}_{y \leftarrow \sigma}$, one cannot guess the original location of y in $\boldsymbol{\pi}$ better than a random guess.

Example 3.11. We turn to consider the case of a general $m \in \mathbb{N}$. Let $\boldsymbol{\pi}$ be the random projection from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$ that is defined as follows: the projection $\boldsymbol{\pi}$ chooses m distinct indices $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in [n]$ and m bits $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_m \in \{0,1\}$ uniformly at random, sets $\boldsymbol{\pi}(x_{\boldsymbol{i}_j}) = y_j^{\boldsymbol{\tau}_j}$ for every $j \in [m]$, and sets all the other variables x_i to random bits. It is clear that $\boldsymbol{\pi}$ is essentially equivalent to the random projection described above. We show that $\boldsymbol{\pi}$ is a $\frac{1}{n-m+1}$-hiding projection. To this end, we should show that for every $i \in [n]$, $j \in [m]$, and $\sigma \in \{0,1\}$ we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi\right] \leq \frac{1}{n-m+1}.$$

For simplicity of notation, we focus on the case where $j = m$. Observe that $\boldsymbol{\pi}_{y_m \leftarrow \sigma}$ reveals the values of $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{m-1}$, and the projection of $\boldsymbol{\pi}_{y_m \leftarrow \sigma}$ on the remaining $n - m + 1$ indices is uniform over $\{0,1\}^{n-m+1}$. Moreover, the latter assertion remains true conditioned on any possible value of $\boldsymbol{i}_m$ (which must be different from the known values of $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{m-1}$). Therefore given $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{m-1}$, the random variable $\boldsymbol{i}_m$ (ranging over all indices other than $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{m-1}$) is independent of the projection of $\boldsymbol{\pi}_{y_m \leftarrow \sigma}$ on the $n - m + 1$ indices other than $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{m-1}$. It follows that for every projection π in the support of $\boldsymbol{\pi}_{y_m \leftarrow \sigma}$ and every index $i \in [n]$, if $\pi(x_{i_1}) = y_1^{\tau_1}, \ldots, \pi(x_{i_{m-1}}) = y_{m-1}^{\tau_{m-1}}$ then:

$$\begin{aligned}\Pr\left[\boldsymbol{\pi}(x_i) \in \{y_m, \overline{y_m}\} \,\middle|\, \boldsymbol{\pi}_{y_m \leftarrow \sigma} = \pi\right] &= \Pr\left[\boldsymbol{i}_m = i \,\middle|\, \boldsymbol{\pi}_{y_m \leftarrow \sigma} = \pi\right] \\ &= \Pr[\boldsymbol{i}_m = i \mid \boldsymbol{i}_1 = i_1, \ldots, \boldsymbol{i}_{m-1} = i_{m-1}] \leq \frac{1}{n-m+1},\end{aligned}$$

as required.

In Section 5, we prove the following shrinkage theorem for hiding projections.

Theorem 3.12 (Shrinkage under hiding projections)**.** *Let ϕ be a formula of size s and depth d, and let $\boldsymbol{\pi}$ be a q-hiding projection. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right] = O\left(m^4 \cdot q^2 \cdot d^2 \cdot s + m^2 \cdot q \cdot \sqrt{s}\right).$$

Applying formula balancing, we can obtain an analog of Corollary 3.8, with an identical proof.

Corollary 3.13. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function with formula complexity s, and let $\boldsymbol{\pi}$ be a q-hiding projection. Then*

$$\mathbb{E}\left[L(f|_{\boldsymbol{\pi}})\right] = m^4 \cdot q^2 \cdot s^{1+O\left(\frac{1}{\sqrt{\log s}}\right)} + m^2 \cdot q \cdot s^{\frac{1}{2}+O\left(\frac{1}{\sqrt{\log s}}\right)}.$$

Remark 3.14. Note that Theorem 3.12 loses a factor of m^4 compared to Theorem 3.6. While this loss might be large in general, it will be of no importance in our applications. The factor m^4 can be improved to m^2 in our actual applications, as discussed in Section 5.1.

The following example shows that the loss of a factor of at least m^2 is necessary for Theorem 3.12. Let $f = \mathsf{Parity}_n$ be the Parity function over $x_1, \ldots, x_n$, and note that the formula complexity of f is $s = \Theta(n^2)$. Let $\boldsymbol{\pi}$ be the random projection from Example 3.11, and recall that it is a q-hiding projection for $q = \frac{1}{n-m+1}$. Then, $f|_{\boldsymbol{\pi}}$ is the Parity function over m bits, and therefore its formula complexity is $\Theta(m^2)$. It follows that when $m \le n/2$,

$$\mathbb{E}\left[L(f|_{\boldsymbol{\pi}})\right] = \Theta(m^2) = \Theta(m^2 \cdot q^2 \cdot s).$$

3.3 Sketch: Cubic formula lower bound for AC^0

We now sketch how our shrinkage theorems can be used to prove the cubic formula lower bound for $\boldsymbol{AC}^0$. We start by sketching how our shrinkage theorems can be used to reprove a quadratic formula lower bound for the surjectivity function Surj (originally due to [6]). Then, we sketch a lower bound for compositions of the form $f \diamond \mathsf{Surj}$, where f is an arbitrary function. Finally, we combine the latter bound with an idea of Andreev [4] to obtain a cubic lower bound for an explicit function in $\boldsymbol{AC}^0$. We note that the first step is the key idea of the proof, whereas the two latter steps are relatively standard. A formal proof following this sketch can be found in Section 8.

3.3.1 Lower bound for the surjectivity function

The surjectivity function $\mathsf{Surj}_n\colon \{0,1\}^n \to \{0,1\}$ [6] takes as input a list of r numbers in $[S]$ for some $S \in \mathbb{N}$, encoded in binary, and outputs 1 if and only if all the numbers in $[S]$ appear in the list. For our purposes, we write $S' = S - 2$ and set $r = \frac{3}{2} \cdot S' + 2$. We sketch a proof that $L(\mathsf{Surj}_n) \approx \Omega(S^2)$. Since the input length of Surj_n, when measured in bits, is $n = O(S \log S)$, this implies that $L(\mathsf{Surj}_n) \approx \Omega(\frac{n^2}{\log^2 n})$.

To this end, we construct a random projection $\boldsymbol{\pi}$ for Surj_n that maps the input variables $x_1, \dots, x_n$ to a single variable y as follows: We choose two distinct numbers $\boldsymbol{i}, \boldsymbol{j} \in [S]$ at random, and then partition the remaining $S - 2$ numbers at random to two sets $\boldsymbol{A}$ and $\boldsymbol{B}$ of size $\frac{1}{2}(S-2) = \frac{1}{2}S'$ each. Then, we create a list of $r = \frac{3}{2} \cdot S' + 2$ numbers in $[S]$ as follows:

1. Initially, all the places in the list are vacant.
2. We put each of the $\frac{1}{2}S' + 1$ numbers in $\boldsymbol{A} \cup \{\boldsymbol{j}\}$ at a single random vacant place in the list.
3. We put each of the $\frac{1}{2}S'$ numbers in $\boldsymbol{B}$ in two random vacant places in the list (for a total of S' places).
4. Note that so far we filled $\frac{3}{2}S' + 1 = r - 1$ vacant places in the list. The remaining vacant place in the list is left empty.

We set $\boldsymbol{\pi}$ such that it fixes all the inputs variables $x_1, \dots, x_n$ to constants according to the list we chose, except for the variables corresponding to the empty place in the list. Then, the $\lceil \log S \rceil$ binary input variables that correspond to the empty place are mapped to $\{0, 1, y, \overline{y}\}$ such that if we substitute $y = 1$ we get the number $\boldsymbol{i}$ and if we substitute $y = 0$ we get the number $\boldsymbol{j}$. Observe that if $y = 1$, then the numbers in $\boldsymbol{A} \cup \{\boldsymbol{i}, \boldsymbol{j}\}$ appear in the list exactly once and the numbers in $\boldsymbol{B}$ appear exactly twice. On the other hand, if $y = 0$, then the numbers in $\boldsymbol{A}$ appear in the list exactly once, the numbers in $\boldsymbol{B} \cup \{\boldsymbol{j}\}$ appear in the list exactly twice, and $\boldsymbol{i}$ does not appear at all.

Observe that $\mathsf{Surj}_n|_{\boldsymbol{\pi}} = y$. Indeed, if $y = 1$ then the foregoing list of numbers contains all the numbers in $[S]$ and thus Surj_n takes the value 1, and if $y = 0$ then the list of numbers does not contain $\boldsymbol{i}$ and thus Surj_n takes the value 0. We claim that $\boldsymbol{\pi}$ is an $O(\frac{1}{S})$-hiding projection. Intuitively, if we substitute a value $\sigma \in \{0, 1\}$ in y, then given the resulting projection $\boldsymbol{\pi}_{y\leftarrow\sigma}$ it is difficult to tell which input variables were mapped to y in the original projection $\boldsymbol{\pi}$. The reason is that in order to tell that, one has to guess the empty place in the original list. To this end, one has to guess $\boldsymbol{i}$ among the $\approx \frac{1}{2}S'$ numbers that appear exactly once in the list (if $\sigma = 1$) or to guess $\boldsymbol{j}$ among the $\approx \frac{1}{2}S'$ numbers that appear exactly twice in the list (if $\sigma = 0$).

A bit more formally, observe that $\boldsymbol{\pi}$ maps an input variable x_h to $\{y, \overline{y}\}$ only if x_h corresponds to the empty place in the list of $\boldsymbol{\pi}$, and that $\boldsymbol{\pi}_{y\leftarrow 0}$ and $\boldsymbol{\pi}_{y\leftarrow 1}$ represent lists of r numbers in $[S]$. Hence, to find an upper bound on the probability that $\boldsymbol{\pi}(x_h) \in \{y, \overline{y}\}$ conditioned on the choice of $\boldsymbol{\pi}_{y\leftarrow\sigma}$ (for $\sigma \in \{0, 1\}$), we need an upper bound on the probability that a place in the list represented by $\boldsymbol{\pi}_{y\leftarrow\sigma}$ was the empty place in the list of $\boldsymbol{\pi}$. We consider two cases:

- If $\sigma = 1$, then the empty place in the list of $\boldsymbol{\pi}$ contains $\boldsymbol{i}$ in the list of $\boldsymbol{\pi}_{y\leftarrow 1}$. From the perspective of someone who only sees $\boldsymbol{\pi}_{y\leftarrow 1}$, the number $\boldsymbol{i}$ could be any number among the $\frac{1}{2}S' + 2$ numbers that appear exactly once in the list of $\boldsymbol{\pi}_{y\leftarrow 1}$. Therefore, conditioned on $\boldsymbol{\pi}_{y\leftarrow 1}$, the empty place in the list of $\boldsymbol{\pi}$ is uniformly distributed among $\frac{1}{2}S' + 2$ places in the list of $\boldsymbol{\pi}_{y\leftarrow 1}$. Hence, any place in the list of $\boldsymbol{\pi}_{y\leftarrow 1}$ has probability of at most $\frac{1}{\frac{1}{2}S'+2} = O\left(\frac{1}{S}\right)$ to be the empty place in $\boldsymbol{\pi}$.

- On the other hand, if $\sigma = 0$, then the empty place in the list of $\boldsymbol{\pi}$ contains $\boldsymbol{j}$ in the list of $\boldsymbol{\pi}_{y\leftarrow 0}$. From the perspective of someone who only sees $\boldsymbol{\pi}_{y\leftarrow 0}$, the number $\boldsymbol{j}$ could be any number among the $\frac{1}{2}S' + 1$ numbers that appear exactly twice in the list of $\boldsymbol{\pi}_{y\leftarrow 0}$. Therefore, conditioned on $\boldsymbol{\pi}_{y\leftarrow 0}$, the empty place in the list of $\boldsymbol{\pi}$ is uniformly distributed among $2 \cdot \left(\frac{1}{2}S' + 1\right)$ places in the list of $\boldsymbol{\pi}_{y\leftarrow 0}$. Hence, any place in the list has probability of at most $\frac{1}{2\cdot\left(\frac{1}{2}S'+1\right)} = O\left(\frac{1}{S}\right)$ to be the empty place in $\boldsymbol{\pi}$.

We showed that $\boldsymbol{\pi}$ is an $O(\frac{1}{S})$-hiding projection, and therefore by our shrinkage theorem for hiding projections (specifically, Corollary 3.13) we get that

$$\mathbb{E}\left[L(\mathsf{Surj}_n|_{\boldsymbol{\pi}})\right] \lesssim \frac{1}{S^2} \cdot L(\mathsf{Surj}).$$

(We ignored the power of $\frac{1}{\sqrt{\log L(\mathsf{Surj}_n)}}$ from Corollary 3.13 in order to simplify the exposition.) On the other hand, as explained above, we have $\mathsf{Surj}|_{\boldsymbol{\pi}} = y$, and therefore $L(\mathsf{Surj}_n|_{\boldsymbol{\pi}}) = 1$. It follows that $L(\mathsf{Surj}_n) \gtrsim S^2$, as required.

We note that the above proof is a special case of a more general phenonmenon. Specifically, in Section 7.1, we show that such a hiding projection can be constructed for every function that has a large (unweighted) adversary bound. As [6] showed, the surjectivity function has such a large adversary bound, and therefore we can construct a hiding projection for it.

3.3.2 Lower bound for $f \diamond \mathsf{Surj}_n$

Let $f\colon \{0,1\}^k \to \{0,1\}$ be any function, and let Surj_n be defined as in the previous section. The block-composition of f and Surj_n, denoted $f \diamond \mathsf{Surj}_n$, is the function that takes k inputs $x_1, \ldots, x_k \in \{0,1\}^n$ for Surj_n and outputs

$$(f \diamond \mathsf{Surj}_n)(x_1, \ldots, x_k) = f\left(\mathsf{Surj}(x_1), \ldots, \mathsf{Surj}(x_k)\right)$$

We sketch a proof that $L(f \diamond \mathsf{Surj}_n) \approx \Omega\left(L(f) \cdot \frac{n^2}{\log^2 n}\right)$. Denote by $x_{i,1}, \ldots, x_{i,n}$ the boolean variables of the input x_i for each $i \in [k]$. Let $\boldsymbol{\pi}$ be the $O(\frac{1}{S})$-hiding projection from the previous section, and let $\boldsymbol{\pi}^k$ be random projection from $x_{1,1}, \ldots, x_{k,n}$ to $y_1, \ldots, y_k$ that is obtained by joining k independent copies of $\boldsymbol{\pi}$. In Section 6 we show that the join operation maintains the hiding property, and therefore $\boldsymbol{\pi}^k$ is an $O(\frac{1}{S})$-hiding projection. Now, it is not hard to see that $f \diamond \mathsf{Surj}_n|_{\boldsymbol{\pi}^k} = f$ and therefore $L(f \diamond \mathsf{Surj}_n|_{\boldsymbol{\pi}^k}) = L(f)$. On the other hand, by our shrinkage theorem for hiding projections (specifically, Corollary 3.13) we get that

$$\mathbb{E}\left[L(f \diamond \mathsf{Surj}_n|_{\boldsymbol{\pi}^k})\right] \lesssim \frac{1}{S^2} \cdot L(f \diamond \mathsf{Surj}_n).$$

(Again, we ignored the power of $\frac{1}{\sqrt{\log L(f \diamond \mathsf{Surj}_n)}}$ from Corollary 3.13 in order to simplify the exposition.) It follows that

$$L(f) \lesssim \frac{1}{S^2} \cdot L(f \diamond \mathsf{Surj}_n),$$

and therefore

$$L(f \diamond \mathsf{Surj}_n) \approx \Omega\left(L(f) \cdot S^2\right) = \tilde{\Omega}\left(L(f) \cdot n^2\right),$$

as required.

3.3.3 Cubic lower bound for AC^0

Let $n = 2^k \cdot (k+1)$ and let $h_n\colon \{0,1\}^n \to \{0,1\}$ be the function that takes as input the truth table of a function $f\colon \{0,1\}^k \to \{0,1\}$ and k inputs $x_1, \ldots, x_k$ for Surj_{2^k}, and outputs

$$F(f, x_1, \ldots, x_k) = (f \diamond \mathsf{Surj})(x_1, \ldots, .x_k)$$

As mentioned above, the function h_n is a variant of the Andreev function in which the parity function is replaced with Surj. It is not hard to show that h_n is in $\boldsymbol{AC}^0$ (see Section 8.2 for details). We sketch a proof that $L(h_n) \approx \Omega(n^{3-o(1)})$.

To see that $L(h_n) \approx \Omega(n^{3-o(1)})$, recall that by a standard counting argument, there exists a function $f\colon \{0,1\}^k \to \{0,1\}$ such that $L(f) = \Theta\left(\frac{2^k}{\log k}\right)$. We now hardwire f as an input to h_n, which turns the function h_n to the function $f \diamond \mathsf{Surj}_{2^k}$. By the lower bound we have for $f \diamond \mathsf{Surj}_{2^k}$, it follows that

$$L(F) \geq L(f \diamond \mathsf{Surj}) \approx \tilde{\Omega}\left(L(f) \cdot 2^{2k}\right) = \tilde{\Omega}\left(\frac{2^k}{\log k} \cdot 2^{2k}\right) = \Omega(n^{3-o(1)}),$$

as required.

Remark 3.15. Note that in the proof of this result, we did not use the shrinkage theorem for fixing projections — indeed, we only used the shrinkage theorem for hiding projections. However, the proof of the shrinkage theorem for hiding projections itself relies on the shrinkage theorem for fixing projections.

4 Proof of shrinkage theorem for fixing projections

In this section, we prove our the shrinkage theorem for fixing projections. Our proof is based on the ideas of [23], but the presentation is different. We start by recalling the definition of fixing projections and the theorem's statement.

Definition 3.4. Let $0 \leq q_0, q_1 \leq 1$. We say that a random projection $\boldsymbol{\pi}$ is a (q_0, q_1)-*fixing projection* if for every projection π, every bit $\sigma \in \{0,1\}$, and every variable x_i, we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i)) \leftarrow \sigma} = \pi\right] \leq q_\sigma \cdot \Pr[\boldsymbol{\pi} = \pi].$$

For shorthand, we say that $\boldsymbol{\pi}$ is a q-fixing projection, for $q = \sqrt{q_0 q_1}$.

Theorem 3.6 (Shrinkage under fixing projections)**.** *Let ϕ be a formula of size s and depth d, and let $\boldsymbol{\pi}$ be a q-fixing projection. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right] = O\left(q^2 \cdot d^2 \cdot s + q \cdot \sqrt{s}\right).$$

Fix a formula ϕ of size s and depth d, and let $\boldsymbol{\pi}$ be a q-fixing projection. We would like to get an upper bound on the expectation of $L(\phi|_{\boldsymbol{\pi}})$. As in [23], we start by giving an upper bound the probability that the projection $\boldsymbol{\pi}$ shrinks a formula to size 1. Specifically, we prove the following lemma in Section 4.1.

Lemma 4.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function, and let $\boldsymbol{\pi}$ be a q-fixing projection. Then,*

$$\Pr[L(f|_{\boldsymbol{\pi}}) = 1] \leq q \cdot \sqrt{L(f)}.$$

Next, we show that to find and upper bound on the expectation of $L(\phi|_{\boldsymbol{\pi}})$, it suffices to give an upper bound on the probability that the projection $\boldsymbol{\pi}$ shrinks two formulas to size 1 simultaneously. In order to state this claim formally, we introduce some notation.

Notation 4.2. Let g be a gate of ϕ. We denote the depth of g in ϕ by $\mathrm{depth}_\phi(g)$, and omit ϕ if it is clear from context (the root has depth 0). If g is an internal gate, we denote the subformulas that are rooted in its left and right children by $\mathrm{left}(g)$ and $\mathrm{right}(g)$, respectively. Here, by "internal gate", we refer to a gate which is an internal node of the tree, i. e., either an AND or an OR gate.

We prove the following lemma, which says that in order to upper-bound $\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right]$ it suffices to upper-bound, for every internal gate g, the probability that $\mathrm{left}(g)$ and $\mathrm{right}(g)$ shrink to size 1 under $\boldsymbol{\pi}$.

Lemma 4.3. *For every projection π it holds that*

$$L(\phi|_\pi) \leq \sum_{\text{internal gate } g \text{ of } \phi} (\text{depth}(g)+2)\cdot 1_{\{L(\text{left}(g)|_\pi)=1 \text{and } L(\text{right}(g)|_\pi)=1\}} + 1_{L(\phi|_\pi)=1}.$$

We would like to use Lemmas 4.1 and 4.3 to prove the shrinkage theorem. As a warm-up, let us make the simplifying assumption that for every two functions $f_1, f_2\colon \{0,1\}^n \to \{0,1\}$, the events $L(f_1|_\pi)=1$ and $L(f_2|_\pi)=1$ are independent. If this were true, we could have given an upper bound on $\mathbb{E}\left[L(\phi|_\pi)\right]$ as follows:

$$\begin{aligned}
\mathbb{E}\left[L(\phi|_\pi)\right] \leq \sum_{\text{int. gate } g \text{ of } \phi} (\text{depth}(g)+2)\cdot \mathbb{E}\left[1_{\{L(\text{left}(g)|_\pi)=1 \text{ and } L(\text{right}(g)|_\pi)=1\}}\right] + \mathbb{E}\left[1_{L(\phi|_\pi)=1}\right]& \\
&\text{(Lemma 4.3)}\\
\leq (d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} \Pr\big[L(\text{left}(g)|_\pi)=1 \text{ and } L(\text{right}(g)|_\pi)=1\big] + \mathbb{E}\left[1_{L(\phi|_\pi)=1}\right]& \\
&(\phi \text{ is of depth } d)\\
= (d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} \Pr[L(\text{left}(g)|_\pi)=1]\cdot \Pr\big[L(\text{right}(g)|_\pi)=1\big] + \mathbb{E}\left[1_{L(\phi|_\pi)=1}\right]& \\
&\text{(simplifying assumption)}\\
\leq (d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} q^2\cdot\sqrt{L(\text{left}(g))\cdot L(\text{right}(g))} + q\cdot\sqrt{s}& \quad \text{(Lemma 4.1)}\\
\leq q^2\cdot(d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} \big(L(\text{left}(g)) + L(\text{right}(g))\big) + q\cdot\sqrt{s}& \quad \text{(AM–GM inequality)}\\
\leq q^2\cdot(d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} \big(\text{size}(\text{left}(g)) + \text{size}(\text{right}(g))\big) + q\cdot\sqrt{s}&\\
= q^2\cdot(d+2)\cdot \sum_{\text{int. gate } g \text{ of } \phi} \text{size}(g) + q\cdot\sqrt{s}.&
\end{aligned}$$

The last sum counts every leaf ℓ of ϕ once for each internal ancestor of ℓ, so the last expression is equal to

$$\begin{aligned}
q^2\cdot(d+2)\cdot &\sum_{\text{leaf } \ell \text{ of } \phi} \text{depth}(\ell) + q\cdot\sqrt{s}\\
&\leq q^2\cdot(d+2)\cdot \sum_{\text{leaf } \ell \text{ of } \phi} d + q\cdot\sqrt{s}\\
&= q^2\cdot(d+2)\cdot d\cdot s + q\cdot\sqrt{s}\\
&= O\left(q^2\cdot d^2\cdot s + q\cdot\sqrt{s}\right),
\end{aligned}$$

which is the bound we wanted. However, the above calculation only works under our simplifying assumption, which is false: the events $L(f_1|_\pi) = 1$ and $L(f_2|_\pi) = 1$ will often be dependent. In particular, in order for the foregoing calculation to work, we need the following inequality to hold:

$$\Pr[L(f_2|_\pi) = 1 \mid L(f_1|_\pi) = 1] \leq q \cdot \sqrt{L(f_2)}.$$

This inequality holds under our simplifying assumption by Lemma 4.1, but may not hold in general. Nevertheless, we prove the following similar statement in Section 4.2.

Lemma 4.4. *Let $\boldsymbol{\pi}$ be a q-fixing projection. Let $f_1, f_2 \colon \{0,1\}^n \to \{0,1\}$, let $\sigma, \tau \in \{0,1\}$, and let y_j be a variable. Then,*

$$\Pr\left[L(f_2|_{\boldsymbol{\pi}_{y_j \leftarrow \sigma}}) = 1 \;\middle|\; f_1|_{\boldsymbol{\pi}} = y_j^\tau\right] \leq q \cdot \sqrt{L(f_2)}.$$

Intuitively, Lemma 4.4 breaks the dependency between the events $L(f_1|_{\boldsymbol{\pi}}) = 1$ and $L(f_2|_{\boldsymbol{\pi}}) = 1$ by fixing in f_2 the single literal to which f_1 has shrunk. We would now like to use Lemma 4.4 to prove the theorem. To this end, we prove an appropriate variant of Lemma 4.3, which allows using the projection $\pi_{y_j \leftarrow \sigma}$ rather than π in the second function. This variant is motivated by the following "one-variable simplification rules" of [23], which are easy to verify.

Fact 4.5 (one-variable simplification rules)**.** *Let $h \colon \{0,1\}^m \to \{0,1\}$ be a function over the variables $y_1, \dots, y_m$, and let $\sigma \in \{0,1\}$. We denote by $h_{y_j \leftarrow \sigma}$ the function obtained from h by setting y_j to the bit σ. Then:*

- *The function $y_j^\sigma \vee h$ is equal to the function $y_j^\sigma \vee h_{y_j \leftarrow \sigma}$.*
- *The function $y_j^\sigma \wedge h$ is equal to the function $y_j^\sigma \wedge h_{y_j \leftarrow \overline{\sigma}}$.*

In order to use the simplification rules, we define, for every internal gate g of ϕ and projection π, an event $\mathcal{E}_{g,\pi}$ as follows: if g is an OR gate, then $\mathcal{E}_{g,\pi}$ is the event that there exists some literal y_j^σ (for $\sigma \in \{0,1\}$) such that $\text{left}(g)|_\pi = y_j^\sigma$ and $L(\text{right}(g)|_{\pi_{y_j \leftarrow \sigma}}) = 1$. If g is an AND gate, then $\mathcal{E}_{g,\pi}$ is defined similarly, except that we replace $\pi_{y_j \leftarrow \sigma}$ with $\pi_{y_j \leftarrow \overline{\sigma}}$. We have the following lemma, which is proved in Section 4.3.

Lemma 4.6. *For every projection π we have*

$$L(\phi|_\pi) \leq \sum_{\textit{internal gate } g \textit{ of } \phi} (\text{depth}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1_{L(\phi|_\pi)=1}.$$

We can now use the following corollary of Lemma 4.4 to replace our simplifying assumption.

Corollary 4.7. *For every internal gate g of ϕ we have*

$$\Pr\left[\mathcal{E}_{g,\boldsymbol{\pi}}\right] \leq q^2 \cdot \sqrt{L(\text{left}(g)) \cdot L(\text{right}(g))}.$$

Proof. Let g be an internal gate of ϕ. We prove the corollary for the case where g is an OR gate, and the proof for the case that g is an AND gate is similar. We have

$$\begin{aligned}
\Pr\left[\mathcal{E}_{g,\pi}\right] &= \Pr\left[\exists \text{ literal } y_j^\sigma : \text{left}(g)|_\pi = y_j^\sigma \text{ and } L(\text{right}(g)|_{\pi_{y_j \leftarrow \sigma}}) = 1\right] \\
&= \sum_{\text{literal } y_j^\sigma} \Pr\left[\text{left}(g)|_\pi = y_j^\sigma \text{ and } L(\text{right}(g)|_{\pi_{y_j \leftarrow \sigma}}) = 1\right] \\
&= \sum_{\text{literal } y_j^\sigma} \Pr\left[L(\text{right}(g)|_{\pi_{y_j \leftarrow \sigma}}) = 1 \middle| \text{left}(g)|_\pi = y_j^\sigma\right] \cdot \Pr\left[\text{left}(g)|_\pi = y_j^\sigma\right] \\
&\leq q \cdot \sqrt{L(\text{right}(g))} \cdot \sum_{\text{literal } y_j^\sigma} \Pr\left[\text{left}(g)|_\pi = y_j^\sigma\right] && \text{(Lemma 4.4)} \\
&= q \cdot \sqrt{L(\text{right}(g))} \cdot \Pr[L(\text{left}(g)|_\pi) = 1] \\
&\leq q^2 \cdot \sqrt{L(\text{left}(g)) \cdot L(\text{right}(g))}, && \text{(Lemma 4.1)}
\end{aligned}$$

as required. □

The shrinkage theorem now follows using the same calculation as above, replacing Lemma 4.3 with Lemma 4.6 and the simplifying assumption with Corollary 4.7:

$$\begin{aligned}
\mathbb{E}\left[L(\phi|_\pi)\right] &\leq (d+2) \cdot \sum_{\text{internal gate } g \text{ of } \phi} \Pr\left[\mathcal{E}_{g,\pi}\right] + q \cdot \sqrt{s} && \text{(Lemma 4.6)} \\
&\leq q^2 \cdot (d+2) \cdot \sum_{\text{internal gate } g \text{ of } \phi} \sqrt{L(\text{left}(g)) \cdot L(\text{right}(g))} + q \cdot \sqrt{s} && \text{(Corollary 4.7)} \\
&\leq q^2 \cdot (d+2) \cdot \sum_{\text{internal gate } g \text{ of } \phi} \left(L(\text{left}(g)) + L(\text{right}(g))\right) + q \cdot \sqrt{s} && \text{(AM–GM ineq.)} \\
&= q^2 \cdot (d+2) \cdot \sum_{\text{internal gate } g \text{ of } \phi} \text{size}(g) + q \cdot \sqrt{s} \\
&\leq O(q^2 \cdot d^2 \cdot s + q \cdot \sqrt{s}).
\end{aligned}$$

In the remainder of this section, we prove Lemmas 4.1, and 4.4, and 4.6.

Remark 4.8. In this paper, we do not prove Lemma 4.3, since we do not actually need it to prove our main results in Section 8. However, this lemma can be established using the proof of Lemma 4.6, with some minor changes.

4.1 Proof of Lemma 4.1

We begin with proving Lemma 4.1, restated next.

Lemma 4.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function, and let $\boldsymbol{\pi}$ be a q-fixing projection. Then,*

$$\Pr[L(f|_{\boldsymbol{\pi}}) = 1] \le q \cdot \sqrt{L(f)}.$$

Let $f\colon \{0,1\}^n \to \{0,1\}$, and let $\mathcal{E}$ be the set of projections π such that $L(f|_\pi) = 1$. We prove that the probability that $\boldsymbol{\pi} \in \mathcal{E}$ is at most $q \cdot \sqrt{L(f)}$. Our proof follows closely the proof of [23, Lemma 4.1].

Let Π be a protocol that solves KW_f and has $L(f)$ leaves (such a protocol exists by Theorem 2.10). Let $\mathcal{A}$ and $\mathcal{B}$ be the sets of projections π for which $f|_\pi$ is the constants 1 and 0, respectively. We extend the protocol Π to take inputs from $\mathcal{A} \times \mathcal{B}$ as follows: when Alice and Bob are given as inputs the projections $\pi^A \in \mathcal{A}$ and $\pi^B \in \mathcal{B}$, respectively, they construct strings $a, b \in \{0,1\}^n$ from π^A, π^B by substituting 0 for all the variables $y_1, \dots, y_m$, and invoke Π on the inputs a and b. Observe that a and b are indeed legal inputs for Π (since $f(a) = 1$ and $f(b) = 0$). Moreover, recall that the protocol Π induces a partition of $\mathcal{A} \times \mathcal{B}$ to combinatorial rectangles, and that we denote the rectangle of the leaf ℓ by $\mathcal{A}_\ell \times \mathcal{B}_\ell$ (see Notation 2.8).

Our proof strategy is the following: We associate with every projection $\pi \in \mathcal{E}$ a leaf of Π, denoted $\operatorname{leaf}(\pi)$. We consider the two disjoint events $\mathcal{E}^+, \mathcal{E}^-$ that correspond to the event that $f|_\pi$ is a single positive literal or a single negative literal, respectively, and show that for every leaf ℓ we have

$$\Pr\left[\boldsymbol{\pi} \in \mathcal{E}^+ \text{ and } \operatorname{leaf}(\boldsymbol{\pi}) = \ell\right] \le q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]} \tag{4.1}$$

$$\Pr[\boldsymbol{\pi} \in \mathcal{E}^- \text{ and } \operatorname{leaf}(\boldsymbol{\pi}) = \ell] \le q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}. \tag{4.2}$$

Together, the two inequalities imply that

$$\Pr[\boldsymbol{\pi} \in \mathcal{E} \text{ and } \operatorname{leaf}(\boldsymbol{\pi}) = \ell] \le 2q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}.$$

The desired bound on $\Pr[\boldsymbol{\pi} \in \mathcal{E}]$ will follow by summing the latter bound over all the leaves ℓ of Π.

We start by explaining how to associate a leaf with every projection $\pi \in \mathcal{E}^+$. Let $\pi \in \mathcal{E}^+$. Then, it must be the case that $f|_\pi = y_j$ for some $j \in [m]$. We define the projections $\pi^1 = \pi_{y_j \leftarrow 1}$ and $\pi^0 = \pi_{y_j \leftarrow 0}$, and observe that $\pi^1 \in \mathcal{A}$ and $\pi^0 \in \mathcal{B}$. We now define $\operatorname{leaf}(\pi)$ to be the leaf to which Π arrives when invoked on inputs π^1 and π^0. Observe that the output of Π at $\operatorname{leaf}(\pi)$ must be a variable x_i that satisfies $\pi(x_i) \in \left\{y_j, \overline{y}_j\right\}$, and thus $\pi_{\mathsf{var}(\pi(x_i)) \leftarrow 1} = \pi^1$.

Next, fix a leaf ℓ. We prove that $\Pr[\boldsymbol{\pi} \in \mathcal{E}^+ \text{ and } \operatorname{leaf}(\boldsymbol{\pi}) = \ell] \le q_1 \cdot \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell]$. Let x_i be the output of the protocol Π at ℓ. Then,

$$\begin{aligned}
\Pr\left[\boldsymbol{\pi} \in \mathcal{E}^+ \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell\right] &\leq \Pr\left[\boldsymbol{\pi}^1 \in \mathcal{A}_\ell \text{ and } \boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i))\leftarrow 1} = \boldsymbol{\pi}^1\right] \\
&\leq \Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i))\leftarrow 1} \in \mathcal{A}_\ell\right] \\
&= \sum_{\pi \in \mathcal{A}_\ell} \Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i))\leftarrow 1} = \pi\right] \\
&\leq q_1 \cdot \sum_{\pi \in \mathcal{A}_\ell} \Pr[\boldsymbol{\pi} = \pi] \qquad\qquad \text{(since } \boldsymbol{\pi} \text{ is } (q_0, q_1)\text{-fixing)} \\
&= q_1 \cdot \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell]\,.
\end{aligned}$$

Similarly, it can be proved that $\Pr[\boldsymbol{\pi} \in \mathcal{E}^+ \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell] \leq q_0 \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]$. Together, the two bounds imply that

$$\Pr\left[\boldsymbol{\pi} \in \mathcal{E}^+ \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell\right] \leq \sqrt{q_1 \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot q_0 \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]} = q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}$$

for every leaf ℓ of Π. We define $\mathrm{leaf}(\pi)$ for projections $\pi \in \mathcal{E}^-$ in an analogous way, and then a similar argument shows that

$$\Pr[\boldsymbol{\pi} \in \mathcal{E}^- \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell] \leq q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}.$$

It follows that

$$\Pr[\boldsymbol{\pi} \in \mathcal{E} \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell] \leq 2q \cdot \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}.$$

Finally, let $\mathcal{L}$ denote the set of leaves of Π. It holds that

$$\begin{aligned}
\Pr[\boldsymbol{\pi} \in \mathcal{E}] &= \sum_{\ell \in \mathcal{L}} \Pr[\boldsymbol{\pi} \in \mathcal{E} \text{ and } \mathrm{leaf}(\boldsymbol{\pi}) = \ell] \\
&\leq 2q \cdot \sum_{\ell \in \mathcal{L}} \sqrt{\Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]} \\
&\leq 2q \cdot \sqrt{|\mathcal{L}|} \cdot \sqrt{\sum_{\ell \in \mathcal{L}} \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]} \qquad \text{(Cauchy-Schwarz – see Fact 2.12)} \\
&= 2q \cdot \sqrt{L(f)} \cdot \sqrt{\sum_{\ell \in \mathcal{L}} \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell]}.
\end{aligned}$$

We conclude the proof by showing that $\sum_{\ell \in \mathcal{L}} \Pr[\boldsymbol{\pi} \in \mathcal{A}_\ell] \cdot \Pr[\boldsymbol{\pi} \in \mathcal{B}_\ell] \leq \frac{1}{4}$. To this end, let $\boldsymbol{\pi}^A, \boldsymbol{\pi}^B$

be two independent random variables that are distributed identically to $\boldsymbol{\pi}$. Then, we have

$$\begin{aligned}
\sum_{\ell\in\mathcal{L}} \Pr[\boldsymbol{\pi}\in\mathcal{A}_\ell]\cdot\Pr[\boldsymbol{\pi}\in\mathcal{B}_\ell] &= \sum_{\ell\in\mathcal{L}} \Pr\left[\boldsymbol{\pi}^A\in\mathcal{A}_\ell\right]\cdot\Pr\left[\boldsymbol{\pi}^B\in\mathcal{B}_\ell\right] \\
&= \sum_{\ell\in\mathcal{L}} \Pr\left[(\boldsymbol{\pi}^A,\boldsymbol{\pi}^B)\in\mathcal{A}_\ell\times\mathcal{B}_\ell\right] && (\boldsymbol{\pi}^A,\boldsymbol{\pi}^B \text{ are independent}) \\
&= \Pr\left[(\boldsymbol{\pi}^A,\boldsymbol{\pi}^B)\in\mathcal{A}\times\mathcal{B}\right] && (\mathcal{A}_\ell\times\mathcal{B}_\ell \text{ are a partition of } \mathcal{A}\times\mathcal{B}) \\
&= \Pr[\boldsymbol{\pi}^A\in\mathcal{A}]\cdot\Pr[\boldsymbol{\pi}^B\in\mathcal{B}] && (\boldsymbol{\pi}^A,\boldsymbol{\pi}^B \text{ are independent}) \\
&= \Pr[\boldsymbol{\pi}\in\mathcal{A}]\cdot\Pr[\boldsymbol{\pi}\in\mathcal{B}] \\
&\le \Pr[\boldsymbol{\pi}\in\mathcal{A}]\cdot(1-\Pr[\boldsymbol{\pi}\in\mathcal{A}]) \le 1/4. && (\mathcal{A},\mathcal{B} \text{ are disjoint})
\end{aligned}$$

4.2 Proof of Lemma 4.4

We turn to proving Lemma 4.4, restated next.

Lemma 4.4. *Let $\boldsymbol{\pi}$ be a q-fixing projection. Let $f_1, f_2\colon \{0,1\}^n \to \{0,1\}$, let $\sigma,\tau\in\{0,1\}$, and let y_j be a variable. Then,*

$$\Pr\left[L(f_2|_{\boldsymbol{\pi}_{y_j\leftarrow\sigma}}) = 1 \,\middle|\, f_1|_{\boldsymbol{\pi}} = y_j^\tau\right] \le q\cdot\sqrt{L(f_2)}. \tag{4.3}$$

Let $\boldsymbol{\pi}$ be a (q_0,q_1)-fixing random projection, and let $q=\sqrt{q_0q_1}$. Let $f_1, f_2\colon \{0,1\}^n \to \{0,1\}$, let $\sigma,\tau\in\{0,1\}$, and let y_j be a variable. We prove Equation 4.3. For simplicity, we focus on the case that $f_1|_{\boldsymbol{\pi}} = y_j$, and the case that $f_1|_{\boldsymbol{\pi}} = \overline{y_j}$ can be dealt with similarly. The crux of the proof is to show that the random projection $\boldsymbol{\pi}_{y_j\leftarrow\sigma}$ is essentially a (q_0,q_1)-fixing projection even when conditioned on the event $f_1|_{\boldsymbol{\pi}} = y_j$, and therefore Equation 4.3 is implied immediately by Lemma 4.1.

In order to carry out this argument, we first establish some notation. Let $\boldsymbol{I}^+ = \boldsymbol{\pi}^{-1}(y_j)$ and $\boldsymbol{I}^- = \boldsymbol{\pi}^{-1}(\overline{y_j})$, and denote by $\boldsymbol{\pi}'$ the projection obtained from $\boldsymbol{\pi}$ by restricting its domain to $\{x_1,\ldots,x_n\}\setminus(\boldsymbol{I}^+\cup\boldsymbol{I}^-)$. We denote by $f_{2,\boldsymbol{I}^+,\boldsymbol{I}^-}$ the function over $\{x_1,\ldots,x_n\}\setminus(\boldsymbol{I}^+\cup\boldsymbol{I}^-)$ that is obtained from f_2 by hard-wiring σ and $\overline{\sigma}$ to the variables in $\boldsymbol{I}^+$ and $\boldsymbol{I}^-$, respectively. Observe that $f_2|_{\boldsymbol{\pi}_{y_j\leftarrow\sigma}} = f_{2,\boldsymbol{I}^+,\boldsymbol{I}^-}|_{\boldsymbol{\pi}'}$, so it suffices to prove that for every two disjoint sets $I^+, I^-\subseteq\{x_1,\ldots,x_n\}$ we have

$$\Pr\left[L(f_{2,\boldsymbol{I}^+,\boldsymbol{I}^-}|_{\boldsymbol{\pi}'}) = 1 \,\middle|\, f_1|_{\boldsymbol{\pi}} = y_j, \boldsymbol{I}^+ = I^+, \boldsymbol{I}^- = I^-\right] \le q\cdot\sqrt{L(f_2)}. \tag{4.4}$$

Let $I^+, I^-\subseteq\{x_1,\ldots,x_n\}$ be disjoint sets, and let $\mathcal{I}$ be the event that $\boldsymbol{I}^+ = I^+$ and $\boldsymbol{I}^- = I^-$. For convenience, let $K = \{x_1,\ldots,x_n\}\setminus(I^+\cup I^-)$ and $Y = \{y_1,\ldots,y_m\}\setminus\{y_j\}$, so $\boldsymbol{\pi}'$ is a random projection from K to Y when conditioned on $\mathcal{I}$. To prove Equation 4.4, it suffices to prove that $\boldsymbol{\pi}'$ is a (q_0,q_1)-fixing projection when conditioned on the events $\mathcal{I}$ and $f_1|_{\boldsymbol{\pi}} = y_j$, and then the inequality will follow from Lemma 4.1. We first prove that $\boldsymbol{\pi}'$ is a (q_0,q_1)-fixing projection when conditioning only on the event $\mathcal{I}$ (and not on $f_1|_{\boldsymbol{\pi}} = y_j$).

Proposition 4.9. *Conditioned on the event $\mathcal{I}$, the projection $\boldsymbol{\pi}'$ is a (q_0,q_1)-fixing projection.*

Proof. We prove that $\boldsymbol{\pi}'$ satisfies the definition of a fixing projection. Let π' be a projection from K to Y, and let $x_i \in K$. Let $\sigma \in \{0,1\}$. We have

$$\begin{aligned}
&\Pr\left[\boldsymbol{\pi}'(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}'_{\mathsf{var}(\boldsymbol{\pi}'(x_i))\leftarrow\sigma} = \pi' \;\middle|\; \mathcal{I}\right] \\
&= \Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}_{\mathsf{var}(\boldsymbol{\pi}(x_i))\leftarrow\sigma}|_K = \pi' \text{ and } \mathcal{I}\right] / \Pr[\mathcal{I}] \\
&= \sum_{\substack{\pi:\, \pi|_K=\pi', \pi^{-1}(y_j)=I^+,\\ \pi^{-1}(\overline{y_j})=I^-}} \Pr\left[\boldsymbol{\pi}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}|_{\mathsf{var}(\boldsymbol{\pi}(x_i))\leftarrow\sigma} = \pi\right] / \Pr[\mathcal{I}] \\
&\le \sum_{\substack{\pi:\, \pi|_K=\pi', \pi^{-1}(y_j)=I^+,\\ \pi^{-1}(\overline{y_j})=I^-}} q_\sigma \cdot \Pr[\boldsymbol{\pi} = \pi] / \Pr[\mathcal{I}] && \text{(since } \boldsymbol{\pi} \text{ is } (q_0,q_1)\text{-fixing)} \\
&= q_\sigma \cdot \Pr[\boldsymbol{\pi}|_K = \pi' \text{ and } \mathcal{I}] / \Pr[\mathcal{I}] \\
&= q_\sigma \cdot \Pr[\boldsymbol{\pi}' = \pi' \mid \mathcal{I}],
\end{aligned}$$

as required. □

We now prove that $\boldsymbol{\pi}'$ remains a (q_0,q_1)-fixing projection when conditioning on $f_1|_{\boldsymbol{\pi}} = y_j$ in addition to $\mathcal{I}$. The crucial observation is that the event $f_1|_{\boldsymbol{\pi}} = y_j$ is essentially a *filter*, defined next.

Definition 4.10. A set of projections $\mathcal{E}$ from $x_1,\dots,x_n$ to $y_1,\dots,y_m$ is a *filter* if it is closed under assignment to variables, i. e., if for every $\pi \in \mathcal{E}$, every variable y_j, and every bit $\tau \in \{0,1\}$, we have $\pi_{y_j\leftarrow\tau} \in \mathcal{E}$.

It turns out that the property of a projection being a (q_0,q_1)-fixing projection is preserved when conditioning on filters. Formally:

Proposition 4.11. *Let $\mathcal{E}$ be a filter and let $\boldsymbol{\pi}^*$ be a (q_0,q_1)-fixing projection. Then, $\boldsymbol{\pi}^*|\mathcal{E}$ is a (q_0,q_1)-fixing projection.*

Proof. Let π^* be a projection, and let x_i be a variable. Let $\sigma \in \{0,1\}$. We would like to prove that

$$\Pr\left[\boldsymbol{\pi}^*(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}^*_{\mathsf{var}(\boldsymbol{\pi}^*(x_i))\leftarrow\sigma} = \pi^* \;\middle|\; \boldsymbol{\pi}^* \in \mathcal{E}\right] \le q_\sigma \cdot \Pr[\boldsymbol{\pi}^* = \pi^* \mid \boldsymbol{\pi}^* \in \mathcal{E}].$$

If $\pi^* \notin \mathcal{E}$, then both sides of the equation are equal to zero: this is obvious for the right-hand side, and holds for the left-hand side since if there is a projection $\pi^0 \in \mathcal{E}$ and a variable y_j such that $\pi^0_{y_j\leftarrow\sigma} = \pi^*$ then it must be the case that $\pi^* \in \mathcal{E}$ by the definition of a filter. Thus, we may assume that $\pi^* \in \mathcal{E}$. Now, it holds that

$$\begin{aligned}
&\Pr\left[\boldsymbol{\pi}^*(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}^*_{\mathsf{var}(\boldsymbol{\pi}^*(x_i))\leftarrow\sigma} = \pi^* \;\middle|\; \boldsymbol{\pi}^* \in \mathcal{E}\right] \\
&\le \Pr\left[\boldsymbol{\pi}^*(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}^*_{\mathsf{var}(\boldsymbol{\pi}^*(x_i))\leftarrow\sigma} = \pi^*\right] / \Pr[\boldsymbol{\pi}^* \in \mathcal{E}] \\
&\le q_\sigma \cdot \Pr[\boldsymbol{\pi}^* = \pi^*] / \Pr[\boldsymbol{\pi}^* \in \mathcal{E}] && (\boldsymbol{\pi}^* \text{ is } (q_0,q_1)\text{-fixing}) \\
&= q_\sigma \cdot \Pr[\boldsymbol{\pi}^* = \pi^* | \boldsymbol{\pi}^* \in \mathcal{E}], && (\pi^* \in \mathcal{E})
\end{aligned}$$

as required. □

Consider the event $f_1|_{\boldsymbol{\pi}} = y_j$. Viewed as a set of projections from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$, this event is not a filter, since it is not closed under assignments to y_j. However, this event is closed under assignments *to all variables except* y_j: when $f_1|_{\boldsymbol{\pi}} = y_j$, the equality continues to hold even if the variables in Y are fixed to constants. Moreover, observe that conditioned on $\mathcal{I}$, the event $f_1|_{\boldsymbol{\pi}} = y_j$ depends only on the values that $\boldsymbol{\pi}$ assigns to K. Thus, we can view the event $f_1|_{\boldsymbol{\pi}} = y_j$ as a set of projections from K to Y, and taking this view, *this event is a filter*. Since $\boldsymbol{\pi}'$ is a (q_0, q_1)-fixing projection from K to $\{y_1, \ldots, y_m\} \setminus \{y_j\}$ when conditioned on $\mathcal{I}$, we conclude that it is a (q_0, q_1)-fixing projection when conditioned on both $\mathcal{I}$ and $f_1|_{\boldsymbol{\pi}} = y_j$. It follows by Lemma 4.1 that

$$\Pr\left[L(f_{2,I^+,I^-}|_{\boldsymbol{\pi}'}) = 1 \;\middle|\; f_1|_{\boldsymbol{\pi}} = y_j, \boldsymbol{I}^+ = I^+, \boldsymbol{I}^- = I^-\right] \le q \cdot \sqrt{L(f_{2,I^+,I^-})} \le q \cdot \sqrt{L(f_2)},$$

as required.

4.3 Proof of Lemma 4.6

Finally, we prove Lemma 4.6, restated next.

Lemma 4.6. *For every projection π we have*

$$L(\phi|_\pi) \le \sum_{\text{internal gate } g \text{ of } \phi} (\mathrm{depth}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1_{L(\phi|_\pi)=1}. \tag{4.5}$$

Let π be a projection. We prove that π satisfies Equation 4.5. Recall that $\mathcal{E}_{g,\pi}$ is the event that there exists some literal y_j^σ such that $\mathrm{left}(g)|_\pi = y_j^\sigma$ and

- $L(\mathrm{right}(g)|_{\pi_{y_j \leftarrow \sigma}}) = 1$ if g is an OR gate, or
- $L(\mathrm{right}(g)|_{\pi_{y_j \leftarrow \overline{\sigma}}}) = 1$ if g is an AND gate.

We prove Equation 4.5 by induction. If ϕ consists of a single leaf, then the upper bound clearly holds. Otherwise, the root of ϕ is an internal gate. Without loss of generality, assume that the root is an OR gate. We denote the subformulas rooted at the left and right children of the root by ϕ_ℓ and ϕ_r, respectively. We consider several cases:

- If $L(\phi|_\pi) = 1$, then the upper bound clearly holds.
- Suppose that $L(\phi_\ell|_\pi) \ge 2$. By the subadditivity of formula complexity (Fact 2.3), we have

$$L(\phi|_\pi) \le L(\phi_\ell|_\pi) + L(\phi_r|_\pi).$$

By the induction hypothesis, we have

$$\begin{aligned}
L(\phi_\ell|_\pi) &\leq \sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi_\ell}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1_{L(\phi_\ell|_\pi)=1} \\
&= \sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi_\ell}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} && (L(\phi_\ell|_\pi) \geq 2) \\
&= \sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi}(g) + 1) \cdot 1_{\mathcal{E}_{g,\pi}},
\end{aligned}$$

where the equality holds since $\text{depth}_{\phi_\ell}(g) = \text{depth}_\phi(g) - 1$ for every gate g of ϕ_ℓ. Since $L(\phi_\ell|_\pi) \geq 2$, at least one of the terms in the last sum must be non-zero, so it holds that

$$\sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi}(g) + 1) \cdot 1_{\mathcal{E}_{g,\pi}} \leq \sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} - 1.$$

Next, by the induction hypothesis we have

$$\begin{aligned}
L(\phi_r|_\pi) &\leq \sum_{\text{internal gate } g \text{ of } \phi_r} (\text{depth}_{\phi_r}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1_{L(\phi_r|_\pi)=1} \\
&\leq \sum_{\text{internal gate } g \text{ of } \phi_r} (\text{depth}_{\phi}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1.
\end{aligned}$$

By combining the two bounds, we get that

$$\begin{aligned}
L(\phi|_\pi) &\leq L(\phi_\ell|_\pi) + L(\phi_r|_\pi) \\
&\leq \sum_{\text{internal gate } g \text{ of } \phi_\ell} (\text{depth}_{\phi}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} - 1 \\
&\qquad + \sum_{\text{internal gate } g \text{ of } \phi_r} (\text{depth}_{\phi}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1 \\
&\leq \sum_{\text{internal gate } g \text{ of } \phi} (\text{depth}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} \\
&\leq \sum_{\text{internal gate } g \text{ of } \phi} (\text{depth}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} + 1_{L(\phi|_\pi)=1},
\end{aligned}$$

as required.

- If $L(\phi_r|_\pi) \geq 2$, then we use the same argument of the previous case by exchanging ϕ_ℓ and ϕ_r.
- Suppose that $L(\phi|_\pi) \geq 2$, $L(\phi_\ell|_\pi) \leq 1$ and $L(\phi_r|_\pi) \leq 1$. Then, it must be the case that $L(\phi|_\pi) = 2$ and also that $L(\phi_\ell|_\pi) = 1$ and $L(\phi_r|_\pi) = 1$ (or otherwise $L(\phi|_\pi) = 1$). In

particular, $\phi_\ell|_\pi$ is equal to some literal y_j^σ. It follows that $\phi|_\pi = y_j^\sigma \vee \phi_r|_\pi$, and by the one-variable simplification rules (Fact 4.5), this function is equal to

$$y_j^\sigma \vee \left(\phi_r|_\pi\right)_{y_j \leftarrow \sigma} = y_j^\sigma \vee \phi_r|_{\pi_{y_j \leftarrow \sigma}}.$$

Thus, it must be the case that $L(\phi_r|_{\pi_{y_j \leftarrow \sigma}}) = 1$ (since $L(\phi|_\pi) = 2$). It follows that if we let g be the root of ϕ, then the event $\mathcal{E}_{g,\pi}$ occurs and so

$$(\mathrm{depth}(g) + 2) \cdot 1_{\mathcal{E}_{g,\pi}} = 2 = L(\phi|_\pi),$$

so the desired upper bound holds.

We proved that the upper bound holds in each of the possible cases, so the required result follows.

5 Proof of shrinkage theorem for hiding projections

In this section we prove our the shrinkage theorem for hiding projections. We start by recalling the definition of hiding projections and their shrinkage theorem.

Definition 3.9. Let $0 \leq q_0, q_1 \leq 1$. We say that a random projection $\boldsymbol{\pi}$ is a (q_0, q_1)-*hiding projection* if for every projection π, every bit $\sigma \in \{0, 1\}$, and all variables x_i, y_j, we have

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \;\middle|\; \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi\right] \leq q_\sigma,$$

whenever the event conditioned on has positive probability. For shorthand, we say that $\boldsymbol{\pi}$ is a q-hiding projection, for $q = \sqrt{q_0 q_1}$.

Theorem 3.12 (Shrinkage under hiding projections)**.** *Let ϕ be a formula of size s and depth d, and let $\boldsymbol{\pi}$ be a q-hiding projection. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right] = O\left(m^4 \cdot q^2 \cdot d^2 \cdot s + m^2 \cdot q \cdot \sqrt{s}\right).$$

The crux of the proof of Theorem 3.12 is that hiding projections can be converted into fixing projections. This is captured by the following result.

Lemma 5.1. *Let $\boldsymbol{\pi}$ be a q-hiding projection from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$. Then, there exists a $(4m^2 \cdot q)$-fixing projection $\boldsymbol{\pi}'$ from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$ and an event $\mathcal{H}$ of probability at least $\frac{1}{2}$ such that $\boldsymbol{\pi}' = \boldsymbol{\pi}$ conditioned on $\mathcal{H}$. Furthermore, the event $\mathcal{H}$ is independent of $\boldsymbol{\pi}$.*

We prove Lemma 5.1 in Section 5.1. By combining Lemma 5.1 with our shrinkage theorem for fixing projections, we obtain the following shrinkage theorem for hiding projections.

Theorem 3.12. *Let ϕ be a formula of size s and depth d, and let $\boldsymbol{\pi}$ be a q-hiding projection. Then*

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right] = O\left(m^4 \cdot q^2 \cdot d^2 \cdot s + m^2 \cdot q \cdot \sqrt{s}\right).$$

Proof. Let $\boldsymbol{\pi}'$ and $\mathcal{H}$ be the $(4m^2 \cdot q)$-fixing projection and event that are obtained from $\boldsymbol{\pi}$ by Lemma 5.1. Since $\Pr[\mathcal{H}] \geq \frac{1}{2}$, we have

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}'})\right] \geq \Pr[\mathcal{H}] \cdot \mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}'}) \,\middle|\, \mathcal{H}\right] \geq \frac{1}{2} \cdot \mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}}) \,\middle|\, \mathcal{H}\right] = \frac{1}{2} \cdot \mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}})\right],$$

where the second inequality holds since conditioned on $\mathcal{H}$ it holds that $\boldsymbol{\pi}' = \boldsymbol{\pi}$, and the last equality holds since the event $\mathcal{H}$ is independent of $\boldsymbol{\pi}$. On the other hand, by applying Theorem 3.6 to $\boldsymbol{\pi}'$, we obtain that

$$\mathbb{E}\left[L(\phi|_{\boldsymbol{\pi}'})\right] = O(m^4 \cdot q^2 \cdot d^2 \cdot s + m^2 \cdot q \cdot \sqrt{s}).$$

The theorem follows by combining the two bounds. □

5.1 Proof of Lemma 5.1

We use the following straightforward generalization of the property of hiding projections.

Claim 5.2. *Let $\boldsymbol{\pi}$ be a (q_0, q_1)-hiding projection, and let $\boldsymbol{\mathcal{E}}$ be a random set of projections that is independent of $\boldsymbol{\pi}$. Then, for every $\sigma \in \{0,1\}$, we have*

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} \in \boldsymbol{\mathcal{E}}\right] \leq q_\sigma.$$

Proof. Let $\boldsymbol{\pi}$ and $\boldsymbol{\mathcal{E}}$ be as in the claim, and let $\sigma \in \{0,1\}$. We have

$$\begin{aligned}
&\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} \in \boldsymbol{\mathcal{E}}\right] \\
&= \sum_{\pi} \Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi \text{ and } \pi \in \boldsymbol{\mathcal{E}}\right] \cdot \Pr\left[\boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} \in \boldsymbol{\mathcal{E}}\right] \\
&= \sum_{\pi} \Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi\right] \cdot \Pr\left[\boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} \in \boldsymbol{\mathcal{E}}\right] \\
&\hspace{20em} (\boldsymbol{\pi} \text{ and } \boldsymbol{\mathcal{E}} \text{ are independent}) \\
&\leq \sum_{\pi} q_\sigma \cdot \Pr\left[\boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi \,\middle|\, \boldsymbol{\pi}_{y_j \leftarrow \sigma} \in \boldsymbol{\mathcal{E}}\right] \qquad (\boldsymbol{\pi} \text{ is } (q_0, q_1)\text{-hiding}) \\
&= q_\sigma.
\end{aligned}$$

□

We turn to proving Lemma 5.1, restated next.

Lemma 5.1. *Let $\boldsymbol{\pi}$ be a q-hiding projection from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$. Then, there exists a $(4m^2 \cdot q)$-fixing projection $\boldsymbol{\pi}'$ from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$ and an event $\mathcal{H}$ of probability at least $\frac{1}{2}$ such that $\boldsymbol{\pi}' = \boldsymbol{\pi}$ conditioned on $\mathcal{H}$. Furthermore, the event $\mathcal{H}$ is independent of $\boldsymbol{\pi}$.*

Suppose that $\boldsymbol{\pi}$ is a (q_0, q_1)-hiding projection from $x_1, \ldots, x_n$ to $y_1, \ldots, y_m$. Let $\boldsymbol{\rho}$ be a $\left(1 - \frac{1}{2m}\right)$-random restriction over $y_1, \ldots, y_m$, i. e., $\boldsymbol{\rho}$ is the random projection from $y_1, \ldots, y_m$ to $y_1, \ldots, y_m$ that assigns each y_j independently as follows:

$$\boldsymbol{\rho}(y_j) = \begin{cases} y_j & \text{with probability } 1 - \frac{1}{2m}, \\ 0 & \text{with probability } \frac{1}{4m}, \\ 1 & \text{with probability } \frac{1}{4m}. \end{cases}$$

For convenience, we define $\boldsymbol{\rho}(0) = 0$, $\boldsymbol{\rho}(1) = 1$, and $\boldsymbol{\rho}(\overline{y_j}) = \overline{\boldsymbol{\rho}(y_j)}$ for every $j \in [m]$. We now choose the random projection $\boldsymbol{\pi}'$ to be the composition $\boldsymbol{\rho} \circ \boldsymbol{\pi}$, and define the event $\mathcal{H}$ to be the event that $\boldsymbol{\rho}(y_j) = y_j$ for every $j \in [m]$. Observe that the event $\boldsymbol{\pi}' = \boldsymbol{\pi}$ occurs whenever $\mathcal{H}$ occurs, and moreover

$$\Pr[\mathcal{H}] = \left(1 - \frac{1}{2m}\right)^m \geq 1 - \frac{m}{2m} = \frac{1}{2},$$

as required. Moreover, the event $\mathcal{H}$ is independent of $\boldsymbol{\pi}$, since $\boldsymbol{\rho}$ is independent of $\boldsymbol{\pi}$. We show that $\boldsymbol{\pi}'$ is a $(4 \cdot m^2 \cdot q_0, 4 \cdot m^2 \cdot q_1)$-fixing projection. To this end, we should show that for every projection π', every bit $\sigma \in \{0,1\}$, and every variable x_i,

$$\Pr\left[\boldsymbol{\pi}'(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\pi}'_{\mathsf{var}(\pi'(x_i)) \leftarrow \sigma} = \pi'\right] \leq 4 \cdot m^2 \cdot q_\sigma \cdot \Pr[\boldsymbol{\pi}' = \pi'] \,. \tag{5.1}$$

This is implied by the following inequality, which we will prove for all $j \in [m]$:

$$\Pr\left[\boldsymbol{\pi}'(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\pi}'_{y_j \leftarrow \sigma} = \pi'\right] \leq 4 \cdot m \cdot q_\sigma \cdot \Pr[\boldsymbol{\pi}' = \pi'] \,. \tag{5.2}$$

Let $j \in [m]$ and let $\boldsymbol{\rho}_{-j}$ be the random projection obtained by restricting $\boldsymbol{\rho}$ to the domain $\{y_1, \ldots, y_m\} \setminus \left\{y_j\right\}$. First, observe that if $\boldsymbol{\rho}(y_j) = \sigma$ then $\boldsymbol{\pi}' = \boldsymbol{\rho} \circ \boldsymbol{\pi} = \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma}$, and therefore

$$\Pr[\boldsymbol{\pi}' = \pi'] \geq \Pr\left[\boldsymbol{\rho}(y_j) = \sigma \text{ and } \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] = \frac{1}{4m} \cdot \Pr\left[\boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] ,$$

where the equality holds since $\boldsymbol{\rho}(y_j)$ is independent of $\boldsymbol{\rho}_{-j}$ and $\boldsymbol{\pi}$. In the rest of this section we will prove the following inequality, which together with the last inequality will imply Equation 5.2:

$$\Pr\left[\boldsymbol{\pi}'(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\pi}'_{y_j \leftarrow \sigma} = \pi'\right] \leq q_\sigma \cdot \Pr\left[\boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] . \tag{5.3}$$

To this end, observe that the event $\boldsymbol{\pi}'(x_i) \in \left\{y_j, \overline{y_j}\right\}$ happens if and only if $\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\}$ and $\boldsymbol{\rho}(y_j) = y_j$, and therefore

$$\begin{aligned}
&\Pr\left[\boldsymbol{\pi}'(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\pi}'_{y_j \leftarrow \sigma} = \pi'\right] \\
&= \Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\rho}(y_j) = y_j \text{ and } \boldsymbol{\pi}'_{y_j \leftarrow \sigma} = \pi'\right] .
\end{aligned}$$

Next, observe that if $\boldsymbol{\rho}(y_j) = y_j$ then $\boldsymbol{\pi}'_{y_j \leftarrow \sigma} = \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma}$. It follows that the latter expression is equal to

$$\begin{aligned}
&\Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\rho}(y_j) = y_j \text{ and } \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] \\
&\leq \Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \text{ and } \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] \\
&= \Pr\left[\boldsymbol{\pi}(x_i) \in \left\{y_j, \overline{y_j}\right\} \;\middle|\; \boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] \cdot \Pr\left[\boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] \\
&\leq q_\sigma \cdot \Pr\left[\boldsymbol{\rho}_{-j} \circ \boldsymbol{\pi}_{y_j \leftarrow \sigma} = \pi'\right] ,
\end{aligned}$$

where the last inequality follows by applying Claim 5.2 with $\boldsymbol{\mathcal{E}}$ being the set of projections π that satisfy $\boldsymbol{\rho}_{-j} \circ \pi = \pi'$. This concludes the proof of Equation 5.3.

Remark 5.3. We can improve the bound m^2 in the statement of Lemma 5.1 to mk, where k is the maximal number of variables $1, \dots, m$ which could appear in any position of $\boldsymbol{\pi}$. The reason is that in the latter case, the transition from Equation 5.2 to Equation 5.1 incurs a factor of k rather than m. This is useful, for example, since the random projections $\boldsymbol{\pi}$ of Section 8.1 have the feature that for each $i \in [n]$ there is a unique $j \in [m]$ such that $\boldsymbol{\pi}(x_i) \in \{0, 1, y_j, \overline{y_j}\}$, and so for these projections $k = 1$.

6 Joining projections

In this section, we define a *join* operation on fixing and hiding projections, and show that it preserves the corresponding properties. This operation provides a convenient tool for constructing random projections, and will be used in our applications.

Definition 6.1. Let α be a projection from $x_1, \dots, x_{n_a}$ to $y_1, \dots, y_{m_a}$, and let β be a projection from $w_1, \dots, w_{n_b}$ to $z_1, \dots, z_{m_b}$. The *join* $\alpha \uplus \beta$ is the projection from $x_1, \dots, x_{n_a}, w_1, \dots, w_{n_b}$ to $y_1, \dots, y_{m_a}, z_1, \dots, z_{m_b}$ obtained by joining α and β together in the obvious way.

Lemma 6.2. *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be independent random projections. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are (q_0, q_1)-fixing projections, then so is $\boldsymbol{\alpha} \uplus \boldsymbol{\beta}$.*

Proof. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be (q_0, q_1)-fixing projections, and let $\boldsymbol{\gamma} = \boldsymbol{\alpha} \uplus \boldsymbol{\beta}$. We prove that $\boldsymbol{\gamma}$ is a (q_0, q_1)-fixing projection. Let α be a projection from $x_1, \dots, x_{n_a}$ to $y_1, \dots, y_{m_a}$, let β be a projection from $w_1, \dots, w_{n_b}$ to $z_1, \dots, z_{m_b}$, and let $\gamma = \alpha \uplus \beta$. We should show that for every $\sigma \in \{0, 1\}$ and every input variable u (either x_i or w_i) we have

$$\Pr\left[\boldsymbol{\gamma}(u) \notin \{0,1\} \text{ and } \boldsymbol{\gamma}_{\mathsf{var}(\boldsymbol{\gamma}(u)) \leftarrow \sigma} = \gamma\right] \le q_\sigma \cdot \Pr\left[\boldsymbol{\gamma} = \gamma\right].$$

Let $\sigma \in \{0, 1\}$ be a bit, and let u be an input variable. Assume that $u = x_i$ for some $i \in [n_a]$ (if $u = w_i$ for some $i \in [n_b]$, the proof is similar). The above equation is therefore equivalent to

$$\Pr\left[\boldsymbol{\alpha}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\alpha}_{\mathsf{var}(\boldsymbol{\alpha}(x_i)) \leftarrow \sigma} = \alpha \text{ and } \boldsymbol{\beta}_{\mathsf{var}(\boldsymbol{\alpha}(x_i)) \leftarrow \sigma} = \beta\right] \le q_\sigma \cdot \Pr\left[\boldsymbol{\alpha} = \alpha \text{ and } \boldsymbol{\beta} = \beta\right].$$

Since the ranges of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are disjoint, the variable $\mathsf{var}(\boldsymbol{\alpha}(x_i))$ does not appear in the range of $\boldsymbol{\beta}$, and therefore $\boldsymbol{\beta}_{\mathsf{var}(\boldsymbol{\alpha}(x_i)) \leftarrow \sigma} = \boldsymbol{\beta}$. The independence of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ shows that the above inequality is equivalent to

$$\Pr\left[\boldsymbol{\alpha}(x_i) \notin \{0,1\} \text{ and } \boldsymbol{\alpha}_{\mathsf{var}(\boldsymbol{\alpha}(x_i)) \leftarrow \sigma} = \alpha\right] \cdot \Pr\left[\boldsymbol{\beta} = \beta\right] \le q_\sigma \cdot \Pr[\boldsymbol{\alpha} = \alpha] \cdot \Pr\left[\boldsymbol{\beta} = \beta\right],$$

which follows from $\boldsymbol{\alpha}$ being (q_0, q_1)-fixing. □

Lemma 6.3. *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be independent random projections. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are (q_0, q_1)-hiding projections, then so is $\boldsymbol{\alpha} \uplus \boldsymbol{\beta}$.*

Proof. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be (q_0, q_1)-hiding projections, and let $\boldsymbol{\gamma} = \boldsymbol{\alpha} \uplus \boldsymbol{\beta}$. We prove that $\boldsymbol{\gamma}$ is a (q_0, q_1)-hiding projection. Let α be a projection from $x_1, \dots, x_{n_a}$ to $y_1, \dots, y_{m_a}$, let β be a projection from

$w_1, \ldots, w_{n_b}$ to $z_1, \ldots, z_{m_b}$, and let $\gamma = \alpha \uplus \beta$. We should show that for every $\sigma \in \{0,1\}$, every input variable u (either x_i or w_i), and every output variable v (either y_j or z_j) we have

$$\Pr\left[\boldsymbol{\gamma}(u) \in \{v, \overline{v}\} \;\middle|\; \boldsymbol{\gamma}_{v \leftarrow \sigma} = \gamma\right] \leq q_\sigma.$$

Let $\sigma \in \{0,1\}$ be a bit, let u be an input variable, and let v be an output variable. Assume that $u = x_i$ for some $i \in [n_a]$ (if $u = w_i$ for some $i \in [n_b]$, the proof is similar). In this case, we may assume that $v = y_j$ for some $j \in [m_a]$, since otherwise the probability on the left-hand side is 0. Thus, the above equation is equivalent to

$$\Pr\left[\boldsymbol{\alpha}(x_i) \in \left\{y_j, \overline{y_j}\right\} \;\middle|\; \boldsymbol{\alpha}_{y_j \leftarrow \sigma} = \alpha \text{ and } \boldsymbol{\beta}_{y_j \leftarrow \sigma} = \beta\right] \leq q_\sigma.$$

Since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are independent, whenever the conditioned event has positive probability, we have

$$\Pr\left[\boldsymbol{\alpha}(x_i) \in \left\{y_j, \overline{y_j}\right\} \;\middle|\; \boldsymbol{\alpha}_{y_j \leftarrow \sigma} = \alpha \text{ and } \boldsymbol{\beta}_{y_j \leftarrow \sigma} = \beta\right] = \Pr\left[\boldsymbol{\alpha}(x_i) \in \left\{y_j, \overline{y_j}\right\} \;\middle|\; \boldsymbol{\alpha}_{y_i \leftarrow \sigma} = \alpha\right] \leq q_\sigma,$$

where the inequality holds since $\boldsymbol{\alpha}$ is (q_0, q_1)-hiding. □

7 Hiding projections from complexity measures

In this section we define a generalization of the unweighted quantum adversary bound due to Ambainis [2], and show how it can be used for constructing hiding projections. We also derive a relationship between this measure to the Khrapchenko bound [33], and conclude that the latter bound can be used for constructing hiding projections as well.

7.1 The soft-adversary method

Ambainis [2] defined the following complexity measure for Boolean functions, called the *unweighted quantum adversary bound*, and proved that it is a lower bound on quantum query complexity (for a definition of quantum query complexity, see [2] or [9]).

Definition 7.1. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant function. Let $R \subseteq f^{-1}(1) \times f^{-1}(0)$, and let A, B be the projections of R into the first and second coordinates, respectively. Let $R(a,B) = \{b \in B : (a,b) \in R\}$, let $R_i(a,B) = \{b \in B : (a,b) \in R, a_i \neq b_i\}$, and define $R(A,b), R_i(A,b)$ analogously. The *unweighted quantum adversary bound of f* is

$$\mathrm{Adv}_u(f) = \max_{R \subseteq f^{-1}(1) \times f^{-1}(0)} \sqrt{\frac{\min_{a \in A} |R(a,B)| \cdot \min_{b \in B} |R(A,b)|}{\max_{\substack{a \in A \\ i \in [n]}} |R_i(a,B)| \cdot \max_{\substack{b \in B \\ i \in [n]}} |R_i(A,b)|}}.$$

Theorem 7.2 ([2]). *The quantum query complexity of a non-constant function $f\colon \{0,1\}^n \to \{0,1\}$ is $\Omega(\mathrm{Adv}_u(f))$.*

We define the following generalization of the unweighted quantum adversary bound.

Definition 7.3. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function. We define the *soft-adversary bound of f*, which we denote Adv_s, to be the maximum of the quantity

$$\sqrt{\min_{\substack{a \in \mathrm{supp}(\boldsymbol{a}) \\ i \in [n]}} \frac{1}{\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a]} \cdot \min_{\substack{b \in \mathrm{supp}(\boldsymbol{b}) \\ i \in [n]}} \frac{1}{\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{b} = b]}}$$

over all distributions $(\boldsymbol{a}, \boldsymbol{b})$ supported on $f^{-1}(1) \times f^{-1}(0)$.

We chose to call this notion "soft-adversary bound" since we view it as a variant of the unweighted adversary bound in which, instead of choosing for each pair (a,b) whether it belongs to R or not (a "hard" decision), we assign each pair (a,b) a probability of being in the relation R (a "soft" decision). We now show that this bound indeed generalizes the unweighted quantum adversary method.

Proposition 7.4. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be non-constant function. Then $\mathrm{Adv}_s(f) \geq \mathrm{Adv}_u(f)$.*

Proof. Let $R \subseteq f^{-1}(1) \times f^{-1}(0)$ be a relation that attains $\mathrm{Adv}_u(f)$. Let $(\boldsymbol{a}, \boldsymbol{b})$ be a uniformly distributed element of R. Using $\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a] = |R_i(a,B)|/|R(a,B)|$ and $\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{b} = b] = |R_i(A,b)|/|R(A,b)|$, it is easy to check that the quantity in the definition of $\mathrm{Adv}_s(f)$ is at least $\mathrm{Adv}_u(f)$. □

Following [53, 38], it can be shown that the soft-adversary bound is a lower bound on formula complexity. We have the following result, which is proved in Section A.

Proposition 7.5. *For any non-constant function $f\colon \{0,1\}^n \to \{0,1\}$ we have $L(f) \geq \mathrm{Adv}_s^2(f)$.*

Our motivation for introducing the soft-adversary bound is that it can be used for constructing hiding projections. We have the following result.

Lemma 7.6. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant Boolean function. There is a q-hiding projection $\boldsymbol{\pi}$ to a single variable y such that $q = 1/\mathrm{Adv}_s(f)$ and such that $f|_{\boldsymbol{\pi}}$ is a non-constant function with probability* 1.

Proof. Let $(\boldsymbol{a}, \boldsymbol{b})$ be a distribution supported on $f^{-1}(1) \times f^{-1}(0)$ which attains $\mathrm{Adv}_s(f)$. We construct a q-hiding projection $\boldsymbol{\pi}$ from $x_1, \ldots, x_n$ to a single variable y for $q = 1/\mathrm{Adv}_s(f)$, such that $\boldsymbol{\pi}_{y \leftarrow 1} = \boldsymbol{a}$ and $\boldsymbol{\pi}_{y \leftarrow 0} = \boldsymbol{b}$. Note that this implies in particular that $f|_{\boldsymbol{\pi}}$ is a non-constant function, since $\boldsymbol{a} \in f^{-1}(1)$ and $\boldsymbol{b} \in f^{-1}(0)$. Specifically, for every input variable x_i of f, we define $\boldsymbol{\pi}$ as follows:

- If $\boldsymbol{a}_i = \boldsymbol{b}_i$ then $\boldsymbol{\pi}(x_i) = \boldsymbol{a}_i$.
- If $\boldsymbol{a}_i = 1$ and $\boldsymbol{b}_i = 0$ then $\boldsymbol{\pi}(x_i) = y$.
- If $\boldsymbol{a}_i = 0$ and $\boldsymbol{b}_i = 1$ then $\boldsymbol{\pi}(x_i) = \overline{y}$.

It is not hard to see that indeed $\boldsymbol{\pi}_{y\leftarrow 1} = \boldsymbol{a}$ and $\boldsymbol{\pi}_{y\leftarrow 0} = \boldsymbol{b}$. We now show that $\boldsymbol{\pi}$ is $1/\mathrm{Adv}_s(f)$-hiding. To this end, we show that $\boldsymbol{\pi}$ is (q_0, q_1)-hiding for

$$q_0 = \max_{\substack{b \in \mathrm{supp}(\boldsymbol{b}) \\ i \in [n]}} \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{b} = b], \qquad q_1 = \max_{\substack{a \in \mathrm{supp}(\boldsymbol{a}) \\ i \in [n]}} \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a].$$

Note that indeed $\sqrt{q_0 q_1} = 1/\mathrm{Adv}_s(f)$. In order to prove that $\boldsymbol{\pi}$ is (q_0, q_1)-hiding, we need to prove that

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \{y, \overline{y}\} \,\middle|\, \boldsymbol{\pi}_{y\leftarrow\sigma} = \pi\right] \leq q_\sigma$$

for every $i \in [n]$ and every $\sigma \in \{0,1\}$. To this end, observe that the event $\boldsymbol{\pi}(x_i) \in \{y, \overline{y}\}$ occurs if and only if $\boldsymbol{a}_i \neq \boldsymbol{b}_i$, and that $\boldsymbol{\pi}_{y\leftarrow 1}, \boldsymbol{\pi}_{y\leftarrow 0}$ are equal to the strings $\boldsymbol{a}, \boldsymbol{b}$ respectively. It follows that

$$\Pr\left[\boldsymbol{\pi}(x_i) \in \{y, \overline{y}\} \,\middle|\, \boldsymbol{\pi}_{y\leftarrow 1} = \pi\right] = \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = \pi] \leq q_1$$
$$\Pr\left[\boldsymbol{\pi}(x_i) \in \{y, \overline{y}\} \,\middle|\, \boldsymbol{\pi}_{y\leftarrow 0} = \pi\right] = \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{b} = \pi] \leq q_0$$

as required. □

Remark 7.7. Several generalizations of Ambainis' original quantum adversary bound have appeared in the literature. Špalek and Szegedy [53] showed that all of these methods are equivalent, and so they are known collectively as the strong quantum adversary bound. The formulation of the strong quantum adversary bound in Ambainis [3] makes it clear that it subsumes our soft version. Laplante, Lee and Szegedy [38] showed that the strong quantum adversary bound is a lower bound on formula complexity, and they came up with an even stronger measure, maxPI^2, that is still a lower bound on formula complexity, but no longer a lower bound on quantum query complexity.

Høyer, Lee and Špalek [28] generalized the strong quantum adversary bound, coming up with a new measure known as the general adversary bound, which is a lower bound both on quantum query complexity and (after squaring) on formula complexity. Reichardt [49, 50, 39, 51] showed that the general adversary bound in fact *coincides* with quantum query complexity (up to constant factors). These results are described in a recent survey by Li and Shirley [40].

Remark 7.8. We note that $\mathrm{Adv}_s(f) \leq n$ (always). In order to prove this, we show that each factor in Definition 7.3 is at most n. In particular, observe that it always holds that $\boldsymbol{a} \neq \boldsymbol{b}$, and hence for every possible value a of $\boldsymbol{a}$, we have

$$\sum_{i=1}^{n} \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a] \geq \Pr[\exists i \in [n] \text{ s.t. } \boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a] = 1.$$

By averaging, there must be a coordinate $i \in [n]$ such that $\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a] \geq \frac{1}{n}$, and therefore

$$\min_{\substack{a \in \mathrm{supp}(\boldsymbol{a}) \\ i \in [n]}} \frac{1}{\Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a]} \leq n\,.$$

The upper bound for the other factor in Definition 7.3 can be proved similarly.

7.2 Khrapchenko bound

Khrapchenko [33] defined the following complexity measure for Boolean functions, and proved that it is a lower bound on formula complexity.

Definition 7.9. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant function. For every two sets $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$, let $E(A,B)$ be the set of pairs $(a,b) \in A \times B$ such that a and b differ on exactly one coordinate. The *Khrapchenko bound of f* is

$$\mathrm{Khr}(f) = \max_{A \subseteq f^{-1}(1), B \subseteq f^{-1}(0)} \frac{|E(A,B)|^2}{|A| \cdot |B|}.$$

Theorem 7.10 ([33])**.** *For every non-constant function $f\colon \{0,1\}^n \to \{0,1\}$ we have $L(f) \geq \mathrm{Khr}(f)$.*

Laplante, Lee, and Szegedy [38] observed that the strong adversary method generalizes the Khrapchenko bound. We show that this holds even for the unweighted adversary bound, up to a constant factor. Specifically, we have the following result, which is proved in Section C.

Proposition 7.11. *For any non-constant function $f\colon \{0,1\}^n \to \{0,1\}$ we have $\mathrm{Adv}_u(f) \geq \frac{\sqrt{\mathrm{Khr}(f)}}{4}$.*

By combining Propositions 7.4 and 7.11 with Lemma 7.6, it follows that the Khrapchenko bound too can be used for constructing hiding projections.

Corollary 7.12. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant Boolean function. There is a q-hiding projection $\boldsymbol{\pi}$ to a single variable y such that $q = \sqrt{4/\mathrm{Khr}(f)}$ and such that $f|_{\boldsymbol{\pi}}$ is a non-constant function with probability 1.*

In Section B, we describe a new complexity measure, *the min-entropy Khrapchenko bound*, which generalizes the Khrapchenko bound in the same way the soft-adversary bound generalizes the unweighted adversary bound. As we show there, the min-entropy Khrapchenko bound provides a simple way to prove lower bounds on the soft-adversary bound.

8 Applications

In this section, we apply our shrinkage theorems to obtain new results regarding the KRW conjecture and the formula complexity of $\boldsymbol{AC}^0$. First, in Section 8.1, we prove the KRW conjecture for inner functions for which the soft-adversary bound is tight. We use this version of the KRW conjecture in Section 8.2 to prove cubic formula lower bounds for a function in $\boldsymbol{AC}^0$. Finally, we rederive some closely related known results in Section 8.3.

8.1 Application to the KRW conjecture

Given two Boolean functions $f\colon \{0,1\}^m \to \{0,1\}$ and $g\colon \{0,1\}^n \to \{0,1\}$, their (block-)composition is the function $f \diamond g\colon \left(\{0,1\}^n\right)^m \to \{0,1\}$ defined by

$$(f \diamond g)(x_1, \ldots, x_m) = f\left(g(x_1), \ldots, g(x_m)\right),$$

where $x_1, \dots, x_m \in \{0,1\}^n$. It is easy to see that $L(f \diamond g) \le L(f) \cdot L(g)$. The KRW conjecture [31] asserts that this is roughly optimal, namely, that $L(f \diamond g) \gtrsim L(f) \cdot L(g)$. In this section, we prove the following related result.

Theorem 8.1. *Let $f\colon \{0,1\}^m \to \{0,1\}$ and $g\colon \{0,1\}^n \to \{0,1\}$ be non-constant Boolean functions. Then,*

$$L(f \diamond g)^{1+O\left(\frac{1}{\sqrt{\log L(f \diamond g)}}\right)} \ge \frac{1}{O(m^4)} \cdot L(f) \cdot \mathrm{Adv}_s^2(g).$$

Proof. Let $\boldsymbol{\pi}$ be the q-hiding projection constructed for g in Lemma 7.6 with $q = 1/\mathrm{Adv}_s(g)$, and recall that $g|_{\boldsymbol{\pi}}$ is non-constant with probability 1. Let $\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_m$ be independent copies of $\boldsymbol{\pi}$, and let $\boldsymbol{\pi}^m$ denote their m-fold join. Observe that $\boldsymbol{\pi}^m$ is q-hiding according to Lemma 6.3. Applying Corollary 3.13 and using the estimate $x + \sqrt{x} = O(x+1)$, we see that $s = L(f \diamond g)$ satisfies

$$\mathbb{E}\left[L\left((f \diamond g)|_{\boldsymbol{\pi}^m}\right)\right] = m^4 \cdot \frac{1}{\mathrm{Adv}_s^2(g)} \cdot s^{1+O\left(\frac{1}{\sqrt{\log s}}\right)} + O(1).$$

On the other hand, it is not hard to see that

$$\left((f \diamond g)|_{\boldsymbol{\pi}^m}\right)(x_1, \dots, x_m) = f\left(g|_{\boldsymbol{\pi}_1}(x_1), \dots, g|_{\boldsymbol{\pi}_m}(x_m)\right),$$

and since $g|_{\boldsymbol{\pi}_1}, \dots, g|_{\boldsymbol{\pi}_m}$ are non-constant, the function f reduces to $(f \diamond g)|_{\boldsymbol{\pi}^m}$. In particular,

$$L(f) \le \mathbb{E}\left[L\left((f \diamond g)|_{\boldsymbol{\pi}^m}\right)\right] = m^4 \cdot \frac{1}{\mathrm{Adv}_s^2(g)} \cdot s^{1+O\left(\frac{1}{\sqrt{\log s}}\right)} + O(1).$$

This means that there exists a constant $C > 0$ such that for all non-constant f and g

$$L(f) \le m^4 \cdot \frac{1}{\mathrm{Adv}_s^2(g)} \cdot L(f \diamond g)^{1+\frac{C}{\sqrt{\log L(f \diamond g)}}} + C.$$

We consider two cases depending on whether $L(f) \ge 2C$ or not. If $L(f) \ge 2C$, then the above inequality implies

$$\frac{L(f)}{2} \le m^4 \cdot \frac{1}{\mathrm{Adv}_s^2(g)} \cdot L(f \diamond g)^{1+\frac{C}{\sqrt{\log L(f \diamond g)}}},$$

and we obtain the theorem by rearranging.

If $L(f) < 2C$, then we have $L(f \diamond g) \ge L(g)$ since g or $\neg g$ is a sub-function of $f \circ g$ (as follows from the fact that f is non-constant). This means that in this case

$$L(f \diamond g) \ge L(g) > \frac{L(f) \cdot L(g)}{2C} \ge \frac{L(f) \cdot \mathrm{Adv}_s^2(g)}{2C}$$

which is stronger than the required conclusion. ◻

A direct consequence of the theorem is the following corollary, which is a special case of the KRW conjecture for inner functions g for which the soft-adversary bound is almost tight.

Corollary 8.2. *Let* $f\colon \{0,1\}^m \to \{0,1\}$ *and* $g\colon \{0,1\}^n \to \{0,1\}$ *be non-constant Boolean functions such that* $\mathrm{Adv}_s^2(g) \geq L(g)^{1-O\left(\frac{1}{\sqrt{\log L(g)}}\right)}$. *Then,*

$$L(f \diamond g) \geq \frac{1}{O(m^4)} \cdot L\left(f\right)^{1-O\left(\frac{1}{\sqrt{\log L(f)}}\right)} \cdot L(g)^{1-O\left(\frac{1}{\sqrt{\log L(g)}}\right)}.$$

Proof. By substituting the assumption on g in the bound of Theorem 8.1, we obtain that

$$L\left(f \diamond g\right)^{1+O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} \geq \frac{1}{O(m^4)} \cdot L(f) \cdot (L(g))^{1-O\left(\frac{1}{\sqrt{\log L(g)}}\right)}.$$

Moreover, since $L(f \diamond g) \leq L(f) \cdot L(g)$, we have

$$\begin{aligned}
L\left(f \diamond g\right)^{1+O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} &\leq L(f \diamond g) \cdot (L(f) \cdot L(g))^{O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} \\
&\leq L(f \diamond g) \cdot L\left(f\right)^{O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} \cdot L\left(g\right)^{O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} \\
&\leq L(f \diamond g) \cdot L\left(f\right)^{O\left(\frac{1}{\sqrt{\log L(f)}}\right)} \cdot L\left(g\right)^{O\left(\frac{1}{\sqrt{\log L(g)}}\right)}.
\end{aligned}$$

The corollary follows by combining the two bounds. ◻

Using the fact that $\mathrm{Khr}(g) \leq \mathrm{Adv}_s^2(g)$ (Proposition 7.11), we obtain the following immediate corollary.

Corollary 8.3. *Let* $f\colon \{0,1\}^m \to \{0,1\}$ *and* $g\colon \{0,1\}^n \to \{0,1\}$ *be non-constant Boolean functions. Then,*

$$L\left(f \diamond g\right)^{1+O\left(\frac{1}{\sqrt{\log L(f\diamond g)}}\right)} \geq \frac{1}{O(m^4)} \cdot L(f) \cdot \mathrm{Khr}(g).$$

Moreover, if $\mathrm{Khr}(g) \geq L(g)^{1-O\left(\frac{1}{\sqrt{\log L(g)}}\right)}$, *then*

$$L(f \diamond g) \geq \frac{1}{O(m^4)} \cdot L\left(f\right)^{1-O\left(\frac{1}{\sqrt{\log L(f)}}\right)} \cdot L(g)^{1-O\left(\frac{1}{\sqrt{\log L(g)}}\right)}.$$

8.2 Formula lower bounds for AC^0

In this section, we derive our second main application: cubic formula lower bounds for AC^0. Formally, we have the following result.

Theorem 1.1. *There exists a family of Boolean functions* $h_n\colon \{0,1\}^n \to \{0,1\}$ *for* $n \in \mathbb{N}$ *such that*

1. h_n *can be computed by uniform depth-4 unbounded fan-in formulas of size* $O(n^3)$.
2. *The formula size of* h_n *is at least* $n^{3-o(1)}$.

The function h_n is constructed similarly to the Andreev function [4], with the parity function replaced with the surjectivity function [6], defined next.

Definition 8.4. Let Σ be a finite alphabet, and let $r \in \mathbb{N}$ be such that $r \geq |\Sigma|$. The *surjectivity function* $\mathsf{Surj}_{\Sigma,r} \colon \Sigma^r \to \{0,1\}$ interprets its input as a function from $[r]$ to Σ, and outputs whether the function is surjective. In other words, $\mathsf{Surj}(\sigma_1, \ldots, \sigma_r) = 1$ if and only if every symbol in Σ appears in $(\sigma_1, \ldots, \sigma_r)$.

In order to prove Theorem 1.1, we will use Theorem 8.1 with the inner function g being Surj_n. To this end, we use the following result of [6].

Lemma 8.5 ([6]). *For every natural number s such that $s \geq 2$, we have* $\mathrm{Adv}_u(\mathsf{Surj}_{[s],2s-2}) = \Omega(s)$.

Note that we can view the inputs of $\mathsf{Surj}_{\Sigma,r}$ as binary strings of length $n = r \cdot \lceil \log |\Sigma| \rceil$ by fixing some arbitrary binary encoding of the symbols in Σ. In what follows, we extend the definition of Surj to every sufficiently large input length n as follows: Given $n \in \mathbb{N}$, we choose s to be the largest number such that $(2s-2) \cdot \lceil \log s \rceil \leq n$, and define Surj_n to be the function that interprets its input as a string in $[s]^{2s-2}$ and computes $\mathsf{Surj}_{[s],2s-2}$ on it. Lemma 8.5 now implies that for every sufficiently large $n \in \mathbb{N}$ we have $\mathrm{Adv}_u(\mathsf{Surj}_n) = \Omega\left(\frac{n}{\log n}\right)$. In particular, this implies the same bound for the soft-adversary bound, i. e., $\mathrm{Adv}_s(\mathsf{Surj}_n) = \Omega\left(\frac{n}{\log n}\right)$. For completeness, we provide an alternative proof for this lower bound on $\mathrm{Adv}_s(\mathsf{Surj}_n)$ in Section B. We turn to prove Theorem 1.1 using our special case of the KRW conjecture (Corollary 8.2).

Proof of Theorem 1.1. We would like to construct, for every sufficiently large $n \in \mathbb{N}$, a function $h_n \colon \{0,1\}^n \to \{0,1\}$ that is computable by a formula with unbounded fan-in of depth 4 and size $O(\frac{n^3}{\log^3 n})$ such that $L(h_n) = \Omega(n^{3-o(1)})$. We start by constructing the function h_n for input lengths n of a special form, and then extend the construction to all sufficiently large input lengths.

First, assume that the input length n is of the form $2^k \cdot (k+1)$, where $k \in \mathbb{N}$ is sufficiently large such that Surj_{2^k} is well-defined. The function $h_n \colon \{0,1\}^n \to \{0,1\}$ takes two inputs: the truth table of a function $f \colon \{0,1\}^k \to \{0,1\}$, and k strings $x_1, \ldots, x_k \in \{0,1\}^{2^k}$. On such inputs, the function F outputs

$$h_n(f, x_1, \ldots, x_k) = (f \diamond \mathsf{Surj}_{2^k})(x_1, \ldots, x_k) = f\left(\mathsf{Surj}_{2^k}(x_1), \ldots, \mathsf{Surj}_{2^k}(x_k)\right).$$

It is easy to see that the function h_n has input length n. We show that F can be computed by a formula with unbounded fan-in of depth 4 and size $O(\frac{n^3}{\log^3 n})$. We start by constructing a formula for Surj_{2^k}. Recall that Surj_{2^k} takes as input (the binary encoding of) a string $(\sigma_1, \ldots, \sigma_r) \in [s]^r$, where $r, s = O(\frac{2^k}{k})$. For simplicity, let us assume that every number in $[s]$ is encoded by exactly one binary string, and that if the binary input to Surj_{2^k} contains a binary string in $\{0,1\}^{\lceil \log s \rceil}$

that does not encode any number in $[s]$, then we do not care what the formula outputs. Now, observe that

$$\mathsf{Surj}_{2^k}(\sigma_1,\dots,\sigma_r) = \bigwedge_{\gamma\in[s]} \bigvee_{j=1}^{r} (\sigma_j = \gamma). \tag{8.1}$$

It is not hard to see that the expression on the right-hand side can be implemented by a formula with unbounded fan-in of depth 3 and size $s \cdot r \cdot \lceil \log s \rceil = O\left(\frac{2^{2k}}{k}\right)$. Next, observe that

$$h_n(f, x_1, \dots, x_k) = \bigvee_{y\in\{0,1\}^k} \left[(f(y) = 1) \wedge \left(\bigwedge_{i=1}^{k} \mathsf{Surj}_{2^k}(x_i) = y_i \right) \right].$$

Using the foregoing formula for surjectivity, it is not hard to see that the expression on the right-hand side can be implemented by a formula with unbounded fan-in of depth 4 and size

$$O\left(2^k \cdot k \cdot \frac{2^{2k}}{k}\right) = O\left(2^{3k}\right) = O\left(\frac{n^3}{\log^3 n}\right),$$

as required.

Finally, we prove that $L(h_n) = \Omega(n^{3-o(1)})$. To this end, let us hardwire the input f to be some function from $\{0,1\}^k$ to $\{0,1\}$ such that $L(f) = \Omega(\frac{2^k}{\log k})$ (such a function exists by a well-known counting argument, see [30, Theorem 1.23]). After hardwiring the input f, the function h_n becomes exactly the function $f \diamond \mathsf{Surj}_{2^k}$. Now, by observing that $\mathrm{Adv}_s^2(\mathsf{Surj}_{2^k}) = \tilde{\Omega}(L(\mathsf{Surj}_{2^k}))$ and applying Corollary 8.2, it follows that

$$\begin{aligned}
L(F) &\geq \tfrac{1}{O(k^4)} \cdot L(f)^{1-O\left(\frac{1}{\sqrt{\log L(f)}}\right)} \cdot L(\mathsf{Surj}_{2^k})^{1-O\left(\frac{1}{\sqrt{\log L(\mathsf{Surj}_{2^k})}}\right)} \\
&\geq \tfrac{1}{O(k^4)} \cdot \left(\frac{2^k}{\log k}\right)^{1-O\left(\frac{1}{\sqrt{k}}\right)} \cdot \left(\frac{2^{2k}}{k^2}\right)^{1-O\left(\frac{1}{\sqrt{k}}\right)} \\
&\geq 2^{3k-o(1)} \\
&= n^{3-o(1)},
\end{aligned}$$

as required.

It remains to deal with input lengths n that are not of the form $2^k \cdot (k+1)$. For such input lengths n, we choose k to be the largest natural number such that $2^k \cdot (k+1) \leq n$, and proceed as before. It can be verified that for this choice of k we have $2^k \cdot (k+1) = \Theta(n)$, and therefore all the foregoing asymptotic bounds continue to hold. □

Remark 8.6. We note that our methods cannot prove formula lower bounds that are better than cubic. As explained in Remark 7.8, it holds that $\mathrm{Adv}_s^2(f) \leq n^2$. Thus, one cannot expect to obtain a lower bound that is better than cubic by combining these measures with Andreev's argument.

8.3 Reproving known formula lower bounds

The proof of Theorem 1.1 combines a lower bound on $\mathrm{Adv}_s(\mathsf{Surj}_n)$ with an upper bound on $L(\mathsf{Surj}_n)$. More generally, we can prove the following result along similar lines.

Theorem 8.7. *Let $g\colon \{0,1\}^n \to \{0,1\}$ be an arbitrary function, and let k be an integer satisfying $k = \log n + O(1)$. Let $F\colon \{0,1\}^{2^k+nk} \to \{0,1\}$ be a function on $N = \Theta(n \log n)$ variables, whose input consists of a function $f\colon \{0,1\}^k \to \{0,1\}$ and k strings $x_1, \ldots, x_k \in \{0,1\}^n$, given by*

$$F(f, x_1, \ldots, x_k) = (f \diamond g)(x_1, \ldots, x_k).$$

*We call F the g-*based Andreev function.

1. *If either $\mathrm{Adv}_s^2(g)$ or $\mathrm{Khr}(g)$ are at least $\Omega(n^{2-o(1)})$ then $L(F) = \Omega(N^{3-o(1)})$.*
2. *If $L(g) = O(n^{2+o(1)})$ then $L(F) = O(N^{3+o(1)})$.*

The proof of Theorem 8.7 is very similar to the proof of Theorem 1.1, and so we leave it to the reader (the only difference is that we may use Corollary 8.3 instead of Corollary 8.2). Using Theorem 8.7, we can derive two known special cases: the original Andreev function (in which g is Parity), and the variant considered in [17]. In particular, using the known facts that $\mathrm{Khr}(\mathsf{Parity}) \geq n^2$ and $\mathrm{Khr}(\mathsf{Majority}) \geq \Omega(n^2)$ [33], we obtain the following results.

Corollary 8.8. *The formula complexity of the Parity_n-based Andreev function is $\Theta(N^{3\pm o(1)})$.*

Corollary 8.9. *For $m \geq 3$ odd, let $\mathsf{Majority}_n\colon \{0,1\}^m \to \{0,1\}$ be the Majority function. The formula complexity of the $\mathsf{Majority}_n$-based Andreev function is $\Omega(N^{3-o(1)})$.*

The best known upper bound on the formula complexity of $\mathsf{Majority}_n$ is only $O(n^{3.91})$ [52], and so we do not expect Corollary 8.9 to be tight.

A Proof of Proposition 7.5

In this appendix, we prove Proposition 7.5, restated next.

Proposition 7.5. *For any $f\colon \{0,1\}^n \to \{0,1\}$ we have $L(f) \geq \mathrm{Adv}_s^2(f)$.*

Preliminaries

While the proof of Proposition 7.5 is fairly simple, it uses some basic concepts from information theory which we review next.

Definition A.1. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be discrete random variables.

- The *entropy* of $\boldsymbol{x}$ is $H(\boldsymbol{x}) = \mathbb{E}_{x \leftarrow \boldsymbol{x}}\left[\log \frac{1}{\Pr[\boldsymbol{x}=x]}\right]$.
- The *conditional entropy* of $\boldsymbol{x}$ given $\boldsymbol{y}$ is $H(\boldsymbol{x} \mid \boldsymbol{y}) = \mathbb{E}_{y \leftarrow \boldsymbol{y}}\left[H(\boldsymbol{x} \mid \boldsymbol{y} = y)\right]$.

- The *mutual information* between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $I(\boldsymbol{x};\boldsymbol{y}) = H(\boldsymbol{x}) - H(\boldsymbol{x} \mid \boldsymbol{y})$.

- The *conditional mutual information* between $\boldsymbol{x}$ and $\boldsymbol{y}$ given $\boldsymbol{z}$ is

$$I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{z}) = H(\boldsymbol{x} \mid \boldsymbol{z}) - H(\boldsymbol{x} \mid \boldsymbol{y},\boldsymbol{z}).$$

We use the following basic facts from information theory (see [11] for proofs).

Fact A.2. *Let $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ be discrete random variables with finite supports, and let X denote the support of $\boldsymbol{x}$.*

- *It holds that $0 \le H(\boldsymbol{x}) \le \log|X|$.*

- *It holds that $0 \le H(\boldsymbol{x} \mid \boldsymbol{y}) \le H(\boldsymbol{x})$, where $H(\boldsymbol{x} \mid \boldsymbol{y}) = 0$ if and only if $\boldsymbol{y}$ determines $\boldsymbol{x}$ (i. e., $\boldsymbol{x}$ is a function of $\boldsymbol{y}$).*

- *If $\boldsymbol{y}$ determines $\boldsymbol{x}$ conditioned on $\boldsymbol{z}$ (i. e., $\boldsymbol{x}$ is a function of $\boldsymbol{y}$ and $\boldsymbol{z}$) then $H(\boldsymbol{y} \mid \boldsymbol{z}) \ge H(\boldsymbol{x} \mid \boldsymbol{z})$.*

- *It holds that $0 \le I(\boldsymbol{x};\boldsymbol{y}) \le H(\boldsymbol{x})$. Similarly, we have $0 \le I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{z}) \le H(\boldsymbol{x} \mid \boldsymbol{z})$, where $I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{z}) = 0$ if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent conditioned on any value of $\boldsymbol{z}$.*

- *It holds that $I(\boldsymbol{x};\boldsymbol{y}) = I(\boldsymbol{y};\boldsymbol{x})$. Similarly, $I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{z}) = I(\boldsymbol{y};\boldsymbol{x} \mid \boldsymbol{z})$.*

- The chain rule: *It holds that*

$$I(\boldsymbol{x};\boldsymbol{y},\boldsymbol{z} \mid \boldsymbol{w}) = I(\boldsymbol{x};\boldsymbol{z} \mid \boldsymbol{w}) + I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{z},\boldsymbol{w}).$$

Finally, we use the following result from the theory of interactive information complexity.

Claim A.3 ([5, Fact 4.15])**.** *Let Π be a deterministic protocol. Suppose we invoke Π on random inputs $\boldsymbol{x}$ and $\boldsymbol{y}$ for Alice and Bob, respectively, and let $\boldsymbol{\ell}$ denote the random leaf that Π reaches on those inputs. Then,*

$$I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\ell}) \ge I(\boldsymbol{x};\boldsymbol{\ell} \mid \boldsymbol{y}) + I(\boldsymbol{y};\boldsymbol{\ell} \mid \boldsymbol{x}).$$

Proof sketch. Let $\boldsymbol{t}$ denote the transcript of the protocol that is associated with $\boldsymbol{\ell}$. We prove that $I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{t}) \ge I(\boldsymbol{x};\boldsymbol{t} \mid \boldsymbol{y}) + I(\boldsymbol{y};\boldsymbol{t} \mid \boldsymbol{x})$, as this is equivalent to the claim. Suppose Alice speaks first, and denote the (random) bit she sends by $\boldsymbol{t}_1$. We show that $I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{t}_1) \ge I(\boldsymbol{x};\boldsymbol{t}_1 \mid \boldsymbol{y}) + I(\boldsymbol{y};\boldsymbol{t}_1 \mid \boldsymbol{x})$. Using the chain rule, we can write

$$I(\boldsymbol{x},\boldsymbol{y};\boldsymbol{t}_1) = I(\boldsymbol{y};\boldsymbol{t}_1) + I(\boldsymbol{x};\boldsymbol{t}_1 \mid \boldsymbol{y}) \ge I(\boldsymbol{x};\boldsymbol{t}_1 \mid \boldsymbol{y}) = I(\boldsymbol{x};\boldsymbol{t}_1 \mid \boldsymbol{y}) + I(\boldsymbol{y};\boldsymbol{t}_1 \mid \boldsymbol{x}),$$

where the last equality follows since $I(\boldsymbol{y};\boldsymbol{t}_1 \mid \boldsymbol{x}) = 0$, as Alice's message $\boldsymbol{t}_1$ is independent of $\boldsymbol{y}$ given her input $\boldsymbol{x}$. Proceeding by induction on the coordinates of $\boldsymbol{t}$ using the chain rule finishes the proof. ∎

The proof of Proposition 7.5

Our proof of Proposition 7.5 generalizes similar arguments in [32, 19]. Let Π be a protocol that solves KW_f. For every leaf ℓ of Π, we denote by i_ℓ the output of the protocol at leaf ℓ, so we have $a_{i_\ell} \neq b_{i_\ell}$ any pair of inputs (a, b) that reach the leaf ℓ. We prove that $L(\Pi) \geq \mathrm{Adv}_s^2(f)$. Let $(\boldsymbol{a}, \boldsymbol{b})$ be a distribution for f that attains $\mathrm{Adv}_s(f)$, and let

$$q_0 = \max_{\substack{b \in \mathrm{supp}(\boldsymbol{b}) \\ i \in [n]}} \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{b} = b], \qquad q_1 = \max_{\substack{a \in \mathrm{supp}(\boldsymbol{a}) \\ i \in [n]}} \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a].$$

Let $\boldsymbol{\ell}$ be the leaf that Π reaches on input $(\boldsymbol{a}, \boldsymbol{b})$. By Fact A.2, we have

$$I(\boldsymbol{\ell}; \boldsymbol{a}, \boldsymbol{b}) \leq H(\boldsymbol{\ell}) \leq \log |L(\Pi)| .$$

On the other hand, Claim A.3 implies that

$$I(\boldsymbol{\ell}; \boldsymbol{a}, \boldsymbol{b}) \geq I(\boldsymbol{\ell}; \boldsymbol{a} \mid \boldsymbol{b}) + I(\boldsymbol{\ell}; \boldsymbol{b} \mid \boldsymbol{a}).$$

Next, observe that $\boldsymbol{a}$ and $\boldsymbol{b}$ together determine $\boldsymbol{\ell}$, and that the event $\boldsymbol{\ell} = \ell$ implies that $\boldsymbol{a}_{i_\ell} \neq \boldsymbol{b}_{i_\ell}$. It follows that

$$\begin{aligned}
I(\boldsymbol{\ell}; \boldsymbol{a} \mid \boldsymbol{b}) &= H(\boldsymbol{\ell} \mid \boldsymbol{b}) - H(\boldsymbol{\ell} \mid \boldsymbol{a}, \boldsymbol{b}) && \text{(by definition)} \\
&= H(\boldsymbol{\ell} \mid \boldsymbol{b}) && (\boldsymbol{a} \text{ and } \boldsymbol{b} \text{ determine } \boldsymbol{\ell}) \\
&= \mathbb{E}_{\ell \leftarrow \boldsymbol{\ell}, b \leftarrow \boldsymbol{b}} \left[\log \frac{1}{\Pr[\boldsymbol{\ell} = \ell \mid \boldsymbol{b} = b]} \middle| \boldsymbol{b} = b \right] && \text{(Definition A.1)} \\
&\geq \mathbb{E}_{\ell \leftarrow \boldsymbol{\ell}, b \leftarrow \boldsymbol{b}} \left[\log \frac{1}{\Pr[\boldsymbol{a}_{i_\ell} \neq \boldsymbol{b}_{i_\ell} \mid \boldsymbol{b} = b]} \middle| \boldsymbol{b} = b \right] \\
&\geq \mathbb{E}_{\ell \leftarrow \boldsymbol{\ell}, b \leftarrow \boldsymbol{b}} \left[\log \frac{1}{q_0} \middle| \boldsymbol{b} = b \right] && \text{(Definition of } q_0\text{)} \\
&= \log \frac{1}{q_0}
\end{aligned}$$

Similarly, it can be shown that $I(\boldsymbol{\ell}; \boldsymbol{b} \mid \boldsymbol{a}) \geq \log \frac{1}{q_1}$. By combining the foregoing equations, it follows that

$$\log |L(\Pi)| \geq I(\boldsymbol{\ell}; \boldsymbol{a} \mid \boldsymbol{b}) + I(\boldsymbol{\ell}; \boldsymbol{b} \mid \boldsymbol{a}) \geq \log \frac{1}{q_0} + \log \frac{1}{q_1},$$

and thus

$$|L(\Pi)| \geq \frac{1}{q_0} \cdot \frac{1}{q_1} = \mathrm{Adv}_s^2(f),$$

as required.

Remark A.4. We note that Proposition 7.5 can also be proved by observing that the soft-adversary bound is a special case of the weighted adversary bound [3], which was shown to be a lower bound on formula complexity by [38]. Alternatively, one could prove Proposition 7.5 (up to a constant factor) by combining Lemma 7.6 along with Lemma 5.1 and Lemma 4.1.

B The min-entropy Khrapchenko bound

In this appendix, we describe a generalization of the Khrapchenko bound that provides a simple lower bound technique for the soft-adversary bound. We then show how this generalization can be used to give an alternative proof for the lower bound on the soft-adversary bound of the surjectivity function. We start by recalling the original Khrapchenko bound.

Definition 7.9. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant function. For every two sets $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$, let $E(A,B)$ be the set of pairs $(a,b) \in A \times B$ such that a and b differ on exactly one coordinate. The *Khrapchenko bound of f* is

$$\mathrm{Khr}(f) = \max_{A\subseteq f^{-1}(1), B\subseteq f^{-1}(0)} \frac{|E(A,B)|^2}{|A|\cdot|B|}.$$

The Khrapchenko measure can be viewed as follows: Consider the subgraph of the Hamming cube that consists only of the cut between A and B. Then, the measure $\frac{|E(A,B)|^2}{|A|\cdot|B|}$ is the product of the average degree of a vertex in A (which is $\frac{|E(A,B)|}{|A|}$) and the average degree of a vertex in B (which is $\frac{|E(A,B)|}{|B|}$). Note that the average degree of A can also be described as the average, over all strings $a \in A$, of the number of coordinates $i \in [n]$ such that if we flip the i-th bit of a we get a string in B. We generalize the Khrapchenko measure as follows:

- Whereas the Khrapchenko bound maximizes over all *cuts* (A,B) of the Hamming cube with $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$, we are maximizing over all *distributions* over edges $(\boldsymbol{a},\boldsymbol{b})$ of the Hamming cube, where $\boldsymbol{a} \in f^{-1}(1)$ and $\boldsymbol{b} \in f^{-1}(0)$.

- Whereas the Khrapchenko bound considers the average number of coordinates i as described above, we consider the min-entropy of the coordinate $\boldsymbol{i}$ on which $\boldsymbol{a},\boldsymbol{b}$ differ.

- Whereas the Khrapchenko bound considers functions whose inputs are binary strings, we consider inputs that are strings over arbitrary finite alphabets.

Our generalization uses the following notion: The *conditional min-entropy* of a random variable $\boldsymbol{x}$ conditioned on a random variable $\boldsymbol{y}$ is $H_\infty(\boldsymbol{x} \mid \boldsymbol{y}) = \min_{x,y} \log \frac{1}{\Pr[\boldsymbol{x}=x|\boldsymbol{y}=y]}$. We can now define our generalization formally:

Definition B.1. Let Σ be a finite alphabet, and let $f\colon \Sigma^n \to \{0,1\}$ be a non-constant Boolean function. We say that a distribution $(\boldsymbol{a},\boldsymbol{b})$ on $f^{-1}(1) \times f^{-1}(0)$ is a *Khrapchenko distribution for f* if $\boldsymbol{a}$ and $\boldsymbol{b}$ always differ on a unique coordinate $\boldsymbol{i} \in [n]$. We define the *min-entropy Khrapchenko bound of f*, which we denote $\mathrm{Khr}_{H_\infty}(f)$, to be the maximum of the quantity

$$2^{H_\infty(\boldsymbol{i}|\boldsymbol{a})+H_\infty(\boldsymbol{i}|\boldsymbol{b})}$$

over all Khrapchenko distributions $(\boldsymbol{a},\boldsymbol{b})$ for f.

As noted above, the min-entropy Khrapchenko bound can be used to lower bound the the soft-adversary bound. In order to define the adversary bound of a function $f\colon \Sigma^n \to \{0,1\}$ over a non-boolean alphabet Σ, we view f as if its input is a binary string in $\{0,1\}^{n\cdot\lceil\log|\Sigma|\rceil}$ that encodes a string in Σ^n via some fixed encoding (the choice of the encoding does not matter for what follows).

Lemma B.2. *For every non-constant function $f\colon \Sigma^n \to \{0,1\}$ we have $\mathrm{Adv}_s(f) \geq \sqrt{\mathrm{Khr}_{H_\infty}(f)}$.*

Proof. Let $(\boldsymbol{a},\boldsymbol{b})$ be a Khrapchenko distribution that attains $\mathrm{Khr}_{H_\infty}(f)$. Let $\boldsymbol{i} \in [n]$ be the unique index at which $\boldsymbol{a},\boldsymbol{b}$ differ, and let $\boldsymbol{a}_{i,j},\boldsymbol{b}_{i,j}$ be the j-th bits of the binary encoding of $\boldsymbol{a}_i,\boldsymbol{b}_i \in \Sigma$. For any $a \in \mathrm{supp}(\boldsymbol{a})$ and $(i,j) \in [n]\times[\lceil\log|\Sigma|\rceil]$, we have

$$\Pr[\boldsymbol{a}_{i,j} \neq \boldsymbol{b}_{i,j} \mid \boldsymbol{a} = a] \leq \Pr[\boldsymbol{a}_i \neq \boldsymbol{b}_i \mid \boldsymbol{a} = a] = \Pr[\boldsymbol{i} = i \mid \boldsymbol{a} = a] \leq 2^{-H_\infty(\boldsymbol{i}|\boldsymbol{a})}.$$

It follows that the first factor in the definition of Adv_s is at least $2^{H_\infty(\boldsymbol{i}|\boldsymbol{a})}$. Similarly, the second factor is at least $2^{H_\infty(\boldsymbol{i}|\boldsymbol{b})}$, and so the entire expression is at least $\sqrt{\mathrm{Khr}_{H_\infty}(f)}$. $\square$

Similarly to the original Khrapchenko bound, the min-entropy Khrapcehnko bound too gives a lower bound on the formula complexity $L(f)$, which can be proved by combining Proposition 7.5 and Lemma B.2. Again, in order to define the formula complexity of a function $f\colon \Sigma^n \to \{0,1\}$ over a non-boolean alphabet, we view f as if its input is a binary string in $\{0,1\}^{n\cdot\lceil\log|\Sigma|\rceil}$.

Corollary B.3. *Let Σ be a finite alphabet, and let $f\colon \Sigma^n \to \{0,1\}$ be a non-constant Boolean function. Then $L(f) \geq \mathrm{Khr}_{H_\infty}(f)$.*

By combining Lemmas 7.6 and B.2, it also follows that the min-entropy Khrapchenko bound too can be used for constructing hiding projections.

Corollary B.4. *Let Σ be a finite alphabet, let $f\colon \Sigma^n \to \{0,1\}$ be a non-constant Boolean function. There is a q-hiding projection $\boldsymbol{\pi}$ to a single variable y such that $q = \sqrt{1/\mathrm{Khr}_{H_\infty}(f)}$ and such that $f|_{\boldsymbol{\pi}}$ is a non-constant function with probability 1.*

B.1 Lower bound on the surjectivity function

In order to demonstrate the usefulness of the min-entropy Khrapchenko bound, we use it to prove a lower bound on the surjectivity function from Section 8.2. This implies a similar bound on the soft-adversary bound of the surjectivity function, and provides an alternative way for proving our main theorem.

Definition 8.4. Let Σ be a finite alphabet, and let $r \in \mathbb{N}$ be such that $r \geq |\Sigma|$. The *surjectivity function* $\mathsf{Surj}_{\Sigma,r}\colon \Sigma^r \to \{0,1\}$ interprets its input as a function from $[r]$ to Σ, and outputs whether the function is surjective. In other words, $\mathsf{Surj}(\sigma_1,\ldots,\sigma_r) = 1$ if and only if every symbol in Σ appears in $(\sigma_1,\ldots,\sigma_r)$.

Lemma B.5. *For every $s \in \mathbb{N}$, we have $\mathrm{Khr}_{H_\infty}(\mathsf{Surj}_{[2s+1],3s+1}) = \Omega(s^2)$.*

Proof. Let $s \in \mathbb{N}$, let $\Sigma = [2s+1]$, and let $\mathsf{Surj} = \mathsf{Surj}_{\Sigma,3s+1}$. We define a Khrapchenko distribution $(\boldsymbol{a}, \boldsymbol{b})$ for Surj as follows. The input $\boldsymbol{b} \in \mathsf{Surj}^{-1}(0)$ is a uniformly distributed string in Σ^{3s+1} in which $s+1$ of the symbols in Σ appear exactly twice, $s-1$ of the symbols in Σ appear exactly once, and the remaining symbol in Σ does not appear at all. The string $\boldsymbol{a}$ is sampled by choosing uniformly at random a coordinate $\boldsymbol{i}$ such that $\boldsymbol{b}_{\boldsymbol{i}}$ is one of the symbols that appear twice in $\boldsymbol{b}$, and replacing $\boldsymbol{b}_{\boldsymbol{i}}$ with the unique symbol in Σ that does not appear in $\boldsymbol{b}$.

We turn to bound $H_\infty(\boldsymbol{i} \mid \boldsymbol{b})$ and $H_\infty(\boldsymbol{i} \mid \boldsymbol{a})$. First, observe that conditioned on any choice of $\boldsymbol{b}$, the coordinate $\boldsymbol{i}$ is uniformly distributed among $2s+2$ coordinates, and therefore

$$H_\infty(\boldsymbol{i} \mid \boldsymbol{b}) = \log(2s+2).$$

In order to bound $H_\infty(\boldsymbol{i} \mid \boldsymbol{a})$, observe that $\boldsymbol{a}$ is distributed like a uniformly distributed string in Σ^{3s+1} in which $s+1$ of the symbols in Σ appear exactly once, and the remaining s symbols in Σ appear exactly twice. Conditioned on any choice of $\boldsymbol{a}$, the coordinate $\boldsymbol{i}$ is uniformly distributed over the $s+1$ coordinates of symbols that appear exactly once. It follows that

$$H_\infty(\boldsymbol{i} \mid \boldsymbol{a}) \geq \log(s+1).$$

We conclude that

$$\mathrm{Khr}_{H_\infty}(\mathsf{Surj}_n) \geq 2^{H_\infty(\boldsymbol{i}|\boldsymbol{a})+H_\infty(\boldsymbol{i}|\boldsymbol{b})} \geq (s+1)\cdot(2s+2) = \Omega\left(s^2\right). \qquad \square$$

Lemma B.5 implies, via Corollary B.3 a quadratic lower bound on the formula complexity of Surj. The same result also follows from the lower bound on the adversary bound of Surj due to [6] (Lemma 8.5).

B.1.1 Comparison to the Khrapchenko bound

We now further compare the min-entropy Khrapchenko bound to the original Khrapchenko bound. Observe that when $\Sigma = \{0,1\}$, the min-entropy $H_\infty(\boldsymbol{i} \mid \boldsymbol{a})$ is exactly the min-entropy of a random neighbor of $\boldsymbol{a}$. In particular, when $(\boldsymbol{a}, \boldsymbol{b})$ is the uniform distribution over a set of edges, the min-entropy $H_\infty(\boldsymbol{i} \mid \boldsymbol{a})$ is the logarithm of the minimal degree of a vertex a. Moreover, if the latter set of edges also induces a regular graph, then the measure $2^{H_\infty(\boldsymbol{i}|\boldsymbol{a})+H_\infty(\boldsymbol{i}|\boldsymbol{b})}$ coincides exactly with the original measure $\frac{|E(A,B)|^2}{|A|\cdot|B|}$ of Khrapchenko. More generally, when $\Sigma = \{0,1\}$, the bound $\mathrm{Khr}_{H_\infty}(f)$ is within a constant factor of the original Khrapchenko bound, as we show in Section C.

Proposition B.6. *For any non-constant function $f\colon \{0,1\}^n \to \{0,1\}$ it holds that*

$$\frac{\mathrm{Khr}(f)}{4} \leq \mathrm{Khr}_{H_\infty}(f) \leq \mathrm{Khr}(f).$$

Unfortunately, when Σ is a larger alphabet, the connection between $\mathrm{Khr}_{H_\infty}(f)$ and the measure $\frac{|E(A,B)|^2}{|A||B|}$ is not so clean. Specifically, the min-entropy $H_\infty(\boldsymbol{i} \mid \boldsymbol{a})$ has no clear connection to the degree of $\boldsymbol{a}$, since the vertex $\boldsymbol{a}$ may have multiple neighbors that correspond to the same coordinate $\boldsymbol{i}$.

B.1.2 Previous work on the Khrapchenko bound

Several versions of the Khrapchenko bound appeared in the literature: Zwick [58] generalized the Khrapchenko bound such that different input coordinates can be given different weights, and Koutsoupias [36] gave a spectral generalization of the bound. The paper of Håstad [23] observed that his analogue of Lemma 4.1 can be viewed as a generalization of the Khrapchenko bound. Ganor, Komargodski, and Raz [18] considered a variant of the Khrapchenko bound in which the edges of the Boolean hypercube are replaced with random walks on the noisy hypercube. Of particular relevance is a paper of Laplante, Lee, and Szegedy [38] that defined a complexity measure that is very similar to our min-entropy Khrapchenko bound, except that the entropy is replaced with Kolmogorov complexity.

B.1.3 An entropy Khrapchenko bound

It is possible to generalize the complexity measure Khr_{H_∞} by replacing the min-entropy in Definition B.1 with Shannon entropy. Such a measure would still give a lower bound on formula complexity — specifically, the proof of Corollary B.3 would go through without a change. However, we do not know how to use such a measure for constructing hiding projections as in Corollary B.4. We note that it is easy to prove that such a measure is an upper bound on Khr_{H_∞}.

C Proof of Propositions B.6 and 7.11

In this appendix we prove Propositions B.6 and 7.11, restated next.

Proposition B.6. *For any $f\colon \{0,1\}^n \to \{0,1\}$ we have $\frac{\mathrm{Khr}(f)}{4} \le \mathrm{Khr}_{H_\infty}(f) \le \mathrm{Khr}(f)$.*

Proposition 7.11. *For any $f\colon \{0,1\}^n \to \{0,1\}$ we have $\sqrt{\frac{\mathrm{Khr}(f)}{4}} \le \mathrm{Adv}_u(f)$.*

We relate $\mathrm{Khr}(f)$ to $\mathrm{Khr}_{H_\infty}(f)$ using an auxiliary measure $\mathrm{Khr}_{\min}(f)$:

$$\mathrm{Khr}_{\min}(f) = \max_{A\subseteq f^{-1}(1), B\subseteq f^{-1}(0)} \left(\min_{a\in A} |E(a,B)| \cdot \min_{b\in B} |E(A,b)|\right),$$

where $E(a,B) = E(\{a\},B)$ and similarly $E(A,b) = E(A,\{b\})$.

We first show that $\mathrm{Khr}_{\min}(f)$ and $\mathrm{Khr}(f)$ are equal up to constants.

Claim C.1. *For any $f\colon \Sigma^n \to \{0,1\}$ we have $\frac{\mathrm{Khr}(f)}{4} \le \mathrm{Khr}_{\min}(f) \le \mathrm{Khr}(f)$.*

Proof. The inequality $\mathrm{Khr}_{\min}(f) \le \mathrm{Khr}(f)$ is simple since $\min_{a\in A} |E(a,B)| \le |E(A,B)|/|A|$ and similarly $\min_{b\in B} |E(A,b)| \le |E(A,B)|/|B|$.

The other direction is more subtle. For ease of notation, for any sets $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$ we write $\mathrm{Khr}(A,B) = \frac{|E(A,B)|^2}{|A||B|}$. Thus,

$$\mathrm{Khr}(f) = \max_{A\subseteq f^{-1}(1), B\subseteq f^{-1}(0)} \mathrm{Khr}(A,B). \tag{C.1}$$

Assume that A and B are sets that maximize $\mathrm{Khr}(A,B)$ in Equation C.1. We show that

$$\forall a \in A: \quad |E(a,B)| \geq \tfrac{|E(A,B)|}{2|A|}, \tag{C.2}$$

$$\forall b \in B: \quad |E(A,b)| \geq \tfrac{|E(A,B)|}{2|B|}. \tag{C.3}$$

In words, the min-degree is at least half the average degree. Before showing why Equations C.2 and C.3 hold, we show that they imply the statement of the claim. Indeed,

$$\mathrm{Khr}_{\min}(f) \geq \left(\min_{a\in A} |E(a,B)| \cdot \min_{b\in B} |E(A,b)|\right) \geq \tfrac{|E(A,B)|}{2|A|} \cdot \tfrac{|E(A,B)|}{2|B|} = \tfrac{\mathrm{Khr}(A,B)}{4} = \tfrac{\mathrm{Khr}(f)}{4}.$$

It remains to show that Equations C.2 and C.3 hold. We focus on Equation C.2 due to symmetry. Assume by contradiction that there exists an $a \in A$ for which

$$|E(a,B)| < \tfrac{|E(A,B)|}{2|A|}.$$

It must be the case that $|A| > 1$, as otherwise A contains only one element and $|E(a,B)| = |E(A,B)|/|A|$ for this element. Consider now the set $A' = A \setminus \{a\}$ — which is non-empty by the above discussion. We claim that $\mathrm{Khr}(A',B) > \mathrm{Khr}(A,B)$, contradicting the choice of A, B. Indeed,

$$\begin{aligned}
\mathrm{Khr}(A',B) &= \frac{|E(A',B)|^2}{|A'||B|} \\
&= \frac{(|E(A,B)| - |E(a,B)|)^2}{(|A|-1)|B|} \\
&> \frac{(|E(A,B)| - \frac{|E(A,B)|}{2|A|})^2}{(|A|-1)|B|} && \text{(By assumption on } a\text{)} \\
&= \frac{|E(A,B)|^2 \cdot (1-\frac{1}{2|A|})^2}{|A||B| \cdot (1-\frac{1}{|A|})} \\
&= \mathrm{Khr}(A,B) \cdot \frac{(1-\frac{1}{2|A|})^2}{(1-\frac{1}{|A|})} \\
&> \mathrm{Khr}(A,B).
\end{aligned}$$

□

It turns out that over the binary alphabet, $\mathrm{Khr}_{\min}(f)$ and $\mathrm{Khr}_{H_\infty}(f)$ coincide.

Claim C.2. *For any $f\colon \{0,1\}^n \to \{0,1\}$ we have $\mathrm{Khr}_{H_\infty}(f) = \mathrm{Khr}_{\min}(f)$.*

Proof. We first show that $\mathrm{Khr}_{H_\infty}(f) \geq \mathrm{Khr}_{\min}(f)$. Let A, B be sets that maximize the expression $(\min_{a\in A} |E(a,B)| \cdot \min_{b\in B} |E(A,b)|)$. We take $(\boldsymbol{a},\boldsymbol{b})$ to be a uniformly distributed pair in $E(A,B)$. It is not hard to see that

$$2^{H_\infty(\boldsymbol{i}|\boldsymbol{a})} = 2^{H_\infty(\boldsymbol{b}|\boldsymbol{a})} = \min_{a\in A} |E(a,B)|,$$

and similarly $2^{H_\infty(\boldsymbol{i}|\boldsymbol{b})} = \min_{b\in B} |E(A,b)|$. We thus get $\mathrm{Khr}_{H_\infty}(f) \geq \mathrm{Khr}_{\min}(f)$.

Next, we show the other direction, $\mathrm{Khr}_{\min}(f) \geq \mathrm{Khr}_{H_\infty}(f)$. Let $(\boldsymbol{a},\boldsymbol{b})$ be a random variable distributed according to a Khrapchenko distribution for f that attains the maximum of $2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})+H_\infty(\boldsymbol{b}|\boldsymbol{a})}$ over all such distributions. Let $A := \mathrm{supp}(\boldsymbol{a})$ and $B := \mathrm{supp}(\boldsymbol{b})$ be the supports of $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. By definition of H_∞ we have $2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})} = \frac{1}{\max_{a,b}\Pr[\boldsymbol{a}=a|\boldsymbol{b}=b]}$. Rearranging, we get that for any b in the support of $\boldsymbol{b}$ it holds that

$$\Pr[\boldsymbol{a} = a \mid \boldsymbol{b} = b] \leq 1/2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})}.$$

In particular, since theses probabilities sum to 1, it must be the case that there are at least $2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})}$ neighbors of b in A — i. e., $|E(A,b)| \geq 2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})}$ for all $b \in B$. Similarly, $|E(a,B)| \geq 2^{H_\infty(\boldsymbol{b}|\boldsymbol{a})}$ for all $a \in A$. The two sets A and B show that $\mathrm{Khr}_{\min}(f) \geq 2^{H_\infty(\boldsymbol{a}|\boldsymbol{b})} \cdot 2^{H_\infty(\boldsymbol{b}|\boldsymbol{a})} = \mathrm{Khr}_{H_\infty}(f)$. ◻

Proposition B.6 follows by combining these two claims. Proposition 7.11 follows from Claim C.2 using the following simple observation.

Claim C.3. *For any* $f\colon \{0,1\}^n \to \{0,1\}$ *we have* $\mathrm{Adv}_u(f) \geq \sqrt{\mathrm{Khr}_{\min}(f)}$.

Proof. Let A, B be sets that attain $\mathrm{Khr}_{\min}(f)$. Define R to be the set of pairs $(a,b) \in A \times B$ that differ at a single coordinate. Thus $|R(a,B)| = |E(a,B)|$ and $|R(A,b)| = |E(A,b)|$. Moreover, $|R_i(a,B)| \leq 1$ since there is a unique b that differs from a only on the i-th coordinate. Similarly, $|R_i(A,b)| \leq 1$. The claim now immediately follows by comparing the definitions of $\mathrm{Khr}_{\min}(f)$ and $\mathrm{Adv}_u(f)$. ◻

Acknowledgements

A. T. would like to thank Igor Carboni Oliveira for bringing the question of proving formula size lower bounds for $\boldsymbol{AC}^0$ to his attention. We are also grateful to Robin Kothari for posing this open question on "Theoretical Computer Science Stack Exchange" [35], and to Kaveh Ghasemloo and Stasys Jukna for their feedback on this question. We would like to thank Anna Gál for very helpful discussions, and Gregory Rosenthal for helpful comments on the manuscript. Finally, we are grateful to anonymous referees for comments that improved the presentation of this work.

This work was partly carried out while the authors were visiting the Simons Institute for the Theory of Computing in association with the DIMACS/Simons Collaboration on Lower Bounds in Computational Complexity, which is conducted with support from the National Science Foundation.

References

[1] MIKLÓS AJTAI: Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. [doi:10.1016/0168-0072(83)90038-6] 4

[2] ANDRIS AMBAINIS: Quantum lower bounds by quantum arguments. *J. Comput. System Sci.*, 64(4):750–767, 2002. Special issue on STOC 2000 (Portland, OR). [doi:10.1006/jcss.2002.1826] 4, 29

[3] ANDRIS AMBAINIS: Polynomial degree vs. quantum query complexity. *J. Comput. System Sci.*, 72(2):220–238, 2006. [doi:10.1016/j.jcss.2005.06.006] 31, 39

[4] ALEXANDER E. ANDREEV: On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow University Mathematics Bulletin*, 42(1):24–29, 1987. Vestnik Moskov. Univ. Ser. 1: Mat. Mekh. 1987(1) 70–73 (Russian original on Math-net.ru) MR 0883632. 2, 12, 35

[5] BOAZ BARAK, MARK BRAVERMAN, XI CHEN, AND ANUP RAO: How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013. [doi:10.1137/100811969] 38

[6] PAUL BEAME AND WIDAD MACHMOUCHI: The quantum query complexity of AC^0. *Quantum Inf. Comput.*, 12(7–8):670–676, 2012. [doi:10.26421/QIC12.7-8-11] 4, 12, 13, 35, 42

[7] MARIA LUISA BONET AND SAMUEL R. BUSS: Size-depth tradeoffs for Boolean formulae. *Inform. Process. Lett.*, 49(3):151–155, 1994. [doi:10.1016/0020-0190(94)90093-0] 7

[8] RICHARD P. BRENT: The parallel evaluation of general arithmetic expressions. *J. ACM*, 21(2):201–206, 1974. [doi:10.1145/321812.321815] 7

[9] HARRY BUHRMAN AND RONALD DE WOLF: Complexity measures and decision tree complexity: a survey. *Theoret. Comput. Sci.*, 288(1):21–43, 2002. [doi:10.1016/S0304-3975(01)00144-X] 29

[10] ANDREW M. CHILDS, SHELBY KIMMEL, AND ROBIN KOTHARI: The quantum query complexity of read-many formulas. In *Proc. 20th Eur. Symp. Algorithms (ESA'12)*, pp. 337–348. Springer, 2012. [doi:10.1007/978-3-642-33090-2_30] 4

[11] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Wiley-Interscience, 1991. [doi:10.1002/047174882X] 38

[12] SUSANNA F. DE REZENDE, OR MEIR, JAKOB NORDSTRÖM, TONIANN PITASSI, AND ROBERT ROBERE: KRW composition theorems via lifting. In *Proc. 61st FOCS*, pp. 43–49. IEEE Comp. Soc., 2020. [doi:10.1109/FOCS46700.2020.00013, ECCC:TR20-099] 4

[13] IRIT DINUR AND OR MEIR: Toward the KRW composition conjecture: Cubic formula lower bounds via communication complexity. *Comput. Complexity*, 27(3):375–462, 2018. [doi:10.1007/s00037-017-0159-x] 4

[14] JEFF EDMONDS, RUSSELL IMPAGLIAZZO, STEVEN RUDICH, AND JIRÍ SGALL: Communication complexity towards lower bounds on circuit depth. *Comput. Complexity*, 10(3):210–246, 2001. [doi:10.1007/s00037-001-8195-x] 4

[15] YUVAL FILMUS, OR MEIR, AND AVISHAY TAL: Shrinkage under random projections, and cubic formula lower bounds for AC0 (extended abstract). In *Proc. 12th Innovations in Theoret. Comp. Sci. Conf. (ITCS'21)*, pp. 89:1–7. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2021. [doi:10.4230/LIPIcs.ITCS.2021.89]

[16] MERRICK L. FURST, JAMES B. SAXE, AND MICHAEL SIPSER: Parity, circuits, and the polynomial-time hierarchy. *Math. Sys. Theory*, 17(1):13–27, 1984. [doi:10.1007/BF01744431] 4

[17] ANNA GÁL, AVISHAY TAL, AND ADRIAN TREJO NUÑEZ: Cubic formula size lower bounds based on compositions with majority. In *Proc. 10th Innovations in Theoret. Comp. Sci. Conf. (ITCS'19)*, pp. 35:1–35:13. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.35] 5, 6, 37

[18] ANAT GANOR, ILAN KOMARGODSKI, AND RAN RAZ: The spectrum of small De Morgan formulas. *Electron. Colloq. Comput. Complexity*, TR12-174, 2012. [ECCC] 43

[19] DMITRY GAVINSKY, OR MEIR, OMRI WEINSTEIN, AND AVI WIGDERSON: Toward better formula lower bounds: The composition of a function and a universal relation. *SIAM J. Comput.*, 46(1):114–131, 2017. [doi:10.1137/15M1018319] 3, 4, 39

[20] MIKAEL GOLDMANN AND JOHAN HÅSTAD: A simple lower bound for monotone clique using a communication game. *Inform. Process. Lett.*, 41(4):221–226, 1992. [doi:10.1016/0020-0190(92)90184-W] 4

[21] MIKA GÖÖS AND TONIANN PITASSI: Communication lower bounds via critical block sensitivity. *SIAM J. Comput.*, 47(5):1778–1806, 2018. [doi:10.1137/16M1082007] 4

[22] JOHAN HÅSTAD: Almost optimal lower bounds for small depth circuits. In S. MICALI, editor, *Randomness and Computation*, volume 5 of *Adv. Computing Research*, pp. 143–170. JAI Press, 1989. DSpace@MIT. Preliminary version in STOC'86. 4

[23] JOHAN HÅSTAD: The shrinkage exponent of De Morgan formulas is 2. *SIAM J. Comput.*, 27(1):48–64, 1998. Preliminary version in FOCS'93. [doi:10.1137/S0097539794261556] 2, 3, 5, 8, 9, 15, 17, 19, 43

[24] JOHAN HÅSTAD, STASYS JUKNA, AND PAVEL PUDLÁK: Top-down lower bounds for depth-three circuits. *Comput. Complexity*, 5(2):99–112, 1995. [doi:10.1007/BF01268140] 3

[25] JOHAN HÅSTAD, BENJAMIN ROSSMAN, ROCCO A. SERVEDIO, AND LI-YANG TAN: An average-case depth hierarchy theorem for Boolean circuits. *J. ACM*, 64(5):35:1–27, 2017. Preliminary version in FOCS'15. [doi:10.1145/3095799] 6

[26] JOHAN HÅSTAD AND AVI WIGDERSON: Composition of the universal relation. In JIN YI CAI, editor, *Advances In Computational Complexity Theory*, volume 13 of *DIMACS Ser. in Discr. Math.*, pp. 119–134. Amer. Math. Soc., 1993. [doi:10.1090/dimacs/013/07] 4

[27] JOHAN TORKEL HÅSTAD: *Computational limitations of small-depth circuits*. Ph. D. thesis, MIT, 1986. Available on author's website, published as a book by MIT Press, 1987. 6

[28] PETER HØYER, TROY LEE, AND ROBERT ŠPALEK: Negative weights make adversaries stronger. In *Proc. 39th STOC*, pp. 526–535. ACM Press, 2007. [doi:10.1145/1250790.1250867] 31

[29] RUSSELL IMPAGLIAZZO AND NOAM NISAN: The effect of random restrictions on formula size. *Random Struct. Algor.*, 4(2):121–134, 1993. [doi:10.1002/rsa.3240040202] 2

[30] STASYS JUKNA: *Boolean Function Complexity – Advances and Frontiers*. Volume 27 of *Algorithms and Combinatorics*. Springer, 2012. [doi:10.1007/978-3-642-24508-4] 4, 36

[31] MAURICIO KARCHMER, RAN RAZ, AND AVI WIGDERSON: Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Comput. Complexity*, 5(3–4):191–204, 1995. [doi:10.1007/BF01206317] 3, 4, 33

[32] MAURICIO KARCHMER AND AVI WIGDERSON: Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Discr. Math.*, 3(2):255–265, 1990. [doi:10.1137/0403021] 3, 4, 7, 39

[33] VALERII MIKHAILOVICH KHRAPCHENKO: A method of obtaining lower bounds for the complexity of π-schemes. *Math. Notes. Acad. Sci. USSR (English translation)*, 10:474–479, 1971. Matem. Zametki 10(1) 83-92, 1971 (Russian original on Math-net.ru). [doi:10.1007/BF01747074] 2, 29, 32, 37

[34] SAJIN KOROTH AND OR MEIR: Improved composition theorems for functions and relations. In *Proc. 22nd Internat. Conf. on Randomization and Computation (RANDOM'18)*, pp. 48:1–18. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.48] 4

[35] ROBIN KOTHARI: Formula size lower bounds for AC0 functions, 2011. Theoretical Computer Science Stack Exchange. 45

[36] ELIAS KOUTSOUPIAS: Improvements on Khrapchenko's theorem. *Theoret. Comput. Sci.*, 116(2):399–403, 1993. [doi:10.1016/0304-3975(93)90330-V] 43

[37] EYAL KUSHILEVITZ AND NOAM NISAN: *Communication Complexity*. Cambridge Univ. Press, 1997. [doi:10.1017/9781108671644] 6

[38] SOPHIE LAPLANTE, TROY LEE, AND MARIO SZEGEDY: The quantum adversary method and classical formula size lower bounds. *Comput. Complexity*, 15(2):163–196, 2006. [doi:10.1007/s00037-006-0212-7] 30, 31, 32, 39, 43

[39] TROY LEE, RAJAT MITTAL, BEN W. REICHARDT, ROBERT ŠPALEK, AND MARIO SZEGEDY: Quantum query complexity of state conversion. In *Proc. 52nd FOCS*, pp. 344–353. IEEE Comp. Soc., 2011. [doi:10.1109/FOCS.2011.75] 31

[40] LILY LI AND MORGAN SHIRLEY: The general adversary bound: A survey, 2021. [arXiv:2104.06380] 31

[41] OR MEIR: Toward better depth lower bounds: Two results on the multiplexor relation. *Comput. Complexity*, 29(1):4, 2020. [doi:10.1007/s00037-020-00194-8] 4

[42] IVAN MIHAJLIN AND ALEXANDER SMAL: Toward better depth lower bounds: The XOR-KRW conjecture. In *Proc. 36th Comput. Complexity Conf. (CCC'21)*, pp. 38:1–24. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2021. [doi:10.4230/LIPIcs.CCC.2021.38] 4

[43] ÈDUARD IVANOVIČ NEČIPORUK: On a Boolean function. *Soviet Math. Doklady*, 7(4):999–1000, 1966. Doklady Akad. Nauk SSSR 169(4), 1966, 765–766 (Russian original on Math-net.ru). 4

[44] MIKE PATERSON AND URI ZWICK: Shrinkage of De Morgan formulae under restriction. *Random Struct. Algor.*, 4(2):135–150, 1993. [doi:10.1002/rsa.3240040203] 2

[45] TONIANN PITASSI AND ROBERT ROBERE: Strongly exponential lower bounds for monotone computation. In *Proc. 49th STOC*, pp. 1246–1255. ACM Press, 2017. [doi:10.1145/3055399.3055478] 4

[46] RAN RAZ AND PIERRE MCKENZIE: Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999. [doi:10.1007/s004930050062] 4

[47] RAN RAZ AND AVI WIGDERSON: Monotone circuits for matching require linear depth. *J. ACM*, 39(3):736–744, 1992. [doi:10.1145/146637.146684] 4

[48] ALEXANDER A. RAZBOROV: Applications of matrix methods to the theory of lower bounds in computational complexity. *Combinatorica*, 10(1):81–93, 1990. [doi:10.1007/BF02122698] 7

[49] BEN W. REICHARDT: Span programs and quantum query complexity: the general adversary bound is nearly tight for every Boolean function. In *Proc. 50th FOCS*, pp. 544–551. IEEE Comp. Soc., 2009. [doi:10.1109/FOCS.2009.55] 31

[50] BEN W. REICHARDT: Reflections for quantum query algorithms. In *Proc. 22nd Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'11)*, pp. 560–569. SIAM, 2011. [doi:10.1137/1.9781611973082.44] 31

[51] BEN W. REICHARDT: Span programs are equivalent to quantum query algorithms. *SIAM J. Comput.*, 43(3):1206–1219, 2014. [doi:10.1137/100792640] 31

[52] IGOR SERGEEVICH SERGEEV: Complexity and depth of formulas for symmetric Boolean functions. *Moscow University Mathematics Bulletin*, 71(3):127–130, 2016. Vestnik Moskov. Univ. Ser. 1: Mat. Mekh. 2016(3) 53–57 (Russian original on Math-net.ru). [doi:10.3103/S0027132216030098] 37

[53] ROBERT ŠPALEK AND MARIO SZEGEDY: All quantum adversary methods are equivalent. *Theory of Computing*, 2(1):1–18, 2006. [doi:10.4086/toc.2006.v002a001] 30, 31

[54] PHILIP M. SPIRA: On time-hardware complexity tradeoffs for Boolean functions. In *Proc. 4th Hawaii International Symposium on System Sciences*, pp. 525–527. Western Periodicals, 1971. Google books. 7

[55] BELLA ABRAMOVNA SUBBOTOVSKAYA: Realizations of linear functions by formulas using $\vee$, &, $^{-}$. *Soviet Math. Doklady*, 2:110–112, 1961. Doklady Akad. Nauk SSSR 136(3), 553–555, 1961 (Russian original on Math-net.ru). 2

[56] AVISHAY TAL: Shrinkage of De Morgan formulae by spectral techniques. In *Proc. 55th FOCS*, pp. 551–560. IEEE Comp. Soc., 2014. [doi:10.1109/FOCS.2014.65] 2, 3

[57] ANDREW CHI-CHIH YAO: Separating the polynomial-time hierarchy by oracles (preliminary version). In *Proc. 26th FOCS*, pp. 1–10. IEEE Comp. Soc., 1985. [doi:10.1109/SFCS.1985.49] 4

[58] URI ZWICK: An extension of Khrapchenko's theorem. *Inform. Process. Lett.*, 37(4):215–217, 1991. [doi:10.1016/0020-0190(91)90191-J] 43

AUTHORS

Yuval Filmus
Associate Professor
Technion — Israel Institute of Technology
The Henry and Marilyn Taub Faculty of Computer Science
Faculty of Mathematics
Haifa 3200003, Israel
yuvalfi@cs.technion.ac.il
https://yuvalfilmus.cs.technion.ac.il/

Or Meir
Associate Professor
Department of Computer Science
University of Haifa
Haifa 3303220, Israel
ormeir@cs.haifa.ac.il
https://cs.haifa.ac.il/~ormeir/

Avishay Tal
Assistant Professor
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720, United States
atal@berkeley.edu
https://www.avishaytal.org/

ABOUT THE AUTHORS

YUVAL FILMUS received his Ph. D. from the University of Toronto under the supervision of Toni Pitassi. Subsequently, he held postdoctoral positions at the Simons Institute for the Theory of Computing and the Institute for Advanced Study. He is interested in Boolean function analysis, complexity theory, and combinatorics.

OR MEIR received his Ph. D. from the Weizmann Institute of Science under the supervision of Oded Goldreich. Subsequently, he held postdoctoral positions at Stanford University, the Institute for Advanced Study, and the Weizmann Institute of Science. He is interested in complexity theory, and in particular in circuit complexity, communication complexity, coding theory, probabilistic proof systems, and the theory of pseudorandomness.

AVISHAY TAL received his M. Sc. from the Technion under the supervision of Amir Shpilka, and his Ph. D. from the Weizmann Institute of Science under the supervision of Ran Raz. Subsequently, he held postdoctoral positions at the Institute for Advanced Study, the Simons Institute for the Theory of Computing, and Stanford University. He is interested in Boolean function analysis, complexity theory and combinatorics, and in particular in circuit complexity, query complexity, computational learning theory, quantum complexity, and the theory of pseudorandomness.