Electronic Colloquium on Computational Complexity, Report No. 30 (2018)

NP-hardness of Minimum Circuit Size Problem for OR-AND-MOD Circuits

Shuichi Hirahara*
The University of Tokyo

Igor C. Oliveira†
University of Oxford

Rahul Santhanam‡
University of Oxford

February 10, 2018

Abstract

The Minimum Circuit Size Problem (MCSP) asks for the size of the smallest boolean circuit that computes a given truth table. It is a prominent problem in NP that is believed to be hard, but for which no proof of NP-hardness has been found. A significant number of works have demonstrated the central role of this problem and its variations in diverse areas such as cryptography, derandomization, proof complexity, learning theory, and circuit lower bounds.

The NP-hardness of computing the minimum numbers of terms in a DNF formula consistent with a given truth table was proved by W. Masek [Mas79] in 1979. In this work, we make the first progress in showing NP-hardness for more expressive classes of circuits, and establish an analogous result for the MCSP problem for depth-3 circuits of the form OR-AND-MOD_2. Our techniques extend to an NP-hardness result for MOD_m gates at the bottom layer under inputs from $(\mathbb{Z}/m\mathbb{Z})^n$.

*hirahara@is.s.u-tokyo.ac.jp
†igor.carboni.oliveira@cs.ox.ac.uk
‡rahul.santhanam@cs.ox.ac.uk

ISSN 1433-8092

Contents

1 Introduction

1.1 The Minimum Circuit Size Problem

In the Minimum Circuit Size Problem (MCSP), we are given the truth table of a Boolean function as input together with a positive integer s, and the question is whether a circuit of size at most s exists for the function represented by this truth table. It is easy to see that MCSP is in NP: simply guess a circuit C of size at most s, and check that C computes each entry of the truth table correctly.

When solving MCSP deterministically, though, it is unclear how to avoid exhaustive search over the space of circuits of size at most s. A natural question arises: is MCSP NP-complete? The answer to this problem remains far from clear. MCSP is one of the very few natural problems in NP for which we have no strong evidence *for* or *against* NP-completeness. This is despite the fact that MCSP has long been recognized as a fundamental problem since the earliest research on complexity theory in the Soviet Union in the 1950s [Tra84]. Indeed, it is reported in [AKRR11] that Levin delayed the publication of his NP-completeness results for Satisfiability because he was hoping to show similar results for MCSP.

The difficulty of showing MCSP to be NP-hard was explicitly addressed in the work of Kabanets and Cai [KC00]. Roughly speaking, suppose we have a polynomial-time reduction f from Satisfiability to MCSP that is "natural", in the sense that the output length and output parameter depend only on the input length, and the input length is polynomially bounded in the output length – this is a property that all standard reductions have. Kabanets and Cai argued that by applying f to a trivial family of unsatisfiable formulas, we can show that the class E of problems solvable in linear exponential time requires superpolynomial circuit size. Given that the question of proving super-polynomial circuit lower bounds for explicit functions is a longstanding open question in complexity theory, this provides a significant obstacle to showing NP-hardness of MCSP via natural reductions. Note, though, that the Kabanets-Cai result does not give any evidence *against* NP-hardness of MCSP – it only suggests that NP-hardness might be hard to *establish*. There has been a long sequence of works [AHK15, MW15, HP15, HW16, AH17] building on this result to give further evidence of the difficulty of showing NP-hardness of MCSP.

One way around the Kabanets-Cai obstacle is to study the complexity of MCSP for circuit classes for which strong circuit lower bounds are *already known*. Given a class $\mathfrak{C}$ of circuits, let $\mathfrak{C}$-MCSP be the problem where, given a truth table and a number s, we wish to know if there is a $\mathfrak{C}$-circuit of size s computing the given truth table.

Studying $\mathfrak{C}$-MCSP for restricted classes $\mathfrak{C}$ of circuits is independently motivated by algorithmic applications in circuit minimization, proof complexity [Kra11, Chapter 30], learning theory (cf. [PV88, AHM+08, Fel09, CIKK16]), and cryptography and circuit lower bounds [RR97] (see also [BR17]). It was shown already in 1979 by Masek [Mas79] that DNF-MCSP is NP-hard.[1] There have been different proofs of this result [Czo99, AHM+08], and extensions to hardness of approximation [AHM+08, Fel09, KS08]. Nevertheless, almost four decades after Masek's result, and despite the significant attention that the MCSP problem has received (see also [ABK+06, AD14, AGM15, OS17]), NP-hardness of $\mathfrak{C}$-MCSP was not known for any natural class $\mathfrak{C}$ of circuits more expressive than DNFs.

Going beyond DNFs, for constant-depth threshold circuits and constant-depth Boolean circuits of large enough depth, there is cryptographic evidence that $\mathfrak{C}$-MCSP is not in polynomial time (see

[1] For a self-contained presentation of a proof of NP-hardness of DNF-MCSP, see [AHM+08].

e.g. [AHM+08]). But for classes extending DNFs that are not known to compute pseudorandom functions, no evidence of any sort for hardness was known. To quote Allender et al. [AHM+08], "Thus an important open question is to resolve the NP-hardness of both learnability results as well as function minimization results above for classes that are stronger than DNF."

1.2 Our Result

The main contribution in this work is the first NP-hardness result for $\mathfrak{C}$-MCSP for a class $\mathfrak{C}$ of depth–3 circuits, namely the class of (unbounded fan-in) $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{MOD}_m$ circuits, where m is any integer.

Theorem 1 (Main Result)**.** *For every $m \geq 2$, given the truth table of a function $f\colon \mathbb{Z}_m^n \to \{0,1\}$, where $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z} = \{0,1,\ldots,m-1\}$, it is NP-hard under polynomial-time deterministic many-one reductions to determine the size of the smallest $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{MOD}_m$ circuit C that computes f, where circuit size is measured as the top fan-in of C.*[2]

A few comments are in order. First, we elaborate on our computational model and complexity measure. We work with circuits which have an OR gate at the top, AND gates at the middle level, and MOD_m gates at the bottom level. We refer to such circuits as OR-AND-MOD circuits, or equivalently, DNF-MOD circuits. Such circuits operate in a natural way on inputs from $\mathbb{Z}_m^n$. We allow arbitrary constants from $\mathbb{Z}_m$ to feed in to gates at the bottom layer, and insist that inputs to the middle AND layer are Boolean. In other words, a MOD_m gate outputs 1 if and only if its corresponding linear equation over $\mathbb{Z}_m$ is satisfied, and the computations beyond the first layer are all Boolean. For $m = 2$, this is precisely the traditional model of DNF of Parities (cf. [CS16], [Juk06], [Juk12, Section 11.9], [ABG+14]).

The complexity measure we use is the top fan-in of the circuit, i.e., fan-in to the top OR gate. The main reason we work with this measure is naturalness and convenience. As argued in [CS16], top fan-in is the preferred measure for OR-AND-MOD_2 circuits because: (i) it measures the number of affine subspaces required to cover the 1s of the function, and thus has a nice combinatorial meaning; (ii) the number of MOD_2 gates feeding in to any middle layer AND gate can be assumed to be at most n without loss of generality, by using basic linear algebra, and thus the top fan-in approximates the total number of gates to within a factor of n; and (iii) the size of a DNF is often measured by the number of terms in it, and analogously it makes sense to measure the size of a DNF of Parities by the top fan-in of the circuit.

Our results are not however critically dependent on the complexity measure we use, and admit different extensions. Indeed, we demonstrate the robustness of our techniques by adapting them to show a hardness result for computing the number of gates in OR-AND-MOD_p formulas, where p is prime (Appendix B). Moreover, we mention that our approach can be modified to show a hardness of approximation result (Appendix C).

The strategy for the proof of Theorem 1 is explained in Section 1.3. In short, we reduce from a variant of the well-known set cover problem [Kar72]. The reduction consists of two stages, and it is initially presented as a randomized reduction. As one ingredient in the derandomization of

[2] As stated, Theorem 1 refers to the complexity of the optimization problem of finding the smallest circuit size for a given truth table, rather than the MCSP decision problem as defined. Note however that these two computational problems are easily seen to be polynomial-time equivalent to each other.

our approach, we show the existence of near-optimal (seed length $O(\log n + \log 1/\varepsilon)$) pseudorandom generators against $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits over $\mathbb{Z}_m^n$ of arbitrary size. This result might be of independent interest, and we refer to the discussion in Section 1.3 for more details.

Before further exploring the ideas of our proof, we give some perspective on the result and the possibility of extending it to more expressive circuit classes. Using the Kabanets-Cai [KC00] connection between NP-hardness and circuit lower bounds mentioned before, it is not hard to show that our reduction yields a $2^{\Omega(n)}$ lower bound on the size of DNF-MOD$_2$ circuits for a function in $\mathsf{E} = \mathsf{DTIME}[2^{O(n)}]$. Such strong exponential lower bounds for explicit functions have long been known for the model we consider (see e.g. [Gro98]). On the other hand, extending the NP-hardness result even to slightly different classes such as depth-3 AC^0 circuits might be a challenge. It is still unknown if E requires depth-3 AC^0 circuits of size $2^{\Omega(n)}$, and using the Kabanets-Cai connection, natural approaches to an NP-hardness result would imply such a lower bound.

What might be more feasible though is showing NP-hardness of $\mathfrak{C}$-MCSP for other related classes $\mathfrak{C}$ of circuits, and under weaker kinds of reductions, such as quasi-polynomial time reductions or non-uniform reductions. For instance, it might be possible to extend our techniques to classes such as $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{MOD}$ and depth-3 AC^0 circuits of small bottom fan-in. In these cases, exponential lower bounds of the form $2^{\Omega(n)}$ have been obtained (cf. [Gro98], [PSZ00]).

More broadly, we believe that showing NP-hardness of MCSP for more expressive classes $\mathfrak{C}$ is an important direction in better *understanding* circuit classes from the perspective of *meta-complexity*, i.e., complexity questions about computational problems involving circuits and algorithms. There are various criteria for measuring our understanding of a circuit class, for example, (*i*) Can we design non-trivial satisfiability algorithms for circuits in the class? (*ii*) Can we unconditionally construct pseudo-random generators secure against circuits in the class? (*iii*) Can we learn the class using membership queries under the uniform distribution? (*iv*) Can we prove lower bounds against proof systems whose lines are encoded by circuits in the class? We suggest that the NP-hardness of $\mathfrak{C}$-MCSP is another strong indication that we understand a circuit class $\mathfrak{C}$ well.

1.3 Overview of the Proof of Theorem 1

The rest of the paper is dedicated to the proof of Theorem 1, which will be completed in Section 4. Here we provide a high-level description of the reduction. For simplicity, our exposition mostly focus on the case $m = 2$. After that, we explain the main difficulties in extending the result to general m, and how these are addressed in our proof.

As mentioned above, Masek [Mas79] was the first to establish the NP-hardness of DNF minimization, and Theorem 1 can be interpreted as an extension of Masek's result to the more expressive DNF-MOD circuits. The structure of our argument follows however a *two-step* reduction introduced by Gimpel (cf. Allender et al. [AHM+08]), brought to our attention thanks to an alternative proof of Masek's result from [AHM+08]. More precisely, their work presents a new proof of the first stage of Gimpel's reduction, and provides a self-contained exposition of the entire argument.

Our NP-hardness proof for DNF-MOD circuits heavily builds on ideas of Gimpel and [AHM+08], but the extension to depth-3 requires new ideas and makes the argument much more involved. Let $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP be the computational problem described in Theorem 1 when $m = 2$, and let $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* be its natural generalization to *partial* boolean functions. In other words, an input to $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* encodes the truth table of a function $f\colon \{0,1\}^n \to \{0,1,*\}$, and we are interested in the size of the minimum $(\mathsf{DNF} \circ \mathsf{XOR})$-circuit that agrees with f on $f^{-1}(\{0,1\})$. Let $r \in \mathbb{N}$ be a large enough constant. Our proof reduces from the NP-complete problem r-Bounded Set

Cover (cf. [GJ79]): Given a set system $\mathcal{S} \subseteq \binom{n}{\leq r}$ that covers $[n]$, determine the minimum number ℓ of sets $S_1, \ldots, S_\ell \in \mathcal{S}$ such that $\bigcup_{i=1}^{\ell} S_i = [n]$. (We refer to Section 2.2 for a precise formulation of these computational problems.)

In a bit more detail, we present a *randomized* (2-approximate) reduction from r-Bounded Set Cover to $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP*, and a *randomized* reduction from $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* to $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP. These reductions are then efficiently derandomized using an appropriate pseudorandom generator. As opposed to previous works on the NP-hardness of DNF minimization, our proof crucially explores the fact that r-Bounded Set Cover is NP-hard even to *approximate* (by roughly a $\ln r$-factor), a result from [Fei98, Tre01] (see Theorem 5, Section 2.2).

We discuss each reduction in more detail now. Common to both of them is a convenient characterization of the sets $C^{-1}(1) \subseteq \{0,1\}^n$ of inputs that can be accepted by non-trivial $\mathsf{AND} \circ \mathsf{XOR}$ circuits C. If m is prime, it is not hard to show that this is precisely the class of affine subspaces of $\{0,1\}^n$. Consequently, for a non-trivial partial function $f\colon \{0,1\}^n \to \{0,1,*\}$, its corresponding $\mathsf{DNF}_{\mathsf{XOR}}(f)$ complexity is exactly the minimum number t of affine subspaces $A_1, \ldots, A_t \subseteq \{0,1\}^n$ such that $f^{-1}(1) \subseteq \bigcup_{i=1}^{t} A_i$ and $\bigcup_{i=1}^{t} A_i \subseteq f^{-1}(\{1,*\})$ (see Section 2.1). The analysis of our polynomial-time reductions, which will not be covered in this section, rely on this characterization in fundamental ways.

Step 1. *A randomized reduction from r-Bounded Set Cover to* $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* (Section 3.1).

Given a set-system $\mathcal{S} \subseteq \binom{n}{\leq r}$, we define a partial boolean function $f\colon \{0,1\}^t \to \{0,1,*\}$, where $t = O(r \log n)$. This function is *probabilistically* constructed as follows. First, we associate to each $i \in [n]$ a *random* vector $v^i \in \{0,1\}^t$. For $S \in \mathcal{S}$, let $v^S = \{v^i \mid i \in S\}$. Then, we let f be 1 on each input v^i, 0 on inputs that are *not* in the linear span of v^S for every $S \in \mathcal{S}$, and $*$ elsewhere.

Using this construction, we are able to show by a delicate analysis that if t is sufficiently large, the following holds with high probability: if $\mathcal{S}$ admits a cover of size K, then $\mathsf{DNF}_{\mathsf{XOR}}(f) \leq K$; moreover, if $\mathsf{DNF}_{\mathsf{XOR}}(f) \leq K$, then $\mathcal{S}$ admits a cover of size $\leq 2K$. (We discuss the intuition for this claim in Section 3.1.) This construction and the hardness of approximation result for r-Bounded Set Cover imply that $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* is NP-hard under many-one randomized reductions.

Step 2. *A randomized reduction from* $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* *to* $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP (Section 3.2).

Let $f\colon \{0,1\}^t \to \{0,1,*\}$ be an instance of $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP*. We *probabilistically* construct from f a related *total* function $g\colon \{0,1\}^t \times \{0,1\}^s \to \{0,1\}$, where $r = t+2$ and $s = O(r+t)$. In more detail, we encode for each $x \in \{0,1\}^t$ its corresponding value $f(x) \in \{0,1,*\}$ as a *boolean function* g_x on a hypercube $\{0,1\}^s$. For an input x such that $f(x) \in \{0,1\}$, we let $g(x0^s) = g_x(0^s) = f(x)$, where $g_x(\cdot) = 0$ elsewhere. On the other hand, if $f(x) = *$, we pick a *random* linear subspace $L_x \subseteq \{0,1\}^s$ of dimension r, and we encode $f(x)$ as the characteristic function of L_x.

Again, a careful argument allows us to establish the following connection between the partial function f and the total function g: with high probability over the choice of the random linear subspaces $(L_x)_{x \in f^{-1}(*)}$, $\mathsf{DNF}_{\mathsf{XOR}}(g) = \mathsf{DNF}_{\mathsf{XOR}}(f) + |f^{-1}(*)|$. (We discuss the intuition for this claim in Section 3.2.) Consequently, it follows from this and the previous reduction that $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP is NP-hard under many-one randomized reductions.

Step 3. *Efficient derandomization of the reductions* (Section 4.1).

It is possible to prove that the first reduction is always correct provided that the collection

of random vectors v^i is *nice* with respect to the set-system $\mathcal{S}$ (Definition 12). Similarly, we can prove that the second reduction is correct whenever the collection $(L_x)_{x \in f^{-1}(*)}$ of linear subspaces is *scattered* (Definition 18). It turns out that both conditions can be checked in polynomial time. This implies that the previously discussed reductions are in fact *zero-error* reductions. Consequently, if we can efficiently construct nice vectors and scattered families of linear subspaces, the reductions can be made deterministic.

In order to achieve this, we use in both cases a subtle derandomization argument that relies on (polynomial-time computable) ε-biased distributions [NN93a]. Recall that such distributions can fool arbitrary linear tests. By a more careful analysis, it is also known that they fool $\mathsf{AND} \circ \mathsf{XOR}$ circuits. We do *not* describe an $\mathsf{AND} \circ \mathsf{XOR}$ circuit to check if a collection of vectors is nice, or to check if a collection of linear subspaces is scattered. Still, we are able to show that if $\varepsilon < 2^{-s}$ then some scattered collection of linear subspaces is encoded by a string in the support of an ε-biased distribution, and that the same holds with respect to a nice collection of vectors if $\varepsilon < 2^{-t}$. In particular, trying all possible seeds of an ε-biased generator produces the combinatorial and algebraic objects that are sufficient to derandomize our reductions. (We refer to Section 4.1 for more details.)

Overall, combining the (derandomized) reductions and using the hardness of approximation result for r-Bounded Set Cover mentioned above, it follows that $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP is NP-hard under many-one deterministic polynomial-time reductions.

The argument for arbitrary $m \geq 2$. Let $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP and $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP* be the corresponding computational problems with respect to an arbitrary $m \geq 2$. (Recall that the input boolean functions in this case are defined over $\mathbb{Z}_m^n$.) As we explain next, additional difficulties are present for general m.

An immediate challenge is that it is no longer clear if the analogue characterization (via affine subspaces) of the class of subsets of $\mathbb{Z}_m^n$ accepted by non-trivial $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits holds, and this is crucially exploited when $m = 2$. The main issue is that, while in the latter case the result can be established by elementary techniques using that $\mathbb{Z}_2^n$ is a *vector space* over $\mathbb{Z}_2$, for an arbitrary m the underlying structure might be just a *module*. Without a *basis*, the result is less clear.

Nevertheless, it is possible to prove that the analogue result for $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits hold (cf. Lemma 2). The alternative and more general argument relies on a property of double orthogonal complements in $\mathbb{Z}_m^n$ (Appendix A), and we refer to Section 2.1 for more details. Armed with this characterization, the reductions discussed before can be adapted to arbitrary m. Finding the right generalization of each definition requires some work, but after that, the *randomized* reductions for $m = 2$ and arbitrary $m \geq 2$ can be presented in a unified and transparent way.

In order to conclude the proof of Theorem 1, we need to derandomize the new reductions. For $m = 2$, the argument was based on an efficient construction of ε-biased distributions supported over $\{0,1\}^n$, and the fact that such distributions are also able to fool $\mathsf{AND} \circ \mathsf{XOR}$ circuits over $\{0,1\}^n$. Without going into further details, we mention that for arbitrary m it is sufficient to use a pseudorandom generator that fools $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits over $\mathbb{Z}_m^n$. However, a generator with *near-optimal* dependency on n and ε is needed if we are hoping to obtain a polynomial-time reduction. We were not able to find such a result in the literature.[3]

[3] Existing generators seem to generate *bits* only, or are restricted to prime modulus, or can handle larger classes of functions but are not efficient enough for our purposes. We refer to [GKM15] and the references therein for related results.

We show in Section 4.2 that, for every $m \geq 2$, there is an efficient pseudorandom generator $G_n \colon \{0,1\}^{O(\log n + \log 1/\varepsilon)} \to \mathbb{Z}_m^n$ that ε-fools $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits of arbitrary size. Our construction relies on the efficient ε-biased generators for $\mathbb{Z}_m^n$ from [AMN98], together with a proof of the following result: If G is an ε-biased generator against $\mathbb{Z}_m^n$, then G $(m\varepsilon)$-fools $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits. Again, we cannot rely on a adaptation of the similar claim for $m = 2$, which requires a basis. Our proof proceeds instead by a careful analysis of certain exponential sums encoding the behaviour of the circuit, and that can be used to connect the distinguishing probability to the guarantees offered by the ε-biased generator. We refer to Section 4.2 for more details.

2 Preliminaries

Notation. For an integer $n \geq 1$, let $[n]$ denote $\{1, \ldots, n\}$.

Some notions from group theory. Let $m \geq 2$ be a constant. Let $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z}$ denote the integers modulo m, where all operations on elements in $\mathbb{Z}_m = \langle +, \{0, 1, \ldots, m-1\} \rangle$ are taken mod m. For any integer $t \geq 1$, we regard $\mathbb{Z}_m^t$ as an additive group with component-wise addition. A non-empty subset $H \subseteq \mathbb{Z}_m^t$ is called a *linear subspace* if H is a subgroup, that is, $0 \in H$ and $x + y \in H$ for any $x, y \in H$. A subset $A \subseteq \mathbb{Z}_m^t$ is called an *affine subspace* if A is a coset, that is, there exist $a \in \mathbb{Z}_m^t$ and a linear subspace $H \subseteq \mathbb{Z}_m^t$ such that $A = H + a := \{ h + a \mid h \in H \}$.

We stress that $\mathbb{Z}_m^t$ gives rise to a *module* and not to a *vector space* when m is a composite number; however, we borrow some standard notation; for example, for a scalar $c \in \mathbb{Z}_m$ and a "vector" $v \in \mathbb{Z}_m^t$, let cv denote the scalar multiplication. Let $\langle x, y \rangle := \sum_{i=1}^t x_i y_i \ \ (\in \mathbb{Z}_m)$ for any $x, y \in \mathbb{Z}_m^t$ and $t \in \mathbb{N}$.

2.1 Circuit Size Measure and Its Characterization

For any integer $m \geq 2$, an $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{MOD}_m$ $(= \mathsf{DNF} \circ \mathsf{MOD}_m)$ circuit is a DNF formula whose terms are $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits. Here, the MOD_m gate is a Boolean function such that $\mathsf{MOD}_m(x) = 1$ if and only if $\sum_{i=1}^n x_i \mod m = 0$ on input $x \in \{0,1\}^n$. We allow multiple input wires and access to constant input bits in the circuit. Note that this allows for more general equations to be computed by a bottom-layer modular gate.

The size of a circuit is usually defined as the number of gates. However, for us it is important to define the size of a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit as the top fan-in of the circuit, or equivalently, its number of $\mathsf{AND} \circ \mathsf{MOD}_m$ terms. (Note that the same size measure was used in [CS16] in the case $m = 2$.) For a Boolean function $f \colon \{0,1\}^t \to \{0,1\}$, define $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ as the minimum number of terms of a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing f, i.e., the fan-in of its OR gate.

In order to present our results in a unified way for any integer $m \geq 2$, we extend the input $\{0,1\}^t$ of a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit to the larder domain $\mathbb{Z}_m^t$ in a natural way: that is, we regard the bottom MOD_m gate as a function $\mathsf{MOD}_m \colon \mathbb{Z}_m^* \to \{0,1\}$ that outputs 1 if and only if the sum of its input elements is congruent to 0 mod m. Again, more general equations can be obtained using multiple input wires and access to constants in $\mathbb{Z}_m$.

An $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C *accepts* the set $X \subseteq \mathbb{Z}_m^t$ if for any $x \in \mathbb{Z}_m^t$, $x \in X$ if and only if C outputs 1 on x. There is a nice combinatorial characterization of the set of inputs that such circuits can accept.

Lemma 2 (Characterization of the power of $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits)**.** *Let $X \subseteq \mathbb{Z}_m^t$ be a nonempty set. Then, an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit accepts X if and only if X is an affine subspace of $\mathbb{Z}_m^t$.*

This is a standard fact when m is a prime power (cf. [CS16] for $m = 2$), in which case $\mathbb{Z}_m^t$ is a vector space. However, the same characterization holds when $m \geq 2$ is an arbitrary composite number, as established below. The proof relies on the following fact about orthogonal complements in the more general context of modules.

Fact 3 (Double orthogonal complement)**.** *Let $H \subseteq \mathbb{Z}_m^t$ be a linear subspace, and let $H^\perp := \{\, x \in \mathbb{Z}_m^t \mid \sum_{i=1}^t x_i y_i = 0$ for any $y \in H \,\}$ be its orthogonal complement. Then, $(H^\perp)^\perp = H$.*

For completeness, we include a proof of this result in Appendix A. Assuming Fact 3, we proceed to a proof of Lemma 2.

Proof of Lemma 2. Let $x := (x_1, \ldots, x_t) \in \mathbb{Z}_m^t$ denote the input to the circuit.

Suppose that an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit $\bigwedge_{k=1}^K C_k$ accepts X, where each C_k is a MOD_m gate. Each MOD_m gate C_k in the circuit defines a linear equation over $(x_1, \ldots, x_t)$. That is, there are coefficients $a_k^1, \ldots, a_k^t \in \mathbb{Z}_m$ and an element $b_k \in \mathbb{Z}_m$ such that $\sum_{i=1}^t a_k^i x_i = b_k$ if and only if C_k accepts the input x. Therefore, the circuit $\bigwedge_{k=1}^K C_k$ accepts the intersection of such linear equations over $\mathbb{Z}_m$. Specifically, for a matrix $A := (a_k^i)_{k\in[K],i\in[t]}$ and a vector $b := (b_k)_{k\in[K]}$, the circuit accepts all inputs $x \in \mathbb{Z}_m^t$ such that $Ax = b$; namely, $X = \{\, x \in \mathbb{Z}_m^t \mid Ax = b \,\}$. Since X is nonempty, we can take some element $x_0 \in X$. Now, we can rewrite X as

$$X = \{\, x \in \mathbb{Z}_m^t \mid A(x - x_0) = 0 \,\} = \{\, y \in \mathbb{Z}_m^t \mid Ay = 0 \,\} + x_0,$$

which is an affine subspace of $\mathbb{Z}_m^t$.

For the converse direction, we use the notion of orthogonal complement. Suppose that $X \subseteq \mathbb{Z}_m^t$ is an affine subspace. By definition, we can decompose X into a linear subspace $H \subseteq \mathbb{Z}_m^t$ and a shift $a \in \mathbb{Z}_m^t$ so that $X = H + a$.

We first claim that H can be accepted by some $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit. To prove this, it is sufficient to show the existence of some matrix $A \in \mathbb{Z}_m^{K\times t}$ such that $H = \{\, x \in \mathbb{Z}_m^t \mid Ax = 0 \,\}$. Since H is a linear subspace, by Fact 3, for any $x \in \mathbb{Z}_m^t$,

$$x \in H \quad \text{if and only if} \quad \sum_{i=1}^t x_i \cdot y_i = 0 \text{ for every } y \in H^\perp.$$

That is, we can define a matrix $A \in \mathbb{Z}_m^{|H^\perp|\times t}$ as $(y_i)_{y\in H^\perp,i\in[t]}$. (In other words, for each $y \in H^\perp$, we add a MOD_m gate that checks if $\sum_{i=1}^t x_i \cdot y_i = 0$, where each coefficient y_i is simulated using multiple input wires.)

To accept X, we just need to shift H by a. Indeed, for a vector $b := Aa$, we have $X = H + a = \{\, x \in \mathbb{Z}_m^t \mid Ax = b \,\}$; thus we can construct an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit accepting X by simulating the condition $Ax = b$. □

As a consequence of Lemma 2, for a function $f\colon \mathbb{Z}_m^t \to \{0, 1\}$, the minimum size of a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing f equals the minimum number S of affine subspaces $T_1, \ldots, T_S \subseteq \mathbb{Z}_m^t$ such that $\bigcup_{i=1}^S T_i = f^{-1}(1)$.

2.2 Computational Problems

The starting point of our NP-hardness results is the set cover problem on instances where each set has size at most r.

Definition 4 (r-Bounded Set Cover Problem)**.** *For an integer $r \in \mathbb{N}$, the r-Bounded Set Cover Problem is defined as follows:*

- Input. *An integer $n \in \mathbb{N}$ and a collection $\mathcal{S} \subseteq 2^{[n]}$ of nonempty subsets of the universe $[n]$ such that $|S| \leq r$ for each $S \in \mathcal{S}$, and $\bigcup_{S \in \mathcal{S}} S = [n]$.*
- Output. *The minimum number ℓ of subsets $S_1, \ldots, S_\ell \in \mathcal{S}$ such that $\bigcup_{i=1}^{\ell} S_i = [n]$.*

For this problem, a tight inapproximability result based on NP-hardness is known.

Theorem 5 (Feige [Fei98], Trevisan [Tre01])**.** *Let r be a sufficiently large constant. It is NP-hard (under polynomial-time many-one reductions) to approximate the solution of the r-bounded set cover problem within a factor of $\ln r - O(\ln \ln r)$. That is, for any language $L \in \mathsf{NP}$, there exists a polynomial-time machine that, on input x, outputs a threshold θ and an instance $\mathcal{S}$ of the r-bounded set cover problem such that if $x \in L$ then $\mathcal{S}$ has a cover of size at most θ, and if $x \notin L$ then $\mathcal{S}$ does not have a cover of size at most $\theta \cdot (\ln r - O(\ln \ln r))$.*

We stress that the *inapproximability* result is essential for us; we will present a reduction from a 2-factor approximation of the r-bounded set cover problem to the minimum $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit minimization problem.

Definition 6 (Minimum Circuit Size Problem for $\mathsf{DNF} \circ \mathsf{MOD}_m$)**.** *For an integer $m \geq 2$, the Minimum Circuit Size Problem for $\mathsf{DNF} \circ \mathsf{MOD}_m$, abbreviated as $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP, is defined as follows:*

- Input. *A Boolean function $f \colon \mathbb{Z}_m^t \to \{0,1\}$, represented as a truth table of length m^t.*
- Output. $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$.

While our final theorem confirms that $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP is NP-hard, we will first prove NP-hardness of the circuit minimization problem on instances of a partial function $f : \mathbb{Z}_m^t \to \{0,1,*\}$. That is, we regard any input $x \in f^{-1}(*)$ as "undefined." For a partial function $f \colon \mathbb{Z}_m^t \to \{0,1,*\}$, we say that a circuit C computes f if $C(x) = f(x)$ for any $x \in f^{-1}(\{0,1\})$. We extend the definition of $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ to the size of the minimum $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing the partial function $f \colon \mathbb{Z}_m^t \to \{0,1,*\}$. The following problem is concerned with the circuit size of partial functions, and we distinguish it from the problem above by adding a superscript $*$.

Definition 7 (Minimum Circuit Size Problem for Partial Functions)**.** *For an integer $m \geq 2$, the Minimum Circuit Size Problem* for $\mathsf{DNF} \circ \mathsf{MOD}_m$, abbreviated as $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP*, is defined as follows:*

- Input. *A Boolean function $f \colon \mathbb{Z}_m^t \to \{0,1,*\}$, represented as a string of length m^t over the alphabet $\{0,1,*\}$.*
- Output. $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$.

3 Hardness of $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP Under Randomized Reductions

3.1 Reduction from r-Bounded Set Cover to $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP*

This subsection is devoted to proving the following theorem.

Theorem 8. $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP* *is* NP*-hard under (zero-error) randomized polynomial-time many-one reductions.*

Let r be a large enough constant so that the approximation factor of $\ln r - O(\ln \ln r)$ in Theorem 5 is larger than 2. We present a reduction from a 2-factor approximation of the r-bounded set cover problem to $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP*.

Let us prepare some notation. Let $\mathcal{S}$ be an instance of the r-bounded set cover problem over the universe $[n]$ (in particular, $\bigcup_{S \in \mathcal{S}} S = [n]$). Let $t \in \mathbb{N}$ be a parameter chosen later. For each $i \in [n]$, pick $v^i \in_R \mathbb{Z}_m^t$ independently and uniformly at random. For any $S \subseteq [n]$, let v^S denote $\{ v^i \mid i \in S \}$. Let $\mathsf{span}(v^S) := \{ \sum_{i \in S} c_i \cdot v^i \mid c_i \in \mathbb{Z}_m \text{ for any } i \in S \}$ denote the linear span of v^S. (Note that $\mathsf{span}(v^S)$ is a linear subspace of $\mathbb{Z}_m^t$ whenever $S \neq \varnothing$.) In our reduction, an element $i \in [n]$ is mapped to a random point v^i of $\mathbb{Z}_m^t$, and a set $S \in \mathcal{S}$ corresponds to a linear subspace $\mathsf{span}(v^S)$.

For any set cover instance $\mathcal{S}$, we define a function $f : \mathbb{Z}_m^t \to \{0, 1, *\}$ as

$$f(x) := \begin{cases} 1 & (\text{if } x = v^i \text{ for some } i \in [n]) \\ 0 & (\text{if } x \notin \bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S)) \\ * & (\text{otherwise}) \end{cases}$$

for any $x \in \mathbb{Z}_m^t$. The truth table of f is the output of our reduction.

It is not hard to see that $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ is at most the minimum set cover size for $\mathcal{S}$ (Claim 9 below). Of course, the difficulty is in proving a circuit lower bound for f (Claim 10 below).

The idea is as follows: For simplicity of the exposition, let us focus on the case of $m = 2$, and moreover let us first consider the case of a $\mathsf{DNF} \circ \mathsf{MOD}_2$ circuit C for f that accepts a union of *linear* subspaces (instead of affine subspaces). More precisely, let $C^{-1}(1)$ be a union of linear subspaces $\{T_k\}_{k \in [K]}$. Then T_k is a subset of $C^{-1}(1) \subseteq f^{-1}(\{1, *\}) = \bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S)$; furthermore, each $\mathsf{span}(v^S)$ is a random linear subspace of small dimension r; therefore, it is possible to show that, with high probability, the set $\{ i \in [n] \mid v^i \in T_k \}$ of points covered by T_k is contained in some legal set $S \in \mathcal{S}$ of the set cover instance; hence the circuit size K is at least the minimum set cover size.

In the case that a circuit C accepts the union of *affine* subspaces, it is no longer true that, for any affine subspace T such that $T \subseteq \bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S)$, the set $\{ i \in [n] \mid v^i \in T \}$ is covered by some legal set $S \in \mathcal{S}$; indeed, for any two points v^i and v^j, the set $\{v^i, v^j\}$ $\left(= v^i \oplus \{0, v^i \oplus v^j\}\right)$ is an affine subspace of $\mathbb{Z}_2^t$, whereas $\{i, j\}$ is not necessarily legal in the set cover instance $\mathcal{S}$. Nonetheless, we can still prove that, with high probability, the set $\{ i \in [n] \mid v^i \in T \}$ is covered by two legal sets $S_1, S_2 \in \mathcal{S}$. As a consequence, the minimum number of affine subspaces needed to cover $v^1, \ldots, v^n$ gives us a 2-factor approximation of the minimum set cover size for $\mathcal{S}$. By Theorem 5, it follows that $(\mathsf{DNF} \circ \mathsf{XOR})$-MCSP* is NP-hard under randomized reductions. Details follow.

Claim 9 (Easy part)**.** *Suppose that $\mathcal{S}$ has a set cover of size K. Then $\mathsf{DNF}_{\mathsf{MOD}_m}(f) \leq K$.*

Proof. Let $\mathcal{C} \subseteq \mathcal{S}$ be a set cover of size K. For each $S \in \mathcal{C}$, by Lemma 2, there exists an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C_S such that C_S accepts $\mathsf{span}(v^S)$. Define a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit $C := \bigvee_{S \in \mathcal{C}} C_S$. It is easy to see that C computes f. □

Conversely, we prove the following:

Claim 10 (Hard part)**.** *For some parameter t such that $m^t = (nm)^{O(r)}$, the following holds with probability at least $\frac{1}{2}$ (over the choice of $(v^i)_{i \in [n]}$):*

Let $K := \mathsf{DNF}_{\mathsf{MOD}_m}(f)$. Then $\mathcal{S}$ has a set cover of size $2K$.

The two claims above imply that $2\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ is a 2-factor approximation for the set cover problem: indeed, let s be the minimum set cover size for $\mathcal{S}$; then we have $s \leq 2\mathsf{DNF}_{\mathsf{MOD}_m}(f) \leq 2s$. It thus remains to prove Claim 10.

To prove Claim 10, let us clarify the desired condition that random objects $(v^i)_{i \in [n]}$ should satisfy. For any $I \subseteq [n]$, define the *affine span* of v^I as

$$\mathsf{affine\text{-}span}(v^I) := \Big\{ \sum_{i \in I} c_i v^i \mid c_i \in \mathbb{Z}_m \text{ for } i \in I \text{ and } \sum_{i \in I} c_i = 1 \Big\}.$$

The important property of the affine span is that, if an affine subspace A covers the set v^I of points in $I \subseteq [n]$, then its affine span must also be covered by A.

Claim 11 (Property of the affine span)**.** *For any affine subspace A of $\mathbb{Z}_m^t$ and any $I \subseteq [n]$, if $v^I \subseteq A$ then $\mathsf{affine\text{-}span}(v^I) \subseteq A$.*

Proof. Let us write $A = H + a$ for some linear space $H \subseteq \mathbb{Z}_m^t$ and vector $a \in \mathbb{Z}_m^t$. Since $v^i \in v^I \subseteq A$ for each $i \in I$, there exists some vector $h^i \in H$ such that $v^i = h^i + a$. Take any coefficients $(c_i)_{i \in I}$ such that $c_i \in \mathbb{Z}_m$ and $\sum_{i \in I} c_i = 1$. Then,

$$\sum_{i \in I} c_i v^i = \sum_{i \in I} c_i (h^i + a) = \sum_{i \in I} c_i h^i + a \ \in \ H + a.$$

□

By Lemma 2, the circuit size of f equals the minimum number of affine subspaces $A_1, \ldots, A_K \subseteq f^{-1}(\{1, *\})$ such that $\bigcup_{i=1}^K A_i \supseteq f^{-1}(1)$. Intuitively, we would like to require that, if the set $v^I \ \big(\subseteq f^{-1}(1)\big)$ of points is covered by some affine subspace $A \subseteq f^{-1}(\{1, *\})$, then there exist two legal sets S_1, S_2 of the set cover instance $\mathcal{S}$ such that $I \subseteq S_1 \cup S_2$. In fact, one of these sets can be taken as a singleton:

Definition 12. *We say that $(v^i)_{i \in [n]}$ is* nice *(with respect to $\mathcal{S}$) if, for any $I \subseteq [n]$,*

$$\mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S) \quad \Longrightarrow \quad I \subseteq S_I \cup \{i_I\} \tag{1}$$

for some $S_I \in \mathcal{S}$ and $i_I \in [n]$.

We will prove that $(v^i)_{i \in [n]}$ is nice with probability at least $\frac{1}{2}$, and that for any nice $(v^i)_{i \in [n]}$, the minimum size of $\mathsf{DNF} \circ \mathsf{MOD}_m$ is a 2-factor approximation of the minimum set cover size. We prove the latter first:

Claim 13. *Let $(v^i)_{i\in[n]}$ be nice, and $K := \mathsf{DNF}_{\mathsf{MOD}_m}(f)$. Then $\mathcal{S}$ has a set cover of size $2K$.*

Proof. Let $C = \bigvee_{k=1}^{K} C_k$ be a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing f, where each $C_k \in \mathsf{AND} \circ \mathsf{MOD}_m$ is nontrivial. By Lemma 2, $C_k^{-1}(1)$ is an affine subspace of $\mathbb{Z}_m^t$. For each C_k, we will choose 2 sets from $\mathcal{S}$ so that the union of all these sets cover the universe $[n]$.

Fix any C_k and let $I_k := \{\, i \in [n] \mid C_k(v^i) = 1 \,\}$ be the set of all points covered by C_k. Since $C_k^{-1}(1)$ is an affine subspace of $\mathbb{Z}_m^t$ and $v^{I_k} \subseteq C_k^{-1}(1)$, we have $\mathsf{affine\text{-}span}(v^{I_k}) \subseteq C_k^{-1}(1)$ by Claim 11. Since the circuit C computes f, $C_k^{-1}(1) \subseteq C^{-1}(1) \subseteq f^{-1}(\{1,*\}) = \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$. Thus we have $\mathsf{affine\text{-}span}(v^{I_k}) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$, which means that the hypothesis of niceness (1) is satisfied; hence there exist some subset $S_{k1} \in \mathcal{S}$ and some element $i_k \in [n]$ such that $I_k \subseteq S_{k1} \cup \{i_k\}$. Take any set $S_{k2} \in \mathcal{S}$ such that $i_k \in S_{k2}$ (such a set S_{k2} must exist because we assumed $\bigcup_{S\in\mathcal{S}} S = [n]$). Then $I_k \subseteq S_{k1} \cup S_{k2}$.

Now we claim that $\bigcup_{k=1}^{K} S_{k1} \cup S_{k2} = [n]$ (and hence the set cover instance $\mathcal{S}$ has a cover of size $2K$). Indeed, for any $i \in [n]$, we have $f(v^i) = 1$ and hence $C(v^i) = 1$, which means that there exists some subcircuit C_k such that $C_k(v^i) = 1$. Thus $i \in I_k \subseteq S_{k1} \cup S_{k2}$ for some $k \in [K]$. □

It remains to show that a random choice of $(v_i)_{i\in[n]}$ is nice with high probability:

Claim 14. *For each $i \in [n]$, pick $v^i \in_R \mathbb{Z}_m^t$ uniformly at random and independently. If $t \geq r + ((r+2)\log n + \log|\mathcal{S}| + 1)/\log m$, then $(v^i)_{i\in[n]}$ is nice with probability at least $\frac{1}{2}$.*

To prove Claim 14, we will use a union bound over all relevant subsets $I \subseteq [n]$; however, the definition of niceness (1) appears to suggest that we need to take a union bound over exponentially many subsets I. The next claim shows that this is in fact *not* the case.

Claim 15 (Characterization of niceness)**.** *$(v^i)_{i\in[n]}$ is not nice (with respect to $\mathcal{S}$) if and only if there exists some subset $I \subseteq [n]$ such that all the following conditions hold:*

1. *$|I| \leq r + 2$,*
2. *$I \not\subseteq S \cup \{i\}$ for any $S \in \mathcal{S}$ and $i \in [i]$, and*
3. *$\mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$.*

In particular, there are at most n^{r+2} subsets $I \subseteq [n]$ over which we need to take a union bound.

Proof. By the definition of niceness, $(v^i)_{i\in[n]}$ is not nice if and only if there exists some subset $I \subseteq [n]$ such that $\mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$ whereas $I \not\subseteq S \cup \{i\}$ for any $S \in \mathcal{S}$ and $i \in [i]$. Therefore, it is clear that the three conditions imply that $(v^i)_{i\in[n]}$ is not nice; we prove below the converse direction (the "only if" part of Claim 15).

A crucial observation is that, for any subset $I \subseteq [n]$ of size at least $r + 2$, the second condition always holds: Indeed, recall that $\mathcal{S}$ is an instance of the r-bounded set cover instance; that is, $|S| \leq r$ for any $S \in \mathcal{S}$. Hence, for any $S \in \mathcal{S}$ and $i \in [n]$, we have $|S \cup \{i\}| \leq r + 1$; thus I cannot be a subset of $S \cup \{i\}$ simply because $|I| \geq r + 2$.

Now suppose that there exists some subset $I \subseteq [n]$ satisfying the second and third conditions, but not the first one, that is, $|I| > r + 2$. Take any subset $I' \subseteq I$ such that $|I'| = r + 2$. We claim that I' satisfies all three conditions: The first condition ($|I'| \leq r + 2$) is obvious. The second condition holds because of the observation above. To see the third condition, by assumption, we have $\mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$; hence, we also have $\mathsf{affine\text{-}span}(v^{I'}) \subseteq \mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$. □

Now let us proceed to a proof of Claim 14.

Proof of Claim 14. We will bound the probability that a random $(v^i)_{i\in[n]}$ is not nice, by using the union bound over all the subsets $I \subseteq [n]$ such that the first and second conditions in Claim 15 hold. To this end, fix any subset $I \subseteq [n]$ such that $|I| \le r+2$ and $I \not\subseteq S \cup \{i\}$ for any $S \in \mathcal{S}$ and $i \in [n]$ (in particular, I is not empty). We would like to bound the probability that the affine subspace of v^I is a subset of $\bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)$.

Take an arbitrary (*e.g.* the smallest) element $i_0 \in I$. Define coefficients $(c_i)_{i\in I}$ as follows: $c_i := 1 \in \mathbb{Z}_m$ for any $i \in I \setminus i_0$ and $c_{i_0} := (2-|I|) \mod m \in \mathbb{Z}_m$. By this definition, we have $\sum_{i\in I} c_i = 1$; hence, $\sum_{i\in I} c_i v^i \in \mathsf{affine\text{-}span}(v^I)$. Therefore,

$$\begin{aligned}\Pr_{v^1,\dots,v^n}\left[\mathsf{affine\text{-}span}(v^I) \subseteq \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)\right] &\le \Pr\left[\sum_{i\in I} c_i v^i \in \bigcup_{S\in\mathcal{S}} \mathsf{span}(v^S)\right] \\ &\le \sum_{S\in\mathcal{S}} \Pr\left[\sum_{i\in I} c_i v^i \in \mathsf{span}(v^S)\right].\end{aligned}$$

By the assumption on I, we have $I \not\subseteq S \cup \{i_0\}$ for any $S \in \mathcal{S}$; that is, there exists some index $j_S \in I \setminus \{i_0\} \setminus S$. Note that $c_{j_S} = 1$ because $j_S \in I \setminus \{i_0\}$. Therefore, the last probability is

$$\begin{aligned}\sum_{S\in\mathcal{S}} \Pr\left[\sum_{i\in I} c_i v^i \in \mathsf{span}(v^S)\right] &= \sum_{S\in\mathcal{S}} \Pr\left[v^{j_S} \in \mathsf{span}(v^S) - \sum_{i\in I\setminus\{j_S\}} c_i v^i\right] \\ &= \sum_{S\in\mathcal{S}} \Pr\left[v^{j_S} = \sum_{i\in S} d_i v^i - \sum_{i\in I\setminus\{j_S\}} c_i v^i \text{ for some } (d_i)_{i\in S}\right] \\ &= \sum_{S\in\mathcal{S}} \sum_{(d_i)_{i\in S}} \Pr\left[v^{j_S} = \sum_{i\in S} d_i v^i - \sum_{i\in I\setminus\{j_S\}} c_i v^i\right] \\ &\le |\mathcal{S}| \cdot m^r \cdot m^{-t},\end{aligned}$$

where the last inequality holds because the random vector v^{j_S} does not appear in the right summations.

Finally, by taking the union bound over all I such that $|I| \le r+2$ (and $I \not\subseteq S\cup\{i\}$ for any $S \in \mathcal{S}$ and $i \in [n]$), the probability that $(v^i)_{i\in[n]}$ is not nice is bounded from above by $n^{r+2} \cdot |\mathcal{S}| \cdot m^{r-t} \le \frac{1}{2}$. □

Given these claims above, it is immediate to complete the whole proof.

Proof of Claim 10. We may assume without loss of generality that $|\mathcal{S}| \le n^r$ since $\mathcal{S}$ is an instance of the r-bounded set cover problem. We set $t \in \mathbb{N}$ to be the smallest integer such that $t \ge r + ((r+2)\log n + \log|\mathcal{S}| + 1)/\log m$; then $t = O(r\log(nm)/\log m)$. (Here the O notation hides only a universal constant.) Combining Claims 13 and 14, we immediately obtain Claim 10. □

Proof of Theorem 8. The encoding of the function $f\colon \mathbb{Z}_m^t \to \{0,1,*\}$ is of size $O(m^t) = (nm)^{O(r)}$, which is a polynomial in the input size $\mathsf{poly}(n, |\mathcal{S}|)$.

Moreover, it is possible to make the reduction zero-error: Indeed, the condition of the niceness can be checked in polynomial time, by using the characterization of Claim 15.

Finally, recall that the r-bounded set cover problem is NP-hard to approximate within a factor of 2 by Theorem 5 for a sufficiently large constant $r \in \mathbb{N}$. Hence, NP-hardness of $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ follows from Claims 9 and 10. □

3.2 Reduction from $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ to $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP

Next, we present a reduction for the minimum circuit size problem for partial functions to that for total functions:

Theorem 16. *There is a (zero-error) randomized polynomial-time many-one reduction from* $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ *to* $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP.

Let $f\colon \mathbb{Z}_m^t \to \{0,1,*\}$ be an instance of $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$. Let $r := t+2$ and $s := \lceil (2r+2t)\log m + 2 \rceil = \lceil 4(t+1)\log m + 2 \rceil$. We encode each value $f(x) \in \{0,1,*\}$ of the partial function f as a function on a "hypercube" $\mathbb{Z}_m^s$: namely, we construct a new *total* function $g : \mathbb{Z}_m^t \times \mathbb{Z}_m^s \to \{0,1\}$ such that $f(x)$ corresponds to $(g(x,y))_{y \in \mathbb{Z}_m^s}$. Specifically, if $f(x) \neq *$, then $f(x)$ is encoded as a hypercube whose origin[4] 0^s is assigned $f(x)$ and other points are assigned 0; if $f(x) = *$, then we pick a random linear subspace $L_x \subseteq \mathbb{Z}_m^s$ of dimension r and we encode $f(x)$ as the characteristic function of L_x.

Formally, for each $x \in f^{-1}(*)$, we pick $v_x^1, \ldots, v_x^r \in_R \mathbb{Z}_m^s$ uniformly and independently at random, and define a random linear subspace $L_x := \mathsf{span}(v_x^1, \ldots, v_x^r)$. Then the output $g : \mathbb{Z}_m^t \times \mathbb{Z}_m^s \to \{0,1\}$ of our reduction is defined as

$$g(x,y) := \begin{cases} f(x) & (\text{if } f(x) \in \{0,1\} \text{ and } y = 0^s) \\ 1 & (\text{if } f(x) = * \text{ and } y \in L_x) \\ 0 & (\text{otherwise}) \end{cases}$$

for any $(x,y) \in \mathbb{Z}_m^t \times \mathbb{Z}_m^s$.

The idea is as follows: Let us imagine how a minimum $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit C computing g looks like. We need to cover $g^{-1}(1)$ by as few affine subspaces as possible. Note that $g^{-1}(1)$ consists of two parts: $\{(x, 0^s)\}$ for each $x \in f^{-1}(1)$, and $\{x\} \times L_x$ for each $x \in f^{-1}(*)$. In order to cover the latter one, it is likely that we need to use the affine subspace $\{x\} \times L_x$ itself for each $x \in f^{-1}(*)$; indeed, since each L_x is a random linear subspace, under our constraints with high probability there is no affine subspace which simultaneously covers (a large fraction of) two random affine subspaces $\{x\} \times L_x$ and $\{x'\} \times L_{x'}$ for $x \neq x' \in f^{-1}(*)$ (Claim 21 below). Therefore, the minimum circuit C should contain a subcircuit which accepts $\{x\} \times L_x$ for each $x \in f^{-1}(*)$. Now it remains to cover $\{(x, 0^s)\}$ for each $x \in f^{-1}(1)$, but here we can *optionally* cover $\{(x, 0^s)\}$ for each $x \in f^{-1}(*)$ (which has been already covered by $\{x\} \times L_x$). This is exactly the same situation as $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$; thus with high probability we have $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)|$. Details follow.

Claim 17. $\mathsf{DNF}_{\mathsf{MOD}_m}(g) \leq \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)|$.

[4] 0^s denotes the zero of $\mathbb{Z}_m^s$ for any $s \in \mathbb{N}$.

Proof. Suppose that a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit $C = \bigvee_{k=1}^{K} C_k$ computes f. For each $x^* \in f^{-1}(*)$, take an $\mathsf{AND} \circ \mathsf{MOD}_m$ formula C_{x^*} such that $C_{x^*}^{-1}(1) = \{x^*\} \times L_{x^*}$ (by Lemma 2). Define $C'(x, y) := \bigvee_{k=1}^{K}(C_k(x) \wedge (y_1 = 0) \wedge \cdots \wedge (y_s = 0)) \vee \bigvee_{x^* \in f^{-1}(*)} C_{x^*}(x, y)$. It is easy to see that $C'(x, y) = g(x, y)$ for any $(x, y) \in \mathbb{Z}_m^t \times \mathbb{Z}_m^s$. □

In order to prove the other direction, let us clarify the desired condition for random linear spaces. We require that $(L_x)_{x \in f^{-1}(*)}$ is pairwise "disjoint" and that each L_x is nondegenerated.

Definition 18. *We say that* $(L_x)_{x \in f^{-1}(*)}$ *is* scattered *if* $|L_x| = m^r$ *and* $L_x \cap L_{x'} = \{0^s\}$ *for any distinct* $x, x' \in f^{-1}(*)$.

It is easy to prove that the collection of random linear spaces satisfies the condition above.

Claim 19. $(L_x)_{x \in f^{-1}(*)}$ *is scattered with probability at least* $\frac{1}{2}$*, provided that* $s \geq (2r + 2t) \log m + 2$.

Proof. We first bound the probability that $(L_x)_{x \in f^{-1}(*)}$ is not pairwise disjoint.

$$\begin{aligned}
&\Pr\left[\ L_x \cap L_{x'} \neq \{0^s\} \text{ for some distinct } x, x' \in f^{-1}(*)\ \right] \\
&\leq \sum_{x \neq x' \in f^{-1}(*)} \Pr\left[L_x \cap L_{x'} \neq \{0^s\}\right] \\
&\leq \sum_{x \neq x' \in f^{-1}(*)} \Pr\left[\sum_{i=1}^{r} c_i v_x^i = \sum_{i=1}^{r} d_i v_{x'}^i \text{ for some nonzero } (c_i)_{i \in [r]}, (d_i)_{i \in [r]}\right] \\
&< m^{2t} \cdot m^{2r} \cdot 2^{-s} \ \leq\ \frac{1}{4},
\end{aligned}$$

where, in the last line, we used the fact that the probability that $\sum_{i=1}^{r} c_i v_x^i = \sum_{i=1}^{r} d_i v_{x'}^i$ is at most 2^{-s} for nonzero (*i.e.* $c_i \neq 0, d_j \neq 0$ for some $i, j \in [r]$) coefficients $(c_i)_{i \in [r]}, (d_i)_{i \in [r]}$.[5]

Next, we bound the probability that $|L_x| < m^r$. Indeed,

$$\begin{aligned}
&\Pr\left[\ |L_x| < m^r \text{ for some } x \in f^{-1}(*)\ \right] \\
&\leq \sum_{x \in f^{-1}(*)} \Pr\left[\sum_{i=1}^{r} c_i v_x^i = 0^s \text{ for some nonzero } (c_i)_{i \in [r]}\right] \\
&\leq m^t \cdot m^r \cdot 2^{-s} \ \leq\ \frac{1}{4}.
\end{aligned}$$

Overall, the probability that $(L_x)_{x \in f^{-1}(*)}$ is not scattered is less than $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. □

Note that the condition of being scattered can be checked in polynomial time. Indeed, for each $x \in f^{-1}(*)$, one can enumerate all the elements of L_x, which are at most polynomially many in the input size $m^{O(t)}$. Thus, our zero-error randomized reduction picks random linear subspaces $(L_x)_{x \in f^{-1}(*)}$ until we obtain a scattered collection of linear subspaces.

In the rest of the proof, we can thus assume that $(L_x)_{x \in f^{-1}(*)}$ is scattered. The next claim gives the reverse inequality of Claim 17.

Claim 20. $\mathsf{DNF}_{\mathsf{MOD}_m}(g) \geq \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)|$ *if* $(L_x)_{x \in f^{-1}(*)}$ *is scattered.*

[5] Note that any equation $ax = b \pmod m$ with $a \neq 0$ is satisfied with probability $\leq 1/2$ over a random choice of x.

Let $C = \bigvee_{k=1}^{K} C_k$ be a minimum $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing g. (In particular, $K = \mathsf{DNF}_{\mathsf{MOD}_m}(g) \leq \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)| \leq m^{t+1}$.) For each $x \in f^{-1}(*)$, we first extract a subcircuit $C_{l(x)}$ that covers (a large fraction of) the random linear subspace L_x. Let $l(x) \in [K]$ be one of the indices such that $|C_{l(x)}^{-1}(1) \cap (\{x\} \times L_x)|$ is maximized. That is, $C_{l(x)}$ covers the largest fraction of the affine subspace $\{x\} \times L_x$; in particular, since $\bigcup_{k \in [K]} C_k^{-1}(1) \supseteq \{x\} \times L_x$, there are at least $|L_x|/K$ $\left(= m^r/K \geq m^{r-t-1} \geq 2\right)$ points in the set $C_{l(x)}^{-1}(1) \cap (\{x\} \times L_x)$. Intuitively, the subcircuits $\{\, C_{l(x)} \mid x \in f^{-1}(*) \,\}$ are supposed to cover random linear subspaces, and the rest of the subcircuits computes f.

To make the intuition formal, we will prove the following two claims. The first asserts that, under our constraints, no affine subspace can cover a large fraction of two distinct random linear subspaces.

Claim 21. *$l\colon f^{-1}(*) \to [K]$ is injective.*

The second claim asserts that, if an affine subspace $C_{l(x')}^{-1}(1)$ covers a large fraction of $\{x'\} \times L_{x'}$, then it cannot cover a point $(x, 0^s)$ such that $f(x) = 1$.

Claim 22. *$C_{l(x')}(x, 0^s) = 0$ for any $x \in f^{-1}(1)$ and $x' \in f^{-1}(*)$.*

Assuming these two claims, it is easy to prove Claim 20.

Proof of Claim 20. For each $k \in [K]$, define an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C'_k as $C'_k(x) := C_k(x, 0^s)$ on input $x \in \mathbb{Z}_m^t$. Define a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit $C' := \bigvee_{k \in [K] \setminus \{\, l(x) \mid f(x) = * \,\}} C'_k$. By Claim 21, the number of subcircuits in C' is $K - |f^{-1}(*)|$.

We claim that C' computes f. Indeed, for any $x \in f^{-1}(1)$, we have $C(x, 0^s) = g(x, 0^s) = f(x) = 1$; hence, there is some $k \in [K]$ such that $C_k(x, 0^s) = 1$, which implies that $C'_k(x) = 1$ by the definition of C'_k. Claim 22 implies $k \notin \{\, l(x') \mid f(x') = * \,\}$; thus $C'(x) = 1$. On the other hand, for any $x \in f^{-1}(0)$, we have $C(x, 0^s) = g(x, 0^s) = f(x) = 0$; in particular, for any $k \in [K]$, $C_k(x, 0^s) = 0$. Thus $C'_k(x) = 0$ for any $k \in [K]$, which implies $C'(x) = 0$. □

It remains to prove Claims 21 and 22. We prove the latter fist.

Proof of Claim 22. Assume, by way of a contradiction, that $C_{l(x')}(x, 0^s) = 1$ for some $x \in f^{-1}(1)$ and $x' \in f^{-1}(*)$. By the definition of $l(x')$, there are at least 2 distinct points (x', a) and (x', b) in $C_{l(x')}^{-1}(1) \cap (\{x'\} \times L_{x'})$. Since $C_{l(x')}^{-1}(1)$ is an affine subspace, we have $(x', a) - (x', b) + (x, 0^s) = (x, a - b) \in C_{l(x')}^{-1}(1)$ (as in the proof of Claim 11). It follows that $C(x, a - b) = 1$. Since C computes g, we also have $g(x, a - b) = 1$, which contradicts the fact that $a - b \neq 0^s$ and the definition of g. □

Proof of Claim 21. Assume that $l(x_1) = l(x_2) =: k$ for distinct inputs $x_1, x_2 \in f^{-1}(*)$. Take any 2 distinct points (x_1, a) and (x_1, b) from $C_k^{-1}(1) \cap (\{x_1\} \times L_{x_1})$ and any point (x_2, c) from $C_k^{-1}(1) \cap (\{x_2\} \times L_{x_2})$. Since $C_k^{-1}(1)$ is an affine subspace, we have $(x_1, a) - (x_1, b) + (x_2, c) = (x_2, a - b + c) \in C_k^{-1}(1)$. We also have $(x_2, a - b + c) \in \{x_2\} \times L_{x_2}$, since $C_k^{-1}(1) \cap (\{x_2\} \times \mathbb{Z}_m^s) \subseteq g^{-1}(1) \cap (\{x_2\} \times \mathbb{Z}_m^s) = \{x_2\} \times L_{x_2}$. Therefore, $a - b + c \in L_{x_2}$. Since $c \in L_{x_2}$ and this is a linear subspace, it follows that $a - b \in L_{x_2}$. On the other hand, by the definition of a and b, we have $0^s \neq a - b \in L_{x_1}$. However, this is a contradiction because $0^s \neq a - b \in L_{x_1} \cap L_{x_2} = \{0^s\}$. □

Proof of Theorem 16. By Claims 17 and 20, we obtain $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)|$ for a scattered collection $(L_x)_{x \in f^{-1}(*)}$. Since $s = O(t \log m)$, the truth table of g is of length $m^{t+s} = m^{O(t \log m)}$, which is a polynomial in the input length for every constant $m \geq 2$. Finally, since it is possible to check whether $(L_x)_{x \in f^{-1}(*)}$ is scattered in polynomial time, the reduction is zero-error. □

On our proof strategy and the restriction to functions over boolean inputs ($m > 2$). The linear-algebraic and probabilistic techniques employed here naturally suggest to view a set of inputs for the input instance f as a subset of the algebraic structure $\mathbb{Z}_m^n$ (a vector space or module, depending on m). In order to establish a similar NP-hardness result with respect to functions on the hypercube and AND-OR-MOD$_m$ circuits, one is tempted to encode elements from the structure $\mathbb{Z}_m^n$ as binary strings, and to consider a bijection $\varphi \colon \mathbb{Z}_m^n \leftrightarrow \Gamma \subseteq \{0,1\}^*$ between vectors and binary strings. However, a binary encoding allows a bottom-layer modular gate to access individual bits of this encoding, and as a consequence, this gate might accept a set $A \subseteq \{0,1\}^*$ that does not correspond under φ to the set of solutions of a modular equation over $\mathbb{Z}_m$. When this is the case, our argument no longer works.

Another natural approach would be to restrict the input function to boolean inputs, and to directly view such inputs as elements in $\{0,1\}^n \subseteq \mathbb{Z}_m^n$. Here certain technical difficulties are transferred to our probabilistic analysis involving affine subspaces of $\mathbb{Z}_m^n$, and it is not immediately clear to us how to modify the argument in this case.

For these reasons, when $m > 2$ our techniques do not seem to be directly applicable to functions defined over boolean inputs only, and a more complicated argument might be necessary. Note however that this does not exclude the existence of different and potentially simpler reductions among these and other intermediary problems.

4 Derandomization and Pseudorandom Generators for AND∘MOD$_m$

In this section, we present a unified way of efficiently derandomizing the zero-error reductions of Section 3. The crucial idea is that certain *subconditions* of being nice or scattered can be checked by $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits over $\mathbb{Z}_m^n$; hence, a pseudorandom generator for $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits can be used to derandomize the reductions.

In order to achieve this, we show that there exists a quick pseudorandom generator with logarithmic seed length that fools any $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit (regardless of its size), a result that might be of independent interest.

Theorem 23. *For every $\epsilon = \epsilon(n) > 0$ and each $m \geq 2$, there exists a quick pseudorandom generator $G = \{G_n \colon [\Gamma_n] \to \mathbb{Z}_m^n\}_{n \in \mathbb{N}}$ that ϵ-fools any $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit over $\mathbb{Z}_m^n$, where $\Gamma_n = \mathsf{poly}(n, 1/\epsilon, m)$ is a positive integer.*

Here we say that, for $\epsilon > 0$ and an integer $m \geq 2$, a function $G_n : [\Gamma_n] \to \mathbb{Z}_m^n$ *ϵ-fools* $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits if $|\mathbb{E}_{\gamma \in_R [\Gamma_n]}[C(G_n(\gamma))] - \mathbb{E}_{v \in_R \mathbb{Z}_m^n}[C(v)]| \leq \epsilon$ for every $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C; such a function G_n is called an *ϵ-pseudorandom generator* for $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits. We say that a family $\{G_n\}_{n \in \mathbb{N}}$ of pseudorandom generators is *quick* if G_n can be computed in $\mathsf{poly}(\Gamma_n)$ time. (Recall that $[\Gamma_n]$ denotes the set $\{1, \ldots, \Gamma_n\}$, which means that the seed-length of G_n is logarithmic in n, m, and $1/\varepsilon$ when its input elements are represented as binary strings.)

4.1 Derandomizing the Reductions

We defer a proof of Theorem 23 to the next subsection, and present its applications first: The pseudorandom generator implies polynomial-time derandomizations of the reductions presented in Section 3.

Theorem 24 (Restatement of Theorem 1)**.** $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}$ *is* NP*-hard under polynomial-time many-one reductions.*

Our basic strategy is as follows: Each reduction of Section 3 employs random variables that take value on $\mathbb{Z}_m^k$, for different choices of k. To derandomize the reductions, we simply replace these random variables by the output of the pseudorandom generator of Theorem 23; then we try all possible Γ_n seeds of G_n, and check whether the generated random variables satisfy the desired condition (which can be done in polynomial time). Below we give details for each reduction, starting with the second.

Derandomizing the second reduction. We start with the reduction from $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ to $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}$. The reduction required a scattered collection of linear subspaces, which is provided by the probabilistic argument of Claim 19. Here we present a deterministic construction of such a collection.

Theorem 25. *For any integer $m \geq 2$, there exists a deterministic algorithm that, on inputs t and r, outputs a* scattered *collection of r-dimensional linear subspaces $(L_h)_{h \in [H]}$ for $H := m^t$. Specifically,*

1. *L_h is a linear subspace of $\mathbb{Z}_m^s$ for $s := \lceil (2r + 2t) \log m + 2 \rceil$,*
2. *$|L_h| = m^r$, and*
3. *$L_h \cap L_{h'} = \{0^s\}$ for any distinct $h, h' \in [H]$.*

The running time of the algorithm is $m^{O((r+t) \log m)}$.

In the proof of Theorem 16, we picked random vectors $v_x^1, \ldots, v_x^r \in_R \mathbb{Z}_m^s$ and defined $L_x := \mathsf{span}(v_x^1, \ldots, v_x^r)$ for each $x \in f^{-1}(*) \subseteq \mathbb{Z}_m^t$. We take a similar approach, but instead of generating vectors uniformly at random, we use the output of the pseudorandom generator as the source of randomness. Specifically, let $\gamma \in [\Gamma_{rsH}]$ be a seed of the pseudorandom generator of G_{rsH}; define vectors $(v_h^1, \ldots, v_h^r)_{h \in [H]} := G_{rsH}(\gamma) \in (\mathbb{Z}_m^{rs})^H$; then, define $L_h := \mathsf{span}(v_h^1, \ldots, v_h^r)$ for each $h \in [H]$. We show that the probabilistic argument of Claim 19 still works even if the randomness is replaced in this way:

Claim 26. *Let G_{rsH} be the pseudorandom generator of Theorem 23 with error parameter $\epsilon = 2^{-s}$. Pick a seed $\gamma \in_R [\Gamma_{rsH}]$ uniformly at random, and define a collection $(L_h)_{h \in [H]}$ of linear subspaces as above. Then, $(L_h)_{h \in [H]}$ is scattered with nonzero probability.*

Proof. Note that union bounds hold for any distribution; hence, by using the union bounds as in Claim 19, the probability that $(L_h)_{h \in [H]}$ is not pairwise disjoint is

$$\begin{aligned} &\Pr\left[\ L_h \cap L_{h'} \neq \{0^s\} \text{ for some distinct } h, h' \in [H]\ \right] \\ &\leq \sum_{h \neq h' \in [H]} \sum_{(c_i),(d_i)} \Pr\left[\sum_{i=1}^{r} c_i v_h^i = \sum_{i=1}^{r} d_i v_{h'}^i \right], \end{aligned} \tag{2}$$

where the second sum is taken over all nonzero coefficient vectors $(c_i)_{i\in[r]}$ and $(d_i)_{i\in[r]}$ with entries $c_i, d_i \in \mathbb{Z}_m$. If the random vectors $(v_h^i)_{h,i}$ were uniformly distributed, the probability in (2) could be bounded by 2^{-s} as in Claim 19; Here the probability is taken over a random seed $\gamma \in_R [\Gamma_{rsH}]$ of the pseudorandom generator G_{rsH}. The condition that $\sum_{i=1}^r c_i v_h^i = \sum_{i=1}^r d_i v_{h'}^i$ can be checked by some $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit that takes $(v_h^i)_{h,i}$ as input; thus the circuit is ϵ-fooled by the pseudorandom generator; as a consequence, the probability (2) is strictly less than $m^{2t} \cdot m^{2r} \cdot (2^{-s} + \epsilon) \le \frac{1}{2}$.

Similarly,

$$\begin{aligned}
&\Pr\left[\ |L_h| < m^r \text{ for some } h \in [H]\ \right] \\
&\le \sum_{h\in[H]} \sum_{(c_i)} \cdot \Pr\left[\sum_{i=1}^r c_i v_h^i = 0^s\right] \\
&< m^t \cdot m^r \cdot (2^{-s} + \epsilon) \ \le\ \frac{1}{2}.
\end{aligned}$$

Overall, the probability that $(L_h)_{h\in[H]}$ is not scattered is strictly less than $\frac{1}{2} + \frac{1}{2} = 1$. □

Proof of Theorem 25. By Claim 26, there exists some seed $\gamma \in [\Gamma_{rsH}]$ such that the output $G_{rsH}(\gamma)$ defines a scattered collection $(L_h)_{h\in[H]}$ of linear subspaces. By exhaustively searching all the seeds, one can enumerate all the outputs of G_{rsH} in time $\mathsf{poly}(\Gamma_{rsH}) = \mathsf{poly}(rsH, 2^s, m)$. Moreover, one can check whether $G_{rsH}(\gamma)$ defines a scattered collection for each $\gamma \in [\Gamma_{rsH}]$ in time $\mathsf{poly}(H, m^s)$. Overall, the running time of our construction is $\mathsf{poly}(m^s) = m^{O((r+t)\log m)}$. □

The randomized reduction of Theorem 16 can be now derandomized, using the deterministic construction of Theorem 25 for $r := t + 2$.

Corollary 27. *There is a polynomial-time ($m^{O(t\log m)}$ time on input length $O(m^t)$) many-one reduction from* $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ *to* $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}$.

Derandomizing the first reduction. We now consider the reduction from the r-bounded set cover problem to $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$. Let $[n]$ be the universe, and $\mathcal{S} \subseteq \binom{[n]}{\le r}$ be an input to the set cover problem. Derandomizing the reduction amounts to a deterministic construction of a *nice* collection $(v^i)_{i\in[n]}$ of vectors. We generate the random vectors using the pseudorandom generator for $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits, and show that the probabilistic argument of Claim 14 still works.

Claim 28 (Revised Claim 14)**.** *Let G_{tn} be the pseudorandom generator of Theorem 23 with error parameter $\epsilon < m^{-t}$. Pick a seed $\gamma \in_R [\Gamma_{tn}]$ uniformly at random. Define $(v^1, \ldots, v^n) := G_{tn}(\gamma) \in (\mathbb{Z}_m^t)^n$. If $t \ge r + ((r+2)\log n + \log|\mathcal{S}| + 1)/\log m$, then $(v^i)_{i\in[n]}$ is nice with nonzero probability.*

Proof. By using union bounds as in Claim 14, it is sufficient to prove

$$n^{r+2} \cdot |\mathcal{S}| \cdot m^r \cdot \Pr\left[v^{j_S} = \sum_{i\in S} d_i v^i - \sum_{i\in I\setminus\{j_S\}} c_i v^i\right] < 1 \tag{3}$$

for coefficients $(c_i)_{i\in I}, (d_i)_{i\in S}$ and $j_S \in I \setminus S$, where the probability is taken over a random seed γ.

The condition $v^{j_S} = \sum_{i \in S} d_i v^i - \sum_{i \in I \setminus \{j_S\}} c_i v^i$ can be checked by an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit that takes $(v^1, \ldots, v^n) \in \mathbb{Z}_m^{tn}$ as input. By Theorem 23, we get

$$\Pr\left[v^{j_S} = \sum_{i \in S} d_i v^i - \sum_{i \in I \setminus \{j_S\}} c_i v^i\right] \leq m^{-t} + \epsilon.$$

Consequently, due to our choice of t and using $\epsilon < m^{-t}$, the left-hand side of (3) is strictly less than

$$n^{r+2} \cdot |\mathcal{S}| \cdot m^r \cdot 2m^{-t} \leq 1,$$

which completes the proof. □

In particular, there exists some seed $\gamma \in [\Gamma_{tn}]$ such that $(v^1, \ldots, v^n) = G_{tn}(\gamma)$ is nice. The number of seeds is at most $\Gamma_{tn} = \mathsf{poly}(tn, 1/\epsilon, m) = \mathsf{poly}(n, m^t) = (nm)^{O(r)}$, which is a polynomial in the input length; hence, in polynomial time, one can try all possible seeds and find a nice collection $(v^i)_{i \in [n]}$ of vectors. Thus the reduction of Theorem 8 can be derandomized:

Corollary 29. $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-$\mathsf{MCSP}^*$ *is* NP*-hard under polynomial-time many-one reductions.*

Proof of Theorem 24. Immediate from Corollaries 29 and 27. □

4.2 Near-Optimal Pseudorandom Generators for $\mathsf{AND} \circ \mathsf{MOD}_m$

This subsection contains a proof of Theorem 23. We assume basic familiarity with concepts from analysis of boolean functions [O'D14]. For simplicity, we first focus on the case of $m = 2$, which admits a simpler proof.

Proof for $m = 2$. An *ϵ-biased generator*, introduced by Naor and Naor [NN93b], is a pseudorandom generator for XOR functions. That is, we say that a function $G\colon \{0,1\}^s \to \{0,1\}^n$ is an ϵ-biased generator if $|\mathbb{E}_{x \in_R \{0,1\}^n}[\chi_S(x)] - \mathbb{E}_{s \in_R \{0,1\}^s}[\chi_S(G(s))]| \leq \epsilon$ for any $S \subseteq [n]$, where $\chi_S(x) := \bigoplus_{i \in S} x_i$. While this definition only requires the generator to fool XOR functions, it can be shown that any Boolean function with small ℓ_1 Fourier norm can be fooled by ϵ-biased generators.

Lemma 30 (see e.g., [DETT09, Lemma 2.5])**.** *Every function $f\colon \{0,1\}^n \to \{0,1\}$ can be $\epsilon \hat{\|} f \hat{\|}_1$ fooled by any ϵ-biased generator. Here, $\hat{\|} f \hat{\|}_1 := \sum_{S \subseteq [n]} |\widehat{f}(S)|$.*

Proof Sketch. Use the Fourier expansion $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$, and apply the triangle inequality. □

Moreover, it is known that any $\mathsf{AND} \circ \mathsf{XOR}$ circuit f has $\hat{\|} f \hat{\|}_1 = 1$.

Lemma 31 (see e.g., [O'D14, Proposition 3.12])**.** $\hat{\|} f \hat{\|}_1 = 1$ *for any Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ computable by a nontrivial* $\mathsf{AND} \circ \mathsf{XOR}$ *circuit.*

Proof Sketch. Let $H + a \subseteq \{0,1\}^n$ be the (nonempty) affine subspace accepted by f. Take a basis of $H^\perp$. Write a characteristic function of f using the basis, and expand it to obtain a Fourier expansion of f. □

Combining these two lemmas, any ϵ-biased generator fools $\mathsf{AND} \circ \mathsf{XOR}$ circuits. Moreover, Naor and Naor [NN93b] gave an explicit construction of an ϵ-biased generator of seed length $O(\log n + \log(1/\epsilon))$, from which Theorem 23 follows when $m = 2$.

In the proof sketched above, we exploited the fact that $\{0,1\}^n = \mathbb{Z}_2^n$ is a vector space: We took a basis of a linear subspace in the proof of Lemma 31. In order to generalize the result to the case of $m \geq 2$, we need a more direct proof which does not rely on a basis.

Proof for any $m \geq 2$. Azar, Motwani and Naor [AMN98] generalized the notion of ϵ-biased generator on $\{0,1\}^n$ to $\mathbb{Z}_m^n$ for any integer $m \geq 2$, and gave an explicit construction. We review the generalized notion and their result below.

Definition 32 ([AMN98])**.** *For a probability distribution $\mathcal{D}$ over $\mathbb{Z}_m^n$ and a vector $a \in \mathbb{Z}_m^n$, $\mathsf{bias}_{\mathcal{D}}(a)$ is defined as follows: for $g := \gcd(a_1, \ldots, a_n, m)$,*

$$\mathsf{bias}_{\mathcal{D}}(a) := \frac{1}{g} \max_{0 \leq k < m/g} \left| \Pr_{x \sim \mathcal{D}} [\langle a, x \rangle = kg] - \frac{g}{m} \right|.$$

*We say that a distribution $\mathcal{D}$ is ϵ-*biased *if $\mathsf{bias}_{\mathcal{D}}(a) \leq \epsilon$ for every $a \in \mathbb{Z}_m^n$. We say that a function $G \colon [\Gamma] \to \mathbb{Z}_m^n$ is an ϵ-*biased generator *if the distribution $G(\gamma)$ for a random seed $\gamma \in_R [\Gamma]$ is ϵ-biased.*

Theorem 33 ([AMN98, Theorem 6.1])**.** *For $m(n) \geq 2$ and $\epsilon = \epsilon(n) > 0$, there exists a quick ϵ-biased generator $G = \{G_n : [\Gamma_n] \to \mathbb{Z}_m^n\}_{n \in \mathbb{N}}$ for some $\Gamma_n = \mathsf{poly}(n, 1/\epsilon, m)$.*

We use the same pseudorandom generator G as in Theorem 33. In what follows, we will show that any ϵ-biased generator $m\epsilon$-fools $\mathsf{AND} \circ \mathsf{MOD}_m$ circuits, which completes the proof of Theorem 23.

Define $e_m \colon \mathbb{Z}_m \to \mathbb{C}^\times$ as $e_m(k) := \exp(2\pi\sqrt{-1} \cdot k/m)$ for $k \in \mathbb{Z}_m$.

Lemma 34. *For any distribution $\mathcal{D}$ on $\mathbb{Z}_m^n$ and any nonzero vector $a \in \mathbb{Z}_m^n$, we have*

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} [e_m(\langle a, x \rangle)] \right| \leq m \cdot \mathsf{bias}_{\mathcal{D}}(a).$$

Proof. The proof follows the same approach of [AMN98, Lemma 4.4]. Let $g := \gcd(a_1, \ldots, a_n, m)$.

$$\begin{aligned}
\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} [e_m(\langle a, x \rangle)] \right| &= \left| \sum_{0 \leq k < m/g} e_m(kg) \Pr_{x \sim \mathcal{D}} [\langle a, x \rangle = kg] \right| \\
&= \left| \sum_{0 \leq k < m/g} e_m(kg) \left(\Pr_{x \sim \mathcal{D}} [\langle a, x \rangle = kg] - \frac{g}{m} \right) \right| \\
&\leq \sum_{0 \leq k < m/g} |e_m(kg)| \cdot \left| \Pr_{x \sim \mathcal{D}} [\langle a, x \rangle = kg] - \frac{g}{m} \right| \\
&\leq \frac{m}{g} \cdot 1 \cdot g \cdot \mathsf{bias}_{\mathcal{D}}(a) \;=\; m \cdot \mathsf{bias}_{\mathcal{D}}(a),
\end{aligned}$$

where the first equality follows from the fact that $\langle a, x \rangle$ is a multiple of g for any $x \in \mathbb{Z}_m^n$, and in the second equality we used that $\sum_{0 \leq k < m/g} e_m(kg) = 0$ for $g < m$, which is true if $a \neq 0^n$. $\square$

As a consequence of the previous lemma, we can prove that any affine function can be "fooled":

Lemma 35. *For any ϵ-biased probability distribution $\mathcal{D}$ on $\mathbb{Z}_m^n$, any vector $a \in \mathbb{Z}_m^n$, and any scalar $b \in \mathbb{Z}_m$,*

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[e_m \left(\langle a, x \rangle + b \right) \right] - \mathop{\mathbb{E}}_{x \in_R \mathbb{Z}_m^n} \left[e_m \left(\langle a, x \rangle + b \right) \right] \right| \leq m\epsilon.$$

Proof. When $a = 0^n$, both expectations are constant, and hence the lemma follows. Otherwise, we have $\mathbb{E}_{x \in_R \mathbb{Z}_m^n} \left[e_m \left(\langle a, x \rangle \right) \right] = 0$, since this expression can be written as a product of expectations, and one of them evaluates to zero. Using Lemma 34, we obtain

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[e_m \left(\langle a, x \rangle + b \right) \right] - \mathop{\mathbb{E}}_{x \in_R \mathbb{Z}_m^n} \left[e_m \left(\langle a, x \rangle + b \right) \right] \right| &= |e_m(b)| \cdot \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[e_m \left(\langle a, x \rangle \right) \right] - \mathop{\mathbb{E}}_{x \in_R \mathbb{Z}_m^n} \left[e_m \left(\langle a, x \rangle \right) \right] \right| \\ &= 1 \cdot \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[e_m \left(\langle a, x \rangle \right) \right] \right| \\ &\leq m \, \mathsf{bias}_{\mathcal{D}}(a) \leq m\epsilon. \end{aligned}$$

□

Theorem 36. *For any ϵ-biased probability distribution $\mathcal{D}$ on $\mathbb{Z}_m^n$ and any function $f : \mathbb{Z}_m^n \to \{0, 1\}$ computable by some $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit,*

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[f(x) \right] - \mathop{\mathbb{E}}_{x \in_R \mathbb{Z}_m^n} \left[f(x) \right] \right| \leq m\epsilon$$

Proof. Suppose that an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit computing f has K MOD_m gates, and, for each $k \in [K]$, let $g_k \colon \mathbb{Z}_m^n \to \mathbb{Z}_m$ denote the *affine function* that corresponds to the kth MOD_m gate. That is, $g_k(x) = \langle a_k, x \rangle + b_k$ for some vector $a_k \in \mathbb{Z}_m^n$ and some scalar $b_k \in \mathbb{Z}_m$; moreover, for any input $x \in \mathbb{Z}_m^n$, $f(x) = 1$ if and only if $g_k(x) = 0$ for all $k \in [K]$.

We employ the following construction. Let $p(z)$ be the polynomial over $\mathbb{C}$ defined as follows.

$$p(z) := \frac{1}{m} \prod_{\alpha \in \mathbb{Z}_m \setminus \{0\}} (z - e_m(\alpha)) \tag{4}$$

$$= \frac{1}{m} \frac{z^m - 1}{z - 1} = \frac{1}{m} \sum_{i=0}^{m-1} z^i, \tag{5}$$

where the second equality holds because the roots of the polynomial $z^m - 1$ are $\{ e_m(\alpha) \mid \alpha \in \mathbb{Z}_m \}$. Useful properties of this polynomial are that, by (4), we have $p(e_m(\alpha)) = 0$ for any $\alpha \in \mathbb{Z}_m \setminus \{0\}$,

and that $p(e_m(0)) = p(1) = 1$ because of (5). Using the polynomial, we can write f as follows:

$$\begin{aligned} f(x) &= \bigwedge_{k\in[K]} [g_k(x) = 0] \\ &= \bigwedge_{k\in[K]} [p(e_m(g_k(x))) = 1] \\ &= \prod_{k\in[K]} p(e_m(g_k(x))) \\ &= \prod_{k\in[K]} \left(\frac{1}{m} \sum_{j=0}^{m-1} e_m(j \cdot g_k(x)) \right) \\ &= \frac{1}{m^K} \prod_{k\in[K]} \sum_{\alpha_k\in\mathbb{Z}_m} e_m(\alpha_k g_k(x)) \\ &= \frac{1}{m^K} \sum_{\alpha\in\mathbb{Z}_m^K} e_m \left(\sum_{k\in[K]} \alpha_k g_k(x) \right). \end{aligned}$$

Now, by using Lemma 35, we obtain

$$\begin{aligned} &\left| \mathop{\mathbb{E}}_{x\sim\mathcal{D}} [f(x)] - \mathop{\mathbb{E}}_{x\in_R\mathbb{Z}_m^n} [f(x)] \right| \\ &\leq \frac{1}{m^K} \sum_{\alpha\in\mathbb{Z}_m^K} \left| \mathop{\mathbb{E}}_{x\sim\mathcal{D}} \left[e_m \left(\sum_{k\in[K]} \alpha_k g_k(x) \right) \right] - \mathop{\mathbb{E}}_{x\in_R\mathbb{Z}_m^n} \left[e_m \left(\sum_{k\in[K]} \alpha_k g_k(x) \right) \right] \right| \\ &\leq m\epsilon, \end{aligned}$$

where in the last inequality we used the fact that $\sum_{k\in[K]} \alpha_k g_k(x)$ is an affine function. □

Proof of Theorem 23. The result is immediate from Theorems 33 and 36. □

References

[ABG+14] Adi Akavia, Andrej Bogdanov, Siyao Guo, Akshay Kamath, and Alon Rosen. Candidate weak pseudorandom functions in $AC^0[2]$. In *Innovations in Theoretical Computer Science* (ITCS), pages 251–260, 2014.

[ABK+06] Eric Allender, Harry Buhrman, Michal Koucký, Dieter van Melkebeek, and Detlef Ronneburger. Power from random strings. *SIAM J. Comput.*, 35(6):1467–1493, 2006.

[AD14] Eric Allender and Bireswar Das. Zero knowledge and circuit minimization. In *Symposium on Mathematical Foundations of Computer Science* (MFCS), pages 25–32, 2014.

[AGM15] Eric Allender, Joshua A. Grochow, and Cristopher Moore. Graph isomorphism and circuit size. *CoRR*, abs/1511.08189, 2015.

[AH17] Eric Allender and Shuichi Hirahara. New insights on the (non-)hardness of circuit minimization and related problems. In *International Symposium on Mathematical Foundations of Computer Science* (MFCS), pages 54:1–54:14, 2017.

[AHK15] Eric Allender, Dhiraj Holden, and Valentine Kabanets. The minimum oracle circuit size problem. In *International Symposium on Theoretical Aspects of Computer Science* (STACS), pages 21–33, 2015.

[AHM+08] Eric Allender, Lisa Hellerstein, Paul McCabe, Toniann Pitassi, and Michael E. Saks. Minimizing disjunctive normal form formulas and AC0 circuits given a truth table. *SIAM J. Comput.*, 38(1):63–84, 2008.

[AKRR11] Eric Allender, Michal Koucký, Detlef Ronneburger, and Sambuddha Roy. The pervasive reach of resource-bounded Kolmogorov complexity in computational complexity theory. *J. Comput. Syst. Sci.*, 77(1):14–40, 2011.

[AMN98] Yossi Azar, Rajeev Motwani, and Joseph Naor. Approximating probability distributions using small sample spaces. *Combinatorica*, 18(2):151–171, 1998.

[BR17] Andrej Bogdanov and Alon Rosen. Pseudorandom functions: Three decades later. In *Tutorials on the Foundations of Cryptography*, pages 79–158. 2017.

[CIKK16] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Learning algorithms from natural proofs. In *Conference on Computational Complexity* (CCC), pages 10:1–10:24, 2016.

[CS16] Gil Cohen and Igor Shinkar. The complexity of DNF of parities. In *Innovations in Theoretical Computer Science* (ITCS), pages 47–58, 2016.

[Czo99] Sebastian Czort. The complexity of minimizing disjunctive normal form formulas. Master's Thesis, University of Aarhus, 1999.

[DETT09] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. *Electronic Colloquium on Computational Complexity* (ECCC), 16:141, 2009.

[Fei98] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998.

[Fel09] Vitaly Feldman. Hardness of approximate two-level logic minimization and PAC learning with membership queries. *J. Comput. Syst. Sci.*, 75(1):13–26, 2009.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[GKM15] Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *Symposium on Foundations of Computer Science* (FOCS), pages 903–922, 2015.

[God] Chris Godsil. Double orthogonal complement of a finite module. MathOverflow (Retrieved 19-01-2018).

[Gro98] Vince Grolmusz. A lower bound for depth-3 circuits with MOD_m gates. *Inf. Process. Lett.*, 67(2):87–90, 1998.

[HP15] John M. Hitchcock and Aduri Pavan. On the NP-completeness of the minimum circuit size problem. In *Conference on Foundation of Software Technology and Theoretical Computer Science* (FSTTCS), pages 236–245, 2015.

[HW16] Shuichi Hirahara and Osamu Watanabe. Limits of minimum circuit size problem as oracle. In *Conference on Computational Complexity* (CCC), pages 18:1–18:20, 2016.

[Juk06] Stasys Jukna. On graph complexity. *Combinatorics, Probability & Computing*, 15(6):855–876, 2006.

[Juk12] Stasys Jukna. *Boolean Function Complexity - Advances and Frontiers*. Springer, 2012.

[Kar72] Richard M. Karp. Reducibility among combinatorial problems. In *Symposium on the Complexity of Computer Computations*, pages 85–103, 1972.

[KC00] Valentine Kabanets and Jin-yi Cai. Circuit minimization problem. In *Symposium on Theory of Computing* (STOC), pages 73–79, 2000.

[Kra11] Jan Krajícek. *Forcing with Random Variables and Proof Complexity*. Cambridge University Press, 2011.

[KS08] Subhash Khot and Rishi Saket. Hardness of minimizing and learning DNF expressions. In *Symposium on Foundations of Computer Science* (FOCS), pages 231–240, 2008.

[Mas79] William J. Masek. Some NP-complete set covering problems. Unpublished Manuscript, 1979.

[MW15] Cody Murray and Ryan Williams. On the (non) NP-hardness of computing circuit complexity. In *Conference on Computational Complexity* (CCC), pages 365–380, 2015.

[NN93a] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[NN93b] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[OS17] Igor Carboni Oliveira and Rahul Santhanam. Conspiracies between learning algorithms, circuit lower bounds, and pseudorandomness. In *Computational Complexity Conference* (CCC), pages 18:1–18:49, 2017.

[PSZ00] Ramamohan Paturi, Michael E. Saks, and Francis Zane. Exponential lower bounds for depth three boolean circuits. *Computational Complexity*, 9(1):1–15, 2000.

[PV88] Leonard Pitt and Leslie G. Valiant. Computational limitations on learning from examples. *J. ACM*, 35(4):965–984, 1988.

[RR97] Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[Sla96] Petr Slavík. A tight analysis of the greedy algorithm for set cover. In *Symposium on the Theory of Computing* (STOC), pages 435–441, 1996.

[Tra84] Boris A. Trakhtenbrot. A survey of Russian approaches to perebor (brute-force searches) algorithms. *IEEE Annals of the History of Computing*, 6(4):384–400, 1984.

[Tre01] Luca Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Symposium on Theory of Computing* (STOC), pages 453–461, 2001.

A Proof of Fact 3 – Double Orthogonal Complement in $(\mathbb{Z}/m\mathbb{Z})^n$

In this section we present the proof of Fact 3, which for convenience is reformulated as Theorem 37 stated below. Our presentation follows the proof outlined in [God].

Recall the following concepts. We consider the Abelian group $G := (\mathbb{Z}/m\mathbb{Z})^n$ equipped with component-wise addition modulo m, and let $\langle x, y\rangle := \sum_{i\in[n]} x_i y_i \mod m$, where $x, y \in G$. For a subgroup V of G, define $V^\perp := \{\, x \in G \mid \langle x, y\rangle = 0 \text{ for all } y \in V \,\}$, which is again a subgroup of G.

Theorem 37 (folklore)**.** *$V^{\perp\perp} = V$ for any subgroup V of $G = (\mathbb{Z}/m\mathbb{Z})^n$.*

It is easy to see $V \subseteq V^{\perp\perp}$: indeed, for any $x \in V$, we have $\langle x, y\rangle = 0$ for each $y \in V^\perp$ by the definition of $V^\perp$; hence $x \in V^{\perp\perp}$. Therefore, it is sufficient to show that the size of $V^{\perp\perp}$ is equal to that of V. To this end, we prove the following claim.

Claim 38. *$|V^\perp| = |G|/|V|$ for any subgroup V of G.*

Note that, applying this claim twice, we obtain $|V^{\perp\perp}| = |G|/|V^\perp| = |G|/(|G|/|V|) = |V|$, which completes the proof of Theorem 37. Claim 38 will be proved by combining the three claims below.

Let H be any finite Abelian group. A *character* of the group H is a homomorphism $\chi\colon H \to \mathbb{C}^\times$ Let $\widehat{H}$ denote the dual group of H, that is, the group of all characters of H. (See e.g. [O'D14, Section 8.5] for more details.) It is known that the order of a group H and the order of its dual group $\widehat{H}$ are the same.

Claim 39 ([O'D14, Corollary of Proposition 8.55 and Exercise 8.35])**.** *$|H| = |\widehat{H}|$ for any finite Abelian group H.*

For any subgroup V of G, define $V^* := \{\, \chi \in \widehat{G} \mid \chi(v) = 1 \text{ for every } v \in V \,\}$.

Claim 40. *$\widehat{G/V} \cong V^*$ for any subgroup V of G.*

Proof. We define an isomorphism $\varphi\colon \widehat{G/V} \to V^*$. Given $\chi \in \widehat{G/V}$, we define $\varphi(\chi)\colon G \to \mathbb{C}^\times$ by $\varphi(\chi)(x) := \chi(x + V)$ for $x \in G$. We claim that $\varphi(\chi)$ is indeed in V^*: First, $\varphi(\chi)\colon G \to \mathbb{C}^\times$ is a homomorphism since $\varphi(\chi)(x + y) = \chi(x + y + V) = \chi((x + V) + (y + V)) = \chi(x + V)\chi(y + V)$ for any $x, y \in G$. Second, $\varphi(\chi)(v) = \chi(v + V) = \chi(V) = 1$ for any $v \in V$. (Here, we used the fact that the homomorphism χ maps the identity $0 + V \in G/V$ to the identity $1 \in \mathbb{C}^\times$.)

We claim that φ is a homomorphism. Indeed, $\varphi(\chi_1\chi_2)(x) = (\chi_1\chi_2)(x+V) = \chi_1(x+V)\chi_2(x+V) = \varphi(\chi_1)(x)\varphi(\chi_2)(x)$ for any $x \in G$ and any $\chi_1, \chi_2 \in \widehat{G/V}$; hence $\varphi(\chi_1\chi_2) = \varphi(\chi_1)\varphi(\chi_2)$.

In order to prove that φ is a bijection, we construct an inverse map $\psi\colon V^* \to \widehat{G/V}$. Given $\chi \in V^*$, define $\psi(\chi)(a+V) := \chi(a)$ for any coset $a+V \in G/V$. Note that this map is well defined since $a+V = b+V$ implies $a-b \in V$, and thus $1 = \chi(a-b) = \chi(a)/\chi(b)$. It is straightforward to see that $\psi = \varphi^{-1}$: indeed, $\psi(\varphi(\chi))(a+V) = \varphi(\chi)(a) = \chi(a+V)$ and $\varphi(\psi(\chi))(a) = \psi(\chi)(a+V) = \chi(a)$ for any $a \in G$. Hence φ is both injective and surjective, and consequently, an isomorphism. □

Claim 41. *$V^* \cong V^\perp$ for any subgroup V of $G = (\mathbb{Z}/m\mathbb{Z})^n$.*

Proof. We first prepare some notation: For any $i \in [n]$, let $e_i \in G$ be the vector whose value is 1 on the ith coordinate and is 0 on the other coordinates. Let $\omega := \exp(2\pi\sqrt{-1}/m) \in \mathbb{C}^\times$ denote the mth root of unity.

We construct an isomorphism $\varphi\colon V^\perp \to V^*$. Given $x \in V^\perp$, define $\varphi(x) \in V^*$ as $\varphi(x)(y) := \omega^{\langle x,y\rangle}$ for any $y \in G$. Note that the image of φ is contained in V^*: indeed, for any $v \in V^\perp$, we have $\varphi(x)(v) = \omega^{\langle x,v\rangle} = \omega^0 = 1$.

We claim that φ is injective. It is easy to see that φ is a homomorphism; thus, it is sufficient to prove that the kernel of φ is just $0 \in V^\perp$. If $\varphi(x)$ is the constant function 1, then $\langle x, y\rangle = 0$ for any $y \in G$; in particular, letting $y \in \{e_1, \dots, e_n\}$, we obtain $x = 0$.

Finally, we claim that φ is surjective. For any $\chi \in V^*$ and any $i \in [n]$, there is some $x_i \in \mathbb{Z}/m\mathbb{Z}$ such that $\chi(e_i) = \omega^{x_i}$: indeed, since $1 = \chi(0) = \chi(m \cdot e_i) = \chi(e_i)^m$, $\chi(e_i)$ is one of the mth roots of unity. Now we define $x := \sum_{i=1}^n x_i e_i \in G$. Then, for any $y \in G$, $\varphi(x)(y) = \omega^{\langle x,y\rangle} = \prod_{i=1}^n \omega^{x_i y_i} = \prod_{i=1}^n \chi(e_i)^{y_i} = \prod_{i=1}^n \chi(y_i e_i) = \chi(\sum_{i=1}^n y_i e_i) = \chi(y)$; hence $\varphi(x) = \chi$ for some $x \in G$. Moreover, for any $v \in V$, we have $\chi(v) = \omega^{\langle x,v\rangle} = 1$ since $\chi \in V^*$; thus we have $\langle x, v\rangle = 0$, which implies that $x \in V^\perp$. □

Combining these three claims, we obtain $|V^\perp| = |V^*| = |\widehat{G/V}| = |G/V| = |G|/|V|$, which completes the proof of Claim 38.

B On Different Complexity Measures for $\mathsf{DNF} \circ \mathsf{MOD}_p$ Circuits

In this section, we provide an example of the robustness of our arguments with respect to variations of the complexity measure. Let $p \geq 2$ be a fixed prime. We sketch the proof of a hardness result for a variant of the $(\mathsf{DNF} \circ \mathsf{MOD}_p)$-$\mathsf{MCSP}^*$ problem, described as follows. We consider layered $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{MOD}_p$ formulas[6] over $\mathbb{Z}_p^n$, and measure complexity by the total number of (non-input) gates in the formula.[7] A bit more precisely, we adapt the proof of Theorem 8 from Section 3.1, and show that this problem is also NP-hard under randomized reductions.

Since $\mathbb{Z}_p^t$ is a vector space over the field $\mathbb{Z}_p$, we can define the dimension of an affine subspace: For a linear subspace $H \subseteq \mathbb{Z}_p^t$, let $\dim(H)$ denote the dimension of H, and let $\operatorname{codim}(H) := \dim(H^\perp) = t - \dim(H)$; then, for any $a \in \mathbb{Z}_p^t$, define the dimension of an affine subspace $H + a$ as $\dim(H + a) := \dim(H)$, and $\operatorname{codim}(H + a) := \dim(H)$. Observe that this notion is well-defined. Using dimension, we can characterize the number of gates in $\mathsf{AND} \circ \mathsf{MOD}_p$ formulas.

[6] Recall that in a formula every non-input gate has fan-out one.

[7] Under our notion of layered formulas, an $(\mathsf{AND} \circ \mathsf{MOD}_p)$-circuit with a single MOD_p gate has size 2. While this is convenient for the exposition, it is not particularly important for the result.

Lemma 42. *Let A be an affine subspace of $\mathbb{Z}_p^t$. Then, the minimum number of gates in any layered $\mathsf{AND} \circ \mathsf{MOD}_p$ formula accepting A is exactly $1 + \operatorname{codim}(A)$.*

Proof Sketch. As in the proof of Lemma 2, a layered $\mathsf{AND} \circ \mathsf{MOD}_p$ formula C with $1+s$ gates accepts the set $A = C^{-1}(1)$ of solutions of s linear equations over MOD_p. Let $B \in \mathbb{Z}_p^{s \times t}$ be the matrix that defines these linear equations. Then, we have $\dim \ker(B) = \dim(A)$, and by the rank-nullity theorem, we obtain $\operatorname{codim}(A) = t - \dim(A) = t - \dim \ker(B) = \operatorname{rank}(B) \leq s$.

Conversely, let $A =: H + a$ for some linear subspace H and some $a \in \mathbb{Z}_p^t$, and let $\gamma_1, \ldots, \gamma_s$ be a basis of $H^\perp$, where $s := \operatorname{codim}(H)$. Then, using orthogonal complements, it is easy to check that $x \in A$ if and only if $\langle \gamma_i, x \rangle = \langle \gamma_i, a \rangle$ for all $i \in [s]$. The latter condition can be written as an $\mathsf{AND} \circ \mathsf{MOD}_p$ layered formula with $1 + s$ gates. □

As a corollary, for any *optimal* layered $(\mathsf{DNF} \circ \mathsf{MOD}_p)$-formula $C = \bigvee_{k=1}^K C_k$ for a function $f \colon \mathbb{Z}_p^n \to \{0,1\}$, where C_k is an $\mathsf{AND} \circ \mathsf{MOD}_p$ circuit for each $k \in [K]$, the total number of gates in the formula is precisely $1 + K + \sum_{k=1}^K \operatorname{codim}(C_k^{-1}(1))$.

For convenience, given a function $f \colon \mathbb{Z}_p^t \to \{0,1,*\}$, let $\operatorname{size}(f)$ denote the complexity of f according to our size measure. Now let us revise the proof of Theorem 8. Given an instance $\mathcal{S} \subseteq \binom{n}{\leq r}$ of the r-bounded set cover instance, we construct a function $f \colon \mathbb{Z}_p^t \to \{0,1,*\}$ in exactly the same way. Below we adapt the corresponding claims from Section 3.1. Then we employ the new claims to argue that the NP-hardness result still holds.

Claim 43 (Adaptation of Claim 9)**.** *Assume that $\mathcal{S}$ has a set cover of size K. Then* $\operatorname{size}(f) \leq (t+1)K + 1$.

Proof. Let $\mathcal{C} \subseteq \mathcal{S}$ be a set cover of size K. For each $S \in \mathcal{C}$, let C_S be an $\mathsf{AND} \circ \mathsf{MOD}_p$ circuit over $\mathbb{Z}_p^t$ that accepts $\mathsf{span}(v^S)$. Define a $\mathsf{DNF} \circ \mathsf{MOD}_p$ circuit $C := \bigvee_{S \in \mathcal{C}} C_S$. Then the circuit size of C is $1 + K + \sum_{i=1}^K \operatorname{codim}(C_S^{-1}(1))$, which is obviously at most $1 + K(t+1)$. □

Claim 44 (Adaptation of Claim 13)**.** *Let $(v^i)_{i \in [n]}$ be nice, and $s := \operatorname{size}(f)$. Then $\mathcal{S}$ has a set cover of size $2(s-1)/(t - r - (\log |\mathcal{S}| / \log p) + 1)$.*

Proof. Let $C = \bigvee_{k=1}^K C_k$ be an optimal $\mathsf{DNF} \circ \mathsf{MOD}_p$ layered formula of size s computing f. Then, as discussed above, we have $s = 1 + K + \sum_{k=1}^K \operatorname{codim}(C_k^{-1}(1))$. On the other hand, the same analysis from Claim 13 shows that $\mathcal{S}$ has a set cover of size $\leq 2K$. It thus remains to give an upper bound on K.

Since C computes f, we have $C_k^{-1}(1) \subseteq C^{-1}(1) \subseteq f^{-1}(\{1,*\}) = \bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S)$. By counting the number of elements in $C_k^{-1}(1)$ and $\bigcup_{S \in \mathcal{S}} \mathsf{span}(v^S)$, we obtain $p^{\dim(C_k^{-1}(1))} \leq |\mathcal{S}| \cdot p^r$. Hence, we have $\operatorname{codim}(C_k^{-1}(1)) \geq t - r - \log |\mathcal{S}| / \log p$; therefore,

$$s \;\geq\; 1 + K + \sum_{k=1}^K \operatorname{codim}(C_k^{-1}(1)) \;\geq\; 1 + K + K(t - r - \log |\mathcal{S}| / \log p),$$

which implies $K \leq (s-1)/(t - r - (\log |\mathcal{S}| / \log p) + 1)$. □

Let K be the minimum size of a cover for $\mathcal{S}$. By the claims above, we have $\mathrm{size}(f) \lesssim tK$ and $K \lesssim 2\mathrm{size}(f)/t$, because t can be taken large enough compared to the other relevant parameters; hence $\mathrm{size}(f)/t$ roughly gives us a 2-factor approximation. More precisely, we have $\mathrm{size}(f) \leq (t+1)K + 1 \leq 2(t+1)K$, and $K \leq 2(\mathrm{size}(f) - 1)/((t+1)/2) \leq 4\mathrm{size}(f)/(t+1)$ for any $t \geq 2r + 2\log|\mathcal{S}|/\log p - 1$. That is, the set cover size K satisfies

$$\frac{\mathrm{size}(f)}{2(t+1)} \;\leq\; K \;\leq\; \frac{4\mathrm{size}(f)}{t+1},$$

which gives an 8-factor approximation of K. Since we can take r to be a sufficiently large constant in Theorem 5, the result holds.

C A Hardness of Approximation Result for $(\mathsf{DNF} \circ \mathsf{MOD}_m)$-MCSP

The reduction from $(\mathsf{DNF}\circ\mathsf{MOD}_m)$-MCSP* to $(\mathsf{DNF}\circ\mathsf{MOD}_m)$-MCSP presented in Section 3 is not *approximation-preserving*: given a partial function $f\colon \mathbb{Z}_m^t \to \{0,1,*\}$, it produces a total function $g\colon \mathbb{Z}_m^{O(t\log m)} \to \{0,1\}$ such that $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = \mathsf{DNF}_{\mathsf{MOD}_m}(f) + |f^{-1}(*)|$. The reduction introduces an additive term $|f^{-1}(*)|$, and hence a (multiplicative) approximation of $\mathsf{DNF}_{\mathsf{MOD}_m}(g)$ does not give a good approximation of $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$. In order to fix this situation, we give an approximation-preserving reduction. Our approach is inspired by a reduction described in [AHM+08].

Theorem 45 (Approximation-preserving version of Corollary 27)**.** *There is a polynomial-time algorithm that, given the truth table of a partial function $f\colon \mathbb{Z}_m^t \to \{0,1,*\}$, produces the truth table of a total function $g\colon \mathbb{Z}_m^{2t+2s} \to \{0,1\}$ such that*

$$\mathsf{DNF}_{\mathsf{MOD}_m}(g) = |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1),$$

where $s := \lceil (6t+4)\log m + 2 \rceil$.

Proof. The idea of the proof is to amplify the circuit size for f; that is, we would like to force any circuit C computing g to also compute sub-functions corresponding to $|f^{-1}(*)|$ copies of f.

We can amplify the circuit size as follows. Let $(L_x)_{x\in f^{-1}(*)}$ be a scattered collection of linear subspaces of $\mathbb{Z}_m^s$. Define a function g' by $g'(x,z,w) := f(x)$ if $z \in f^{-1}(*)$ and $w \in L_z$; otherwise $g'(x,z,w) := 0$. Then, under an appropriate choice of parameters, it can be shown that $\mathsf{DNF}_{\mathsf{MOD}_m}(g') = |f^{-1}(*)| \cdot \mathsf{DNF}_{\mathsf{MOD}_m}(f)$. By combining an analogous reduction and the idea behind the proof of Theorem 16, we can obtain a total function g such that $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = \mathsf{DNF}_{\mathsf{MOD}_m}(g') + |f^{-1}(*)| = |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1)$.[8] Details follow.

We first obtain a scattered collection $(L_x)_{x\in f^{-1}(*)}$ of r-dimensional linear subspaces of $\mathbb{Z}_m^s$ by using Theorem 25 for $r := 2t+2$. Then we define $g\colon \mathbb{Z}_m^{2t+2s} \to \{0,1\}$ as

$$g(x,y,z,w) := \begin{cases} f(x) & (\text{if } f(x) \in \{0,1\} \text{ and } y = 0^s \text{ and } f(z) = * \text{ and } w \in L_z) \\ 1 & (\text{if } f(x) = * \text{ and } y \in L_x) \\ 0 & (\text{otherwise}) \end{cases}$$

for any $((x,y),(z,w)) \in (\mathbb{Z}_m^s \times \mathbb{Z}_m^t)^2$.

[8] A black-box application of Corollary 27 produces a function g such that $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = \mathsf{DNF}_{\mathsf{MOD}_m}(g') + |g'^{-1}(*)|$, which is not sufficient for our purpose because $|g'^{-1}(*)|$ is larger than $|f^{-1}(*)|$.

Claim 46 (Analogue of Claim 17)**.** $\mathsf{DNF}_{\mathsf{MOD}_m}(g) \le |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1)$.

Proof. Suppose that a $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit $C = \bigvee_{k=1}^{K} C_k$ computes f. For each $x^* \in f^{-1}(*)$, take an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C_{x^*} accepting $\{x^*\} \times L_{x^*}$ (by Lemma 2). Define

$$C'(x, y, z, w) := \bigvee_{z^* \in f^{-1}(*)} \bigvee_{k=1}^{K} (C_k(x) \wedge y_1 = 0 \wedge \cdots \wedge y_s = 0 \wedge C_{z^*}(z, w)) \vee \bigvee_{x^* \in f^{-1}(*)} C_{x^*}(x, y).$$

It is easy to see that C' computes g. □

The rest of the proof is devoted to the reverse direction.

Claim 47 (Analogue of Claim 20)**.** $\mathsf{DNF}_{\mathsf{MOD}_m}(g) \ge |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1)$.

Let $C = \bigvee_{k=1}^{K} C_k$ be a minimum $\mathsf{DNF} \circ \mathsf{MOD}_m$ circuit computing g. In particular, $K = \mathsf{DNF}_{\mathsf{MOD}_m}(g) \le |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1) \le m^{2t+1}$. For each $x \in f^{-1}(*)$, let $l(x) \in [K]$ be one of the indices such that $|C_{l(x)}^{-1}(1) \cap (\{x\} \times L_x \times \mathbb{Z}_m^{t+s})|$ is maximized. Since $\bigcup_{k \in [K]} C_k^{-1}(1) \supseteq \{x\} \times L_x \times \mathbb{Z}_m^{t+s}$, there are at least $|L_x| \cdot m^{t+s}/K \ge m^{r+t+s}/m^{2t+1} \ge 2$ points in the set $C_{l(x)}^{-1}(1) \cap (\{x\} \times L_x \times \mathbb{Z}_m^{t+s})$.

Define $T_0 := \{\, C_{l(x)} \mid f(x) = * \,\}$. For each $z \in f^{-1}(*)$, let T_z be the set of all C_k such that $k \in [K]$ and C_k accepts at least 2 elements from $\{(x, 0^s, z)\} \times L_z$ for some $x \in f^{-1}(1)$. We will show that the sets T_0, $\{T_z\}_{z \in f^{-1}(*)}$ are pairwise disjoint, and hence $K \ge |T_0| + \sum_{z \in f^{-1}(*)} |T_z|$. We will also prove that $|T_0| = |f^{-1}(*)|$ and $|T_z| \ge \mathsf{DNF}_{\mathsf{MOD}_m}(f)$, which completes the proof.

Claim 48. *$l \colon f^{-1}(*) \to [K]$ is injective (hence $|T_0| = |f^{-1}(*)|$).*

Claim 49. *$T_0 \cap T_z = \varnothing$ for any $z \in f^{-1}(*)$.*

Since the proofs of these claims are essentially the same as in Claims 21 and 22, respectively (except that we have extra coordinates taking values in $\mathbb{Z}_m^t \times \mathbb{Z}_m^s$), we omit them.

Claim 50. *$T_{z_1} \cap T_{z_2} = \varnothing$ for any distinct elements $z_1, z_2 \in f^{-1}(*)$.*

Proof. The proof is basically the argument from Claim 21. For completeness, we briefly repeat it here. Towards a contradiction, assume that there exists a circuit C_k in $T_{z_1} \cap T_{z_2}$. By the definition of T_{z_1} and T_{z_2}, there exist elements $x_1, x_2 \in f^{-1}(1)$, $a \ne b \in L_{z_1}$, and $c \in L_{z_2}$ such that $C_k(x_1, 0^s, z_1, a) = C_k(x_1, 0^s, z_1, b) = C_k(x_2, 0^s, z_2, c) = 1$. Since $C_k^{-1}(1)$ is an affine subspace, we have $(x_1, 0^s, z_1, a) - (x_1, 0^s, z_1, b) + (x_2, 0^s, z_2, c) = (x_2, 0^s, z_2, a - b + c) \in C_k^{-1}(1)$. Since $C_k^{-1}(1) \cap (\{(x_2, 0^s, z_2)\} \times \mathbb{Z}_m^s) \subseteq \{(x_2, 0^s, z_2)\} \times L_{z_2}$, we get $a - b + c \in L_{z_2}$. However, given that $c \in L_{z_2}$, we obtain $0^s \ne a - b \in L_{z_1} \cap L_{z_2}$, which contradicts $L_{z_1} \cap L_{z_2} = \{0^s\}$. □

Fix any $z \in f^{-1}(*)$. For each $C_k \in T_z$, define an $\mathsf{AND} \circ \mathsf{MOD}_m$ circuit C'_k so that $C'^{-1}_k(1) = \{\, x \in \mathbb{Z}_m^t \mid C_k(x, 0^s, z, w) = 1 \text{ for some } w \in \mathbb{Z}_m^s \,\}$. (Note that a projection of an affine subspace $C_k^{-1}(1)$ is again an affine subpace because a projection is a homomorphism.) Now define $C_z := \bigvee_{C_k \in T_z} C'_k$.

Claim 51. *C_z computes f for any $z \in f^{-1}(*)$. (In particular, $|T_z| \ge \mathsf{DNF}_{\mathsf{MOD}_m}(f)$.)*

Proof. Fix any $x \in f^{-1}(1)$. Since $\{(x, 0^s, z)\} \times L_z$ is covered by $\bigcup_{k\in[K]} C_k^{-1}(1)$, and $|L_z| = m^r$, $K \le m^{2t+1}$, and $r = 2t + 2$, there exists $k \in [K]$ such that there are at least 2 elements in $(\{(x, 0^s, z)\} \times L_z) \cap C_k^{-1}(1)$; hence, by the definition of T_z, we have $C_k \in T_z$. Moreover, $C'_k(x) = 1$ by the definition of C'_k; thus $C_z(x) = \bigvee_{C_k\in T_z} C'_k(x) = 1$.

Now fix any $x \in f^{-1}(0)$. Since $g(x, 0^s, z, w) = 0$ for every $w \in \mathbb{Z}_m^s$, we get $C_k(x, 0^s, z, w) = 0$ for any $C_k \in T_z$; thus $C'_k(x) = 0$, which implies that $C_z(x) = 0$. □

Combining the claims above, we obtain

$$\mathsf{DNF}_{\mathsf{MOD}_m}(g) = K \ge |T_0| + \sum_{z\in f^{-1}(*)} |T_z| \ge |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1).$$

This completes the proof of Theorem 45. □

We can then establish a hardness of approximation result for computing $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$. For a function $f\colon \mathbb{Z}_m^t \to \{0, 1\}$, define $|f| := m^t$, which is the number of entries in the truth table of a function f.

Theorem 52. *There exists a constant $c > 0$ such that if there is a quasipolynomial-time algorithm which approximates $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ to within a factor of $c \log\log|f|$, then $\mathsf{NP} \subseteq \mathsf{DTIME}(2^{(\log n)^{O(1)}})$.*

Proof. As noted by Trevisan [Tre01], by choosing the parameters of Feige's reduction [Fei98], one can obtain hardness of approximation results for the r-bounded set cover problem. While Trevisan only analyzed the case when r is constant (cf. Theorem 5), a similar analysis[9] shows that it is NP-hard (under quasipolynomial-time many-one reductions) to approximate the $r(n)$-bounded set cover problem on n points within a factor of $\gamma \log r(n)$ $(= \gamma \log\log n)$ for $r(n) := \log n$ and some small constant $\gamma > 0$.

Suppose that $\mathsf{DNF}_{\mathsf{MOD}_m}(g)$ can be approximated to within a factor of $(\gamma/6)\log\log|g|$ by an algorithm A, where $g\colon \mathbb{Z}_m^t \to \{0, 1\}$ is a total function. We show below that if A runs in quasipolynomial time, then $\mathsf{NP} \subseteq \mathsf{DTIME}(2^{(\log n)^{O(1)}})$.

First, note that in order to conclude this it is enough to describe a quasipolynomial-time algorithm B that approximates r-Bounded Set Cover to within a factor of $\gamma \log r(n)$ for $r(n) = \log n$. Let $([n], \mathcal{S})$ be an instance of the r-Bounded Set Cover Problem. Algorithm B applies the deterministic $n^{O(r(n))}$-time reduction provided by Corollary 29 to produce a partial Boolean function $f\colon \mathbb{Z}_m^{O(r(n)\log n)} \to \{0, 1, *\}$. It then invokes the deterministic reduction from Theorem 45 to construct from f a total function $g\colon \mathbb{Z}_m^{O(r(n)\log n)} \to \{0, 1\}$. Finally, B uses the approximation algorithm A to compute a $(\gamma/6)\log\log|g|$ approximation to $\mathsf{DNF}_{\mathsf{MOD}_m}(g)$. Let $\widetilde{g} \in \mathbb{N}$ be the value output by A. Algorithm B outputs $\widetilde{K} := 2\widetilde{g}/|f^{-1}(*)|$.

Note that B runs in quasipolynomial time under our assumptions. It remains to show that it approximates the solution of the original set cover problem within a factor of $\gamma \log\log n$. Let K be the cost of an optimal solution to the initial set cover instance. Recall that $2\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ is

[9] Specifically, for the parameters and notation in [Fei98], given a 3CNF-5 formula on n variables, let k be a sufficiently large constant, $m := \sqrt{\log n}$, and $\ell := c \log\log m$ for a large constant c. Then the output of Feige's reduction is an instance of the set cover problem on N $(:= m(5n)^\ell)$ points such that each set is of size at most $m2^{O(\ell)} \le r(N) = \log N$, and the gap between yes instances and no instances is $(1 - \frac{4}{k}) \ln m = \Omega(\log\log N)$.

a 2-factor approximation for K; that is, $K \leq 2 \cdot \mathsf{DNF}_{\mathsf{MOD}_m}(f) \leq 2K$. On the other hand, the guarantees of the algorithm A imply that

$$\mathsf{DNF}_{\mathsf{MOD}_m}(g) \;\leq\; \widetilde{g} \;\leq\; \mathsf{DNF}_{\mathsf{MOD}_m}(g) \cdot (\gamma/6) \log\log |g|.$$

Since $\mathsf{DNF}_{\mathsf{MOD}_m}(g) = |f^{-1}(*)| \cdot (\mathsf{DNF}_{\mathsf{MOD}_m}(f) + 1)$, we get

$$K \;\leq\; \frac{2\widetilde{g}}{|f^{-1}(*)|} \;\leq\; (\gamma/6) \log\log |g| \cdot (K+1)$$

Therefore, for large enough n and on non-trivial instances (i.e. $K \geq 1$), the value $\widetilde{K}$ output by B approximates K to within a factor of $2 \cdot (\gamma/6) \log\log |g| \leq (\gamma/3) \cdot (\log r(n) + \log\log n + O(\log m)) \leq (\gamma/3) \cdot 3 \log\log n$. □

Finally, we note that when m is prime, it is possible to design a quasipolynomial-time approximation algorithm for $\mathsf{DNF}_{\mathsf{MOD}_m}(f)$ with an approximation factor of $O(\log |f|)$.

Theorem 53. *Let p be a prime number. There is a quasipolynomial-time algorithm which approximates $\mathsf{DNF}_{\mathsf{MOD}_p}(f)$ to within a factor of $\ln |f|$.*

Proof. Let $|f| = p^t$ be the number of entries in the truth table of f, the input function. By the results of Section 2.1, computing $\mathsf{DNF}_{\mathsf{MOD}_p}(f)$ is equivalent to solving a set cover instance. Recall that set cover admits a polynomial-time approximation algorithm that achieves an approximation factor of $\ln N$ on instances over a universe of size N (cf. [Sla96]). Consequently, in order to prove the result it is enough to verify that computing $\mathsf{DNF}_{\mathsf{MOD}_p}(f)$ reduces to a set cover instance with domain size $N_f := |f^{-1}(1)| \leq |f|$ and of size at most quasipolynomial in $|f|$.

Indeed, for a non-zero function $f \colon \mathbb{Z}_p^t \to \{0,1\}$, $\mathsf{DNF}_{\mathsf{MOD}_p}(f)$ is exactly the minimum number of affine subspaces that cover $f^{-1}(1)$. Therefore, by relabelling elements, computing $\mathsf{DNF}_{\mathsf{MOD}_p}(f)$ reduces to a set cover instance $([N_f], \mathcal{S}_f)$, where a set $S \in S_f$ if and only if S viewed as a subset of $\mathbb{Z}_p^t$ is an affine subspace contained in $f^{-1}(1)$. Each such affine subspace has dimension at most t, and can be explicitly described by a basis $v_1, \ldots, v_\ell \in \mathbb{Z}_p^t$, where $\ell \leq t$, and a vector $b \in \mathbb{Z}_p^t$. Hence there are at most $p^{O(t^2)}$ such spaces, and consequently, $|\mathcal{S}_f| \leq p^{O(t^2)}$. In other words, we get a set cover instance over a ground set of size $\leq |f|$, and this instance contains at most $|f|^{O(\log |f|)}$ sets.

Finally, since the sets in $\mathcal{S}_f$ can be generated in time at most $|f|^{O(\log |f|)}$, and the set cover approximation algorithm runs in time polynomial in its input length, the result holds. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il