Violating Constant Degree Hypothesis Requires Breaking Symmetry

Piotr Kawałek
TU Wien, Austria
Jagiellonian University in Kraków, Poland

Armin Weiß
University of Stuttgart, Germany

Abstract

The Constant Degree Hypothesis was introduced by Barrington et. al. [5] to study some extensions of q-groups by nilpotent groups and the power of these groups in a computation model called NuDFA (non-uniform DFA). In its simplest formulation, it establishes exponential lower bounds for $\mathrm{MOD}_q \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ circuits computing AND of unbounded arity n (for constant integers d, m and a prime q). While it has been proved in some special cases (including $d = 1$), it remains wide open in its general form for over 30 years.

In this paper we prove that the hypothesis holds when we restrict our attention to symmetric circuits with m being a prime. While we build upon techniques by Grolmusz and Tardos [23], we have to prove a new symmetric version of their *Degree Decreasing Lemma* and use it to simplify circuits in a symmetry-preserving way. Moreover, to establish the result, we perform a careful analysis of automorphism groups of $\mathrm{MOD}_m \circ \mathrm{AND}_d$ subcircuits and study the periodic behaviour of the computed functions. Our methods also yield lower bounds when d is treated as a function of n.

Finally, we present a construction of symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ circuits that almost matches our lower bound and conclude that a symmetric function f can be computed by symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits of quasipolynomial size if and only if f has periods of polylogarithmic length of the form $p^k q^\ell$.

2012 ACM Subject Classification Theory of computation → Circuit complexity; Theory of computation → Complexity classes

Keywords and phrases Circuit lower bounds, constant degree hypothesis, permutation groups, CC^0-circuits

Digital Object Identifier 10.4230/LIPIcs.STACS.2025.58

Related Version *Full Version*: `https://arxiv.org/abs/2311.17440` [32]

Funding *Piotr Kawałek*: This research was funded in whole or in part by National Science Centre, Poland #2021/41/N/ST6/03907. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. Funded by the European Union (ERC, POCOCOP, 101071674). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.
Armin Weiß: Partially funded by German DFG Grant WE 6835/1-2.

1 Introduction

Establishing strong lower bounds for general Boolean circuits represents one of the paramount and yet unattained objectives in the field of Computational Complexity Theory. Whenever such lower bounds can be obtained, it is usually in some very restricted setting. One of the standard limitations imposed on circuits in this context is the restriction of their depth. Some strong results were obtained when the circuits have depth bounded by a constant h

42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025).
Editors: Olaf Beyersdorff, Michał Pilipczuk, Elaine Pimentel, and Nguyễn Kim Thắng;
Article No. 58; pp. 58:1–58:21
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and are built of unbounded fan-in Boolean AND/OR gates and unary ¬ gates (so-called AC^0 circuits). By a classical result of Furst, Saxe and Sipser [19], proved independently by Ajtai [1], polynomial-size AC^0 circuits cannot compute the PARITY function (i.e., the sum of the input bits modulo 2). In fact, a followup paper by Yao [38] strengthens the lower bound for n-ary PARITY to be of the form $2^{\Omega(n^c)}$, with a final result of Håstad [26] finding a precise $c = \frac{1}{h-1}$. Interestingly, extending the AC^0 lower bounds from PARITY to MOD_m (i.e. the characteristic function of addition modulo an arbitrary integer m) can be achieved by the very same proof as in [26]. For a precise formulation and even more general results in this direction, see the subsequent work by Smolensky [36].

Here, a natural dual question arises: can modulo counting gates represent the n-ary Boolean AND_n function in the bounded-depth setting? To be more precise, a $\mathsf{CC}_h[m]$ circuit is a circuit of depth h using only (unbounded fan-in) MOD_m^R gates. Each such gate sums the inputs modulo m and outputs 1 if the sum belongs to the set R (we allow different $R \subseteq [m]$ for different gates), otherwise it outputs 0. Thus, the question is, after fixing h and m, what size does a $\mathsf{CC}_h[m]$ circuit require to compute AND_n? Is there a polynomial-size construction for AND_n, making the class ACC^0 collapse to CC^0 (where $\mathsf{CC}^0 = \bigcup_{h,m} \mathsf{CC}_h[m]$)? The first question has a trivial answer when m is a prime power, as then $\mathsf{CC}_h[m]$ circuits can express only bounded-arity AND (see [5] or [29] for more details). Surprisingly, for m having multiple prime divisors, only slightly super-linear lower bounds are known [12] and only for the number of wires – even more: to the best of our knowledge it is consistent with the current understanding that $\mathsf{NP} \subseteq \mathsf{CC}_2[6]$. At the same time, the current best construction for AND_n has size $2^{O(n^c)}$ [11, 29] for some constant c depending on h and m.

This huge gap between lower and upper bounds suggests that the problem of establishing lower bounds in this context is very difficult. Hence, one can consider simpler computational models before answering the above more general questions. Interestingly, group theory outlines in-between steps which can be considered in this context. Barrington, Straubing and Thérien [5] studied a model of non-uniform DFA (NuDFA) over finite groups (or, more generally, monoids), which they used to recognize Boolean languages. They discovered that, if a group is an extension of a p-group by an abelian group, then its corresponding NuDFA can recognize all languages (however, most of them in exponential size). Nevertheless, such NuDFAs cannot compute AND_n unless they have size at least $2^{\Omega(n)}$ [5]. Later this result was restated in a circuit language, saying that, if m is an integer and q is a prime, then any 2-level $\mathrm{MOD}_q \circ \mathrm{MOD}_m$ circuit computing AND_n requires size $2^{\Omega(n)}$ [23, 22, 37] (here, as usual, the circuits have to be read that the MOD_q gate is the output gate – other than e.g. in [29]). The equivalence of the two statements is due to the fact that these (solvable) groups have an internal structure based on modulo counting. The authors of [5] conjectured that this $2^{\Omega(n)}$ lower bound generalizes to NuDFAs over extensions of nilpotent groups by p-groups. This again can be reformulated to a $2^{\Omega(n)}$ lower bound for $\mathrm{MOD}_q \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ circuits computing AND_n (see [23]). This conjecture is known as Constant Degree Hypothesis (CDH for short), whose name corresponds to adding a layer of constant-arity AND_d gates on the input level to a $\mathrm{MOD}_q \circ \mathrm{MOD}_m$ circuit. Interestingly enough, recently in [30] it was proven that all the other groups (which do no correspond to CDH) do not admit this lower bound, i.e. one can construct AND_n of size $2^{O(n^c)}$ for some $c < 1$ using NuDFAs (or corresponding circuits) over these groups. In particular, it follows from [30] as well as the related work [3, 29] that the only $\mathrm{MOD}_{m'} \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ circuits (where m, m' are arbitrary integers) for which subexponential constructions of AND_n are *not* known are either the ones described by CDH (i.e., m, m' prime) or such circuits with $m = p^\alpha, m' = p^\alpha q^\beta$, where p^α, q^β are powers of different primes. However, in the latter case, replacing $m' = p^\alpha q^\beta$ with just q^β does not meaningfully change the expressive power of the related circuits (based on [29]). As a result, $\mathrm{MOD}_q \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ circuits are really the only (algebraically) natural subclass of CC^0 circuits for which these strong $2^{\Omega(n)}$ lower bounds remain to be proven (or disproven).

Low-level CC^0 circuits have many surprising connections. For instance, the techniques used in the construction of relatively small CC^0 circuits for the OR_n function (equivalently, AND_n) found in [3] are useful in constructing small explicit Ramsey-type graphs [21, 20]. These constructions are also used to produce better locally-decodable error-correcting codes [17, 15], private information retrieval schemes [16], and secret sharing schemes [33]. The lower bounds for codes considered in [17] imply lower bounds for certain CC^0 circuits. On the contrary, good lower bounds for low-level CC^0 circuits imply faster algorithms for solving equations in solvable groups [30], faster algorithms for certain algebraic versions of circuit satisfiability problems [28] and also faster algorithms for some variants of the Constraint Satisfaction Problem with Global Constraints [8].

These diverse interconnections encourage to put even more effort to find the correct sizes for optimal modulo-counting circuits computing AND_n. In this pursue, proving (or disproving) CDH plays a central role. The hypothesis is already proven in several special cases: in particular, the case $d = 1$ was confirmed in the very same paper the hypothesis was defined. Moreover, if there is a bound on the number of AND_d gates that are wired to each MOD_m gate, the desired lower bound is also true [23]. More precisely, the number of such connections is required to be $o(\frac{n^2}{\log n})$. The technique used in this case is based on the so-called *Degree Decreasing Lemma*, whose name corresponds to gradually decreasing the degree d, which eventually leads to the $d = 1$ case. The Degree Decreasing Lemma can also be used when the polynomials over $\mathbb{Z}_m$ corresponding to the $\mathrm{MOD}_m \circ \mathrm{AND}_d$ part of the circuit can be written using a sublinear number of binary multiplications [21].

In many studies of different circuit complexity classes, *symmetry* seems to play an important role. In this context both symmetric circuits, as well as symmetric functions were considered. Here, symmetry for a circuit/function means that permuting its inputs/variables does not change the considered circuit/function. Let us here mention the recent results on lower bounds for symmetric arithmetic circuits for the permanent and also a construction of short symmetric circuits for the determinant [13] as well as the lower bound from [27] for computing a certain entry in a product of matrices (here, symmetry means invariance under permuting rows and columns of matrices).

Symmetry seems to play also a special role for $\mathsf{CC}_h[m]$ circuits. The remarkable construction of relatively small circuits for AND_n in [3] uses symmetric polynomials as an intermediate object before translating them to circuits. This translation, when done carefully, leads also to symmetric circuits. Similarly, some of the newer, more optimal constructions of two level $\mathsf{CC}_2[m]$ circuits for AND_n can be performed fully symmetrically [11, 29]. Additionally, [23, 22, 37] analyze the periodic behaviour of the symmetric functions that can be represented by small (not necessarily symmetric) $\mathrm{MOD}_q \circ \mathrm{MOD}_m$ circuits. A value of a symmetric Boolean function $f(x_1, \ldots, x_n)$ is determined by the number of ones among $x_1, \ldots, x_n$. Hence, for an integer $0 \leq m \leq n$, we can naturally define $f(m)$ as $f(1^m 0^{n-m})$ and say that an integer r is a period of f whenever $f(m + r) = f(m)$ for all $0 \leq m \leq n - r$. It follows from [23, 37] that the only symmetric functions that have representations as $\mathrm{MOD}_q \circ \mathrm{MOD}_m$ circuits of subexponential size must have periods of the form $m \cdot q^k$ with $m \cdot q^k \leq n$. In particular, AND_n must have exponential-size circuits.

The dual question, namely the behaviour of symmetric functions computed by small AC^0 circuits, has been studies quite a lot: In [14], polynomial-size symmetric AC^0 circuits of arity n are shown to represent only functions that are constant on the interval $\{n^\varepsilon, \ldots, n - n^\varepsilon\}$ (for large enough n).

The same result has been obtained in [18] also showing that, if a symmetric function f is constant on the interval $\{\log^k n, \dots, n - \log^k n\}$ (for some k and for large enough n), then it is in AC^0. Soon after, [9] showed that the latter condition is actually an if and only if.

These results were extended to $\mathsf{AC}^0[p]$ circuits of quasipolynomial size by Lu [34]: $f = (f_n)_{n\in\mathbb{N}}$ is symmetric with $f \in \mathsf{qAC}^0[p]$ if and only if f_n has period $p^{t(n)} = \log^{O(1)} n$ except at both ends of length $\log^{O(1)} n$. Here, for a function $f : \{0,1\}^* \to \{0,1\}$, we write $f = (f_n)_{n\in\mathbb{N}}$ where $f_n : \{0,1\}^n \to \{0,1\}$ is the restriction of f to $\{0,1\}^n$. As usual we say that f is computed by a family of circuits if for each n there is a circuit computing f_n.

For further results in this direction allowing threshold or majority gates see [4, 24, 39]. Recently, a new technique called torus polynomials were introduced [6] as a possible method to separating TC^0 from ACC^0 and was shown that MAJORITY cannot be approximated by small-degree *symmetric* torus polynomials.

Contribution. In this paper we prove that symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits computing AND_n have exponential size. A key to the proof is to analyze the periodic behaviour of the functions computed by such circuits. Our techniques work also when d is unbounded and is considered as a function of n. The following theorem characterizes the periodic behaviour of such functions.

▶ **Theorem 1.** *Let p and q be primes and $n \geq 13$ and let $1 \leq d \leq n$. Then any function computed by an n-input symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit of size $s < 2^{n/9}$ has a period $p^{k_p}q^{k_q}$ given that $p^{k_p} > d$ and $q^{k_q} > \log s + 1$.*

To fully understand the periodic behaviour of $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits, we would also like to construct relatively small circuits given a function f with period of the form $p^{k_p}q^{k_q}$. We present such a construction below in Proposition 19. The most interesting consequence of this construction is that we get a tight characterization of the periodic behaviour of functions computed by quasipolynomial-size $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits (recall that a quasipolynomial is a function of the form $2^{\log^k n}$ for some constant k).

▶ **Corollary 2.** *Let $p \neq q$ be primes and $d : \mathbb{N} \to \mathbb{N}$ with $d(n) \leq n/2$ for all n. A function $f = (f_n)_{n\in\mathbb{N}}$ (with $f_n : \{0,1\}^n \to \{0,1\}$) can be computed by symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_{d(n)}$ circuits of quasipolynomial size if and only if, for each n, f_n has a period $p^{k_p(n)}q^{k_q(n)} \in \log^{O(1)}(n)$ for some functions $k_p, k_q \colon \mathbb{N} \to \mathbb{N}$.*

Next let us consider the case of the AND_n function more carefully. The following theorem is a careful application of Theorem 1.

▶ **Theorem 3.** *Let p and q be primes, let n be a large enough integer, and let $d \leq \sqrt{n}$. Then every symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit computing the AND_n function has size at least $2^{n/(2dpq)}$.*

Note that the restriction $d \leq \sqrt{n}$ still includes the most interesting case. Indeed, for $\sqrt{n} \leq d \leq n - \sqrt{n}$ we get an almost trivial lower bound of $2^{\sqrt{n}}$ (see Theorem 21). Moreover, Theorem 3 suggests an interesting trade-of between the degree and the size at $d \approx \sqrt{n}$. Then we can reach a lower bound for the size of the form $2^{\Omega(\sqrt{n})}$.

As a direct consequence of Theorem 3 we get the desired result for AND_n.

▶ **Corollary 4.** *For constant d, and primes p, and q every symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit for AND_n has size at least $2^{\Omega(n)}$. Thus, CDH holds for symmetric circuits with p being prime.*

Before we go to the more technical part, let us briefly mention an opposite perspective on the results of this paper. Although current evidence seems to support CDH and lower bounds for AND_n for general $\mathsf{CC}_h[m]$ circuits, it is known that $\mathsf{CC}_h[m]$ circuits using $O(\log n)$ random bits are able to compute AND_n in polynomial size [25]. This was even improved in [31], by showing that $\mathrm{MOD}_q \circ \mathrm{MOD}_p$ circuits can also be used for representing AND_n in this probabilistic model. This might be interpreted as a argument against lower bounds, because now it is enough to derandomize the construction for $\mathrm{MOD}_q \circ \mathrm{MOD}_p$ circuits. We already understand these 2-level circuits relatively well (to the point that we can prove strong lower bounds for them for the AND_n function itself). Our Corollary 4 implies that one cannot construct AND_n with polynomial-size symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits. Hence, to make short deterministic constructions one needs to either go beyond the symmetric setting or consider larger depths.

Outline. The paper is organized as follows: in Section 2 we fix our notation on circuits as well as hypergraphs and group actions. These notions are essential in the later study of the symmetric structure of circuits. In Section 3, we describe how to rewrite a circuit into a nicer form that we use throughout the paper. Then in Section 4 we present our key lemmas including proofs or short proof sketches and show how to derive our main results. The missing proofs can be found in the full version on arXiv [32]. In Section 5 we add some discussion of our results.

2 Preliminaries

Hypergraphs. For $d \in \mathbb{N}$ we write $[d]$ for the set of integers $\{1, \ldots, d\}$. For a set X we denote its power set by $\mathcal{P}(X)$. A *hypergraph* on a set of vertices V is a pair (V, E) with $E \subseteq \mathcal{P}(V) \setminus \{\emptyset\}$. A $\mathbb{F}_p$*-labeled hypergraph* is a pair $G = (V, \lambda)$ where $\lambda\colon (\mathcal{P}(V) \setminus \{\emptyset\}) \to \mathbb{F}_p$. We obtain an (unlabeled) hypergraph by setting $E = \{e \subseteq V \mid \lambda(e) \neq 0\}$ and call each e with $\lambda(e) \neq 0$ an *edge* of G. Thus, an $\mathbb{F}_p$-labeled hypergraph is indeed a hypergraph where we assign to each edge a number from $\mathbb{F}_p \setminus \{0\}$. Moreover, (V, λ) is called an $\mathbb{F}_p$*-labeled* d*-hypergraph* if for all $e \in \mathcal{P}(V)$ with $|e| > d$ we have $\lambda(e) = 0$. We write $\mathcal{H}_p^d(V)$ for the set of $\mathbb{F}_p$-labeled d-hypergraphs on V. For $C \subseteq V$ we write $\overline{C} = V \setminus C$ for the complement of C.

If $G = (V, \lambda)$ and $H = (V, \zeta)$ are $\mathbb{F}_p$-labeled hypergraphs on the same set of vertices V, we define $G + H$ (resp. $G - H$) as $(V, \lambda + \zeta)$ (resp. $(V, \lambda - \zeta)$) where $\lambda + \zeta$ denotes the point-wise addition. We interpret any subset $E \subseteq \mathcal{P}(V) \setminus \{\emptyset\}$ as a hypergraph by setting $\lambda(e) = 1$ if $e \in E$ and $\lambda(e) = 0$ otherwise (be aware of the slight ambiguity as V is not uniquely defined by E – but it always will be clear from the context). Thus, we have defined the addition $G + E$ (resp. $G - E$). We extend this to $G + e = G + \{e\}$.

Permutation groups. For any set V we denote the group of permutations on V by $\mathrm{Sym}(V)$ (i.e., the symmetric group). For an integer n we write $\mathrm{Sym}(n)$ or S_n for the abstract symmetric group acting on any n-element set. Any subgroup $\Gamma \leq \mathrm{Sym}(V)$ *acts* faithfully on V and is called a permutation group. A subset $U \subseteq V$ is called an *orbit* of the action of Γ on V if $U = G \cdot x$ for some $x \in V$. If there is only one orbit, the action of Γ on V is called transitive. Clearly, the orbits form a partition of V; moreover, if $U_1, \ldots, U_k \subseteq V$ are the orbits of the action of $\Gamma \leq \mathrm{Sym}(V)$ on V, then $\Gamma \leq \mathrm{Sym}(U_1) \times \cdots \times \mathrm{Sym}(U_k)$ where × denotes the direct product of groups.

Finally, let $\Gamma' \leq \Gamma$ be a subgroup. A *left-transversal* (in the following simply *transversal*) of Γ' in Γ is a subset $R \subseteq \Gamma$ such that R is a system of representatives of Γ/Γ' – in other words, if $R\Gamma' = \Gamma$ and $r\Gamma' \cap s\Gamma' = \emptyset$ for $r, s \in R$ with $r \neq s$. For further details on permutation groups, we refer to [10].

Actions on hypergraphs. Given an action of $\mathrm{Sym}(V)$ on V, it induces an action on $\mathcal{P}(V)$. Moreover, this extends to an action on $\mathcal{H}_p^d(V)$ where a permutation $\pi \in \mathrm{Sym}(V)$ maps (V, λ) to $\pi((V, \lambda)) = (V, {}^{\pi}\lambda)$ for ${}^{\pi}\lambda$ defined by $({}^{\pi}\lambda)(e) = \lambda(\pi^{-1}(e))$. Be aware that the ${}^{-1}$ is not by accident but rather guarantees that if some $e \in \mathcal{P}(V)$ has label $\gamma = \lambda(e)$, then $\pi(e)$ has label ${}^{\pi}\lambda(\pi(e)) = \lambda(e)$. Note that two labeled hypergraphs with vertices V are isomorphic if and only if they are in the same orbit under $\mathrm{Sym}(V)$. A permutation $\pi \in \mathrm{Sym}(V)$ is called an *automorphism* of $G = (V, \lambda)$ if $\pi(G) = G$ – with other words, if $\lambda(\pi(e)) = \lambda(e)$ for all $e \in \mathcal{P}(V)$. For a labeled hypergraph G, we denote its group of automorphisms by $\mathrm{Aut}(G)$.

Circuits. A circuit is usually defined as a directed acyclic graph with labels on its vertices that inform what kind of operation (like for instance $\wedge, \vee, \neg$, MOD_p^R) a given vertex (gate) computes. We allow multiple edges between any pair of gates. A depth-d circuit of arity n is a circuit that consists of n inputs gates $x_1, \ldots, x_n$ and d layers (or levels) $G_1, \ldots, G_d$ of inner gates (we do not count the input gate as a level). Between neighbour layers G_{i-1} and G_i there is a layer of wires W_i which contains directed edges between $g \in G_{i-1}$ and $h \in G_i$ (where $G_0 = \{x_0, \ldots, x_n\}$). We allow for multiple (directed) edges between the same pair of gates. Moreover, gates are labeled with necessary information which allows to compute a function they represent. In our case we use MOD_p^R gates where p is a prime and $R \subseteq \mathbb{F}_p$. A MOD_p^R with inputs $y_1, \ldots, y_k$ outputs 1 if and only if the sum of its inputs modulo p is contained in R. A circuit is called an *expression* if it is a tree when removing the input layer. A subexpression of an expression is a subgraph containing for every gate also all its predecessors (towards the input gates). For circuits C, D with n inputs we write $C \equiv D$ if for all inputs $\bar{b} \in \{0,1\}^n$ they evaluate to the same value. We define the *size* of a circuit as its number of non-input gates.

In this article we consider $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits: such a circuit consist of 3-levels. On level 1 there are AND_d gates each of which receives inputs from at most d input gates. The second level G_2 consists of MOD_p^R gates – each of them is labeled with an accepting set $R \subseteq \{0, \ldots, p-1\}$. The output layer G_3 contains only one MOD_q^R gate, which sums all the wires from W_3 modulo q.

We say that a circuit C is symmetric if no permutation of the input wires changes the circuit. Note that here the word *symmetric* refers to a *syntactic* structure of a circuit, rather than a semantic property of the function computed by it. More formally, a circuit C on inputs $x_1, \ldots, x_n$ is called *symmetric* if for any $\pi \in \mathrm{Sym}(\{x_1, \ldots, x_n\})$ there is a permutation π' on the set of gates extending π (meaning that $\pi(x_i) = \pi'(x_i)$ for all $i \in [n]$) such that there are k wires connecting gate i to gate j if and only if there are k wires connecting gates $\pi'(i)$ to gate $\pi'(j)$.

3 Preparation: Circuits, Expressions and Hypergraphs

For a simpler notation of expressions, let us denote MOD_p^R with inputs $y_1, \ldots, y_k$ instead by $\mathbf{b}(\sum_{i=1}^k y_i; R)$ for $R \subseteq \mathbb{F}_p$, where $\mathbf{b}$ computes the function

$$\mathbf{b}(y; R) = \begin{cases} 1 & \text{if } y \in R \\ 0 & \text{if } y \notin R. \end{cases}$$

Be aware that we use $\mathbf{b}$ for different domains, i.e. as a function $\mathbb{F}_p \to \{0,1\}$ and $\mathbb{F}_q \to \{0,1\}$. The domain will be clear from the context.

From circuits to expressions. Any 2-level $\mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit corresponds to polynomial over the field $\mathbb{F}_p$. Indeed, the AND_d gates act like a multiplications on the two element domain $\{0,1\} \subseteq \mathbb{F}_p$ and the MOD_p^R gate sums the results and checks whether the sum belong to the accepting set R. Because our circuits are of constant-depth, we can unfold the circuits to obtain expressions. This means, if a gate g has an outgoing wire to several other gates, we create multiple copies of g, so that each gate has only a single output wire. Note that this might lead to a polynomial blow-up in size (more precisely, a circuit with size s and depth bounded by h is converted to an expression of size at most s^{h-1} – thus, in our case s^2). Moreover, note that unfolding the circuit does not destroy the property of being symmetric. Hence, every symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit yields also a symmetric expression

$$\mathbf{b}\Big(\sum_{i=0}^{l} \alpha_i \,\mathbf{b}(\mathbf{p}_i(\overline{x}); R_i); R\Big) \tag{1}$$

for suitable $\alpha_i \in \mathbb{F}_q$, $R_i \subseteq \mathbb{F}_p$ and polynomials $\mathbf{p}_i$ of degree bounded by d for $i \in \{1, \dots, l\}$ and $R \subseteq \mathbb{F}_q$ which computes the same function. Here, l is the number of MOD_p gates used in the $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit, while α_i tells us how many times a given MOD_p gate is wired to the MOD_q gate. Let us take a closer look at what being symmetric means for an expression of the form (1): for each $\pi \in \mathrm{Sym}(n)$ there exist $\pi' \in S_l$ such that for all $i \in \{1, \dots, l\}$ we have $\alpha_i = \alpha_{\pi'(i)}$, $R_i = R_{\pi'(i)}$, and $\mathbf{p}_i(x_1, \dots, x_n) = \mathbf{p}_{\pi'(i)}(x_{\pi(1)}, \dots, x_{\pi(n)})$ (here $=$ refers to equality in the polynomial ring $\mathbb{F}_p[x_1, \dots, x_n]$).

Next, observe that if we omit the outer $\mathbf{b}$ of the expression (1), the function computed by the resulting expression certainly does not have any new (smaller) periods than the one of the complete expression. Therefore, as we are aiming for an upper bound on the periods of the considered symmetric circuits, we will now concentrate on the symmetric expressions of the form

$$\mathbf{f} = \sum_{i=0}^{l} \alpha_i \,\mathbf{b}(\mathbf{p}_i(\overline{x}); R_i) \tag{2}$$

with $R_i \subseteq \mathbb{F}_p$, $\alpha_i \in \mathbb{F}_q$ and polynomials $\mathbf{p}_i$ of degree bounded by d. Indeed, every period of $\mathbf{f}$ is a period of $\mathbf{b}(\mathbf{f})$, so for proving lower bounds, it is enough to consider the periods of $\mathbf{f}$. An expression of the form (2) is called a $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression and each $\mathbf{b}(\mathbf{p}_i(\overline{x}); R_i)$ is an *elementary subexpression* of $\mathbf{f}$.

In the following, let us write $\mathbf{b}(\mathbf{p}(\overline{x}); r)$ for $\mathbf{b}(\mathbf{p}(\overline{x}); \{r\})$. Using this notation we have $\mathbf{b}(\mathbf{p}(\overline{x}); R) = \sum_{r \in R} \mathbf{b}(\mathbf{p}(\overline{x}); r)$. Moreover, we always assume that for $i \neq j$ we have $(\mathbf{p}_i, r_i) \neq (\mathbf{p}_j, r_j)$ as otherwise we can replace $\alpha_i \,\mathbf{b}(\mathbf{p}_i(\overline{x}); r_i) + \alpha_j \,\mathbf{b}(\mathbf{p}_i(\overline{x}); r_j)$ by $\alpha_{ij} \,\mathbf{b}(\mathbf{p}_i(\overline{x}); r_i)$ where $\alpha_{ij} = \alpha_i + \alpha_j$. Thus, using $\mathrm{pol}(n,d)$ to denote the set of multilinear polynomials in $\mathbb{F}_p[x_1, \dots, x_n]$ with degree bounded by d, we rewrite $\mathbf{f}$ in (2) as

$$\mathbf{f} = \sum_{\mathbf{p} \in \mathrm{pol}(n,d)} \sum_{r \in \mathbb{F}_p} \alpha_{\mathbf{p},r} \,\mathbf{b}(\mathbf{p}(\overline{x}); r). \tag{3}$$

Note that to compute the size of $\mathbf{f}$ we only need to count the non-zero $\alpha_{\mathbf{p},r}$ (plus the number of AND gates computing the polynomials $\mathbf{p}$).

Polynomials and hypergraphs. Let us take a closer look at the $\mathrm{MOD}_p \circ \mathrm{AND}_d$ part of a $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit or expression. As any such expression is represented by a polynomial of degree d, we will need to deal with these polynomials and their symmetries. Notice that without loss of generality, we can assume that the polynomial corresponding to a $\mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit is multi-linear since, because the values of variables are restricted to $\{0,1\}$ each occurrence of a higher power x^k of a variable x can be simply replaced by x.

In order to deal better with the combinatorics and symmetries of polynomials, we think of polynomials as hypergraphs. A multilinear polynomial $\mathbf{p} \in \mathbb{F}_p[x_1, \ldots, x_n]$ with the degree bounded by d can be naturally identified with an $\mathbb{F}_p$-labeled d-hypergraph $G = (V, \lambda)$ as follows:

1. Treat each variable x_i in $\mathbf{p}(x_1, \ldots, x_n)$ as a vertex in the graph $G_{\mathbf{p}}$. Thus, $V = \{x_1, \ldots, x_n\}$, which we also identify with the set $[n]$.
2. Each monomial $\gamma \cdot x_1 \cdot \ldots \cdot x_d$ is represented by a hyperedge with a label γ, i.e., we have $\lambda(\{x_1, \ldots, x_d\}) = \gamma$.

Thus, we get a one-to-one correspondence between multilinear polynomials over $\mathbb{F}_p$ of degree at most d and $\mathbb{F}_p$-labeled d-hypergraphs. This means that we can also do the reverse – for each labeled graph G we can create its corresponding polynomial $\mathbf{p}_G$. Moreover, note that also the arithmetic operations we defined on hypergraphs as well as the group actions agree with those on polynomials. Therefore, in the following, we use polynomials and hypergraphs interchangeably.

Now, we can use our graph notation for polynomials in a more general setting and denote each expression $\mathbf{b}(\mathbf{p}(\overline{x}); r)$ by $\mathbf{b}(G_{\mathbf{p}}; r)$ or simply $\mathbf{b}(G; r)$ (when we start with a hypergraph representing a given polynomial). Thus, we can reformulate any expression of the form (3) as $\sum_{\mathbf{p} \in \mathrm{pol}(n,d)} \sum_{r \in \mathbb{F}_p} \alpha_{\mathbf{p},r}\, \mathbf{b}(G_{\mathbf{p}}; r) = \sum_{G \in \mathcal{H}_p^d(V)} \sum_{r \in \mathbb{F}_p} \alpha_{G,r}\, \mathbf{b}(G; r)$.

Symmetric expressions induced by hypergraphs. Now we define several notions, useful in analysing symmetric expressions. For $G = (V, \lambda)$ and $\pi \in \mathrm{Sym}(V)$, let us write $\mathbf{b}^{\pi}(G; R)$ for $\mathbf{b}(\pi G; R)$. The action of $\mathrm{Sym}(V)$ on V now extends naturally to an action on expressions of the form $\mathbf{f} = \sum_{G \in \mathcal{H}_p^d(V)} \sum_{r \in \mathbb{F}_p} \alpha_{G,r}\, \mathbf{b}(G; r)$ by setting

$$\pi(\mathbf{f}) = \sum_{G \in \mathcal{H}_p^d(V)} \sum_{r \in \mathbb{F}_p} \alpha_{G,r}\, \mathbf{b}^{\pi}(G; r).$$

Now, $\mathbf{f}$ being symmetric can be simply expressed as the fact that for each $\pi \in \mathrm{Sym}(V)$ we have $\pi(\mathbf{f}) = \mathbf{f}$.

▶ **Definition 5.** *Let $G = (V, \lambda)$ be a labeled d-hypergraph. Let $\mathrm{Aut}(G)$ be its group of automorphisms and let $\pi_1, \ldots, \pi_k$ be a transversal of $\mathrm{Sym}(V)/\mathrm{Aut}(G)$. For a given $r \in \mathbb{F}_p$, define $\mathbf{s}(G; r)$ to be the following $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression*

$$\mathbf{s}(G; r) = \sum_{i=0}^{k} \mathbf{b}^{\pi_i}(G, r). \tag{4}$$

One needs to check that the above definition does not depend on the choice of the transversal, as there is a choice in picking the specific traversal $\pi_1, \ldots, \pi_k$ which we use to create $\mathbf{s}(G; r)$. However, as G is invariant under its automorphisms, no matter how we choose the specific $\pi_1, \ldots, \pi_k$, we get the same expression in the end. In fact, every symmetric $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression containing $\mathbf{b}(G; r)$ as subexpression, must also contain $\mathbf{s}(G; r)$ as subexpression. So $\mathbf{s}(G; r)$ is a symmetric closure of the basic expression $\mathbf{b}(G; r)$. Let us summarize this as follows:

▶ Remark 6. For every labeled d-hypergraph G and every $r \in \mathbb{F}_p$, the expression $\mathbf{s}(G; r)$ is symmetric. Moreover, it is the smallest symmetric expression that contains $\mathbf{b}(G; r)$ as an elementary subexpression.

▶ **Fact 7.** *Every symmetric $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression* $\mathbf{f}$ *can be written as a sum*

$$\mathbf{f}(\overline{x}) = \sum_{G \in \mathcal{H}_p^d(V)} \sum_{r \in \mathbb{F}_p} \beta_{G,r} \cdot \mathbf{s}(G;r)$$

for $\beta_{G,r} \in \mathbb{F}_q$ (recall that $\mathcal{H}_p^d(V)$ denotes the set of labeled d-hypergraphs on V).

Proof. If a symmetric $\mathbf{f}$ has some $\beta \cdot \mathbf{b}(G,r)$ as an elementary subexpression, it must also have $\beta \cdot \mathbf{s}(G;r)$ as a subexpression (see Remark 6). But now $\mathbf{f} - \beta \cdot \mathbf{s}(G;r)$ is a symmetric expression which is shorter than $\mathbf{f}$, and hence we can use induction to prove the desired decomposition for $\mathbf{f}(\overline{x})$, by adding $\beta \cdot \mathbf{s}(G;r)$ to the decomposition of $\mathbf{f} - \beta \cdot \mathbf{s}(G;r)$. ◀

4 Description of the Proof

We now start with an expression as in Fact 7 and prove our main theorems. For this, we need several definitions and intermediate results. For some of these intermediate results, the full proofs are deferred to the full version [32]; instead, we present short proof sketches, give some high-level ideas how the respective results are used, and then, in Section 4.4, show how our main results follow from the intermediate results. As every symmetric expression $\mathbf{f}$ is decomposed into an appropriate sum of elements of the form $\alpha \cdot \mathbf{s}(G;r)$, we need a deeper understanding of each $\mathbf{s}(G;r)$. We investigate these expressions $\mathbf{s}(G;r)$ in three main steps:

1. we analyze the symmetries of G to find a large so-called *fully symmetric set* (see Lemma 10),
2. we process the hypergraph G further to make it *symmetry purified* (see Definition 13 and Lemma 14) applying two versions of the Degree Decreasing Lemma (Lemma 11 and Lemma 12),
3. we analyze the periods of the resulting expressions $\mathbf{s}(G;r)$ (see Theorem 16).

4.1 Symmetries of Hypergraphs

Recall that one of our goals is to prove exponential lower bounds on the size of symmetric circuits/expressions computing AND_n. In Lemma 10 we are going to show that, if in an expression $\mathbf{f}$ we find a very asymmetric graph G, we know that the size of $\mathbf{f}$ must be relatively large. This is because the automorphism group of G is small and, hence, the length of the expression of the form (4) induced by G, i.e. $\mathbf{s}(G;R)$, must be large (more precisely, k as defined above is large). On the other hand, for highly symmetric graphs G, we can find a big, very regular substructure of G, which we will call a *pseudo-clique.*

▶ **Definition 8.** *Let G be an $\mathbb{F}_p$-labeled hypergraph $G = (V, \lambda)$ (i.e. $\lambda : \mathcal{P}(V) \setminus \{\emptyset\} \to \mathbb{F}_p$). We say that a subset $C \subseteq V$ is* fully symmetric*, if for each pair of subsets $e_1, e_2 \subseteq V$ with $|e_1| = |e_2|$ and $e_1 \cap \overline{C} = e_2 \cap \overline{C}$ we have $\lambda(e_1) = \lambda(e_2)$.*

Moreover, an $\mathbb{F}_p$-labeled hypergraph $G = (V, \lambda)$ is called a pseudo-clique *if $\mathrm{Aut}(G) = \mathrm{Sym}(V)$ – or, equivalently, if for each $d \in [n]$ there is some λ_d such that $\lambda(e) = \lambda_d$ all $e \subseteq V$ with $|e| = d$.*

Note that an induced subgraph on a fully symmetric subset of vertices is a pseudo-clique. We obtain the following easy observations.

▶ **Fact 9.** *Let G be an $\mathbb{F}_p$-labeled hypergraph $G = (V, \lambda)$.*

- *A subset $C \subseteq V$ is fully symmetric if and only if $\mathrm{Sym}(C) \leq \mathrm{Aut}(G)$.*
- *If $C, D \subseteq V$ are fully symmetric sets with $C \cap D \neq \emptyset$, then so is $C \cup D$.*
- *If $C \subseteq V$ is a maximal fully symmetric set with $|C| > |V|/2$, then $\mathrm{Aut}(G) = \mathrm{Sym}(C) \times \Gamma$ for some $\Gamma \leq \mathrm{Sym}(\overline{C})$.*

Now, we are ready to present a key lemma, which allows us to restrict our attention only to very symmetric hypergraphs.

▶ **Lemma 10.** *Let $0 < \varepsilon < 1/8$. Every $\mathbb{F}_p$-labeled hypergraph $G = (V, \lambda)$ with $n = |V| \geq 13$ has either*

- *a fully symmetric subset on at least $n - \lfloor \varepsilon n \rfloor$ vertices, or*
- *its automorphism group satisfies $|\operatorname{Sym}(V)/\operatorname{Aut}(G))| > 2^{\lfloor \varepsilon n \rfloor}$.*

Proof. Let us write $\Gamma = \operatorname{Aut}(G)$. We say that Γ is *small* if $|\Gamma| \leq n!/2^{\lfloor \varepsilon n \rfloor}$. Let us first show that either Γ is small or G contains a pseudo-clique on at least $n - \lfloor \varepsilon n \rfloor$ vertices (in a second step, we will show that this pseudo-clique, indeed, is fully symmetric).

We start by observing that Γ is a subgroup of $\operatorname{Sym}(k_1) \times \cdots \times \operatorname{Sym}(k_m)$, where k_i are the sizes of the orbits of the action on G. If $k_i < n - \lfloor \varepsilon n \rfloor$ for all i, then $|\Gamma| < (n - \lfloor \varepsilon n \rfloor)! \cdot \lfloor \varepsilon n \rfloor!$ and Γ is small because of

$$\frac{n!}{|\Gamma|} > \frac{n!}{(n - \lfloor \varepsilon n \rfloor)! \cdot \lfloor \varepsilon n \rfloor!} = \binom{n}{\lfloor \varepsilon n \rfloor} \geq \left(\frac{n}{\lfloor \varepsilon n \rfloor}\right)^{\lfloor \varepsilon n \rfloor} > 2^{\lfloor \varepsilon n \rfloor}.$$

So from now on there is one orbit $C \subseteq V$ consisting of at least $n - \lfloor \varepsilon n \rfloor$ vertices. Then we have $\Gamma \leq \operatorname{Sym}(C) \times \operatorname{Sym}(\overline{C})$ and we denote by $\varphi : \Gamma \to \operatorname{Sym}(C)$ the projection to the first coordinate.

Suppose $\tilde{\Gamma} = \varphi(\Gamma)$ does not act primitively on C meaning that there is an $\tilde{\Gamma}$-invariant partition of C with r classes each of which consists of $1 < m < |C|$ vertices (as $\tilde{\Gamma}$ acts transitively on C, it must acts transitively on the classes; hence, they all have the same size). Thus, $\tilde{\Gamma}$ is isomorphic to a subgroup of the wreath product $\operatorname{Sym}(m) \wr \operatorname{Sym}(r)$ with $rm = |C|$ (see [10, Theorem 1.8]). Hence,

$$\begin{aligned}\frac{|C|!}{|\tilde{\Gamma}|} &\geq \frac{|C|!}{(m!)^r \cdot r!} = \frac{1 \cdot \ldots \cdot m \cdot \qquad \ldots \qquad \cdot |C|}{(1 \cdots m) \cdot 1 \cdot (1 \cdots m) \cdot 2 \cdot \ldots \cdot (1 \cdots m) \cdot r} = \frac{\prod_{i=1}^{r-1} \prod_{j=1}^{m-1} (im + j)}{((m-1)!)^{r-1}} \\ &\geq 2^{(m-1)(r-1)} \geq 2^{|C|/4} \geq 2^{\lfloor \varepsilon n \rfloor}.\end{aligned}$$

Here the last inequality is because $\varepsilon < 1/8$ (in particular $1 - \varepsilon \geq 1/2$), the second last inequality is due to the assumption $\varepsilon \leq 1/4$ and $m - 1 \geq m/2$ and $r - 1 \geq r/2$. The third last inequality is because $\left(\prod_{j=1}^{m-1}(im + j)\right)/(m-1)! = \prod_{j=1}^{m-1}(im + j)/j \geq 2^{m-1}$ as $i \geq 1$. Since the index of Γ in $\operatorname{Sym}(V)$ is at least the index of $\tilde{\Gamma} = \varphi(\Gamma)$ in $\operatorname{Sym}(C)$, again Γ is small.

Hence, it remains to consider the case that $\varphi(\Gamma) \leq \operatorname{Sym}(C)$ acts primitively on C. Thus, writing $k = |C|$, according to [7, 35] (see also [2]), there are three possibilities: $\varphi(\Gamma)$ is either A_k (the alternating group on k elements) or $S_k \cong \operatorname{Sym}(C)$ or $|\varphi(\Gamma)| \leq 4^k$. First, let us consider the last case. As $n \geq 13$, we have $k \geq n - \lfloor \varepsilon n \rfloor \geq 11$. Therefore, we conclude that we have $|\varphi(\Gamma)| \leq 4^k \leq k!/2^{k/4}$ (which holds for all $k \geq 11$) and, as above, the index of Γ in $\operatorname{Sym}(V)$ is at least $2^{k/4} \geq 2^{(n - \lfloor \varepsilon n \rfloor)/4} \geq 2^{\lfloor \varepsilon n \rfloor}$, meaning that Γ is small.

In the former two cases (i.e., that $\varphi(\Gamma)$ is A_k or S_k), $\varphi(\Gamma)$ acts set-transitively on C (meaning that for each pair of subsets $A, B \subseteq C$ with $|A| = |B|$ there is a permutation π mapping A to B); hence, C is a pseudo-clique.

To see that C is, indeed, fully symmetric if Γ is not small, we proceed as follows: Now, let $N \leq \Gamma$ denote the kernel of the projection $\Gamma \to \operatorname{Sym}(\overline{C})$ to the second component (i.e. the pointwise stabilizer of $\overline{C}$). Then we have $\varphi(N) = N$ (when identifying $\operatorname{Sym}(C)$ with the corresponding subgroup of $\operatorname{Sym}(C) \times \operatorname{Sym}(\overline{C})$). As A_k is simple and the index of N is at most $(n-k)!$ in $\varphi(\Gamma)$ (which is either A_k or S_k), we have $N = A_k$ or $N = S_k$ (as N is also normal in $\varphi(\Gamma)$). In both cases N acts set-transitively on C and; hence, C is fully symmetric (which, by Fact 9, also excludes the case $N = A_k$). ◀

4.2 Reduction Based on the Degree Decreasing Lemma

One of the few examples of lower bounds for circuits using modulo counting are due to Grolmusz and Tardos [23, 22]. The authors prove lower bounds for $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits with restrictions put on connections between AND_d layer and MOD_p gates. More precisely, [22] shows that if the number of multiplications needed to compute the polynomial corresponding to each $\mathrm{MOD}_p \circ \mathrm{AND}_d$ subcircuit is bounded by cn for small enough c, then a $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit requires exponential size to compute AND_n. One of their key tools is the so-called Degree Decreasing Lemma:

▶ **Lemma 11** (Degree Decreasing Lemma). *Let $p \neq q$ be prime numbers. Then every function $f : \mathbb{F}_p^3 \mapsto \mathbb{F}_q$ represented by a 3-ary $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_2$ expression $\mathbf{b}(\gamma \cdot z_1 \cdot z_2 + y; t)$ can be also represented by an expression of the form*

$$\sum_{(j_1,j_2,j_3)\in\mathbb{F}_p^3} \sum_{r\in\mathbb{F}_p} \beta^{(r)}_{j_1,j_2,j_3} \cdot \mathbf{b}(j_1 z_1 + j_2 z_2 + j_3 y;\, r)$$

where $\beta^{(r)}_{j_1,j_2,j_3}$ are some coefficients from $\mathbb{F}_q$ (also depending on γ and t).

The Lemma is a consequence of the result by Grolmusz [22, Lemma 6] (note that the statement there does not include the factor γ, so formally, to obtain Lemma 11, one needs to apply [22, Lemma 6] several times). One can also see it as a consequence of [29, Fact 3.3]. This very simple lemma allows us to navigate through the space of different representations for a given function f by allowing a local change of its corresponding expression. The power of the lemma comes from the fact that we can substitute arbitrary polynomials for z_1, z_2, y and obtain many different equivalences.

We will need a more regular version of the Degree Decrasing Lemma when the multiplication inside $\mathbf{b}$ has bigger arity. The price we pay for a nicer form is that the represented function has a smaller (partially Boolean) domain, which slightly reduces the scope of applicability of the lemma (as we cannot substitute any polynomial for the variables); however, it still suffices for our purposes.

▶ **Lemma 12** (Symmetric Degree Decreasing Lemma). *Let $p \neq q$ be prime. Let $\gamma \in \mathbb{F}_p \setminus \{0\}$. Then every function $f : \{0,1\}^d \times \mathbb{F}_p \mapsto \mathbb{F}_q$ represented by a $d+1$-ary $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression*

$$\mathbf{b}(\gamma \cdot x_1 \cdot \ldots \cdot x_d + y; t)$$

can be also represented by an expression

$$\mathbf{h}(\overline{x}, y; t) = \mathbf{b}(y; t) + \sum_{r\in\mathbb{F}_p} \beta_{t,r} \sum_{S\subseteq[d]} \alpha_{|S|} \cdot \mathbf{b}(\gamma \cdot \sum_{s\in S} s + y; r)$$

for $\alpha_{|S|} = (-1)^{|S|}$ and some coefficients $\beta_{t,r} \in \mathbb{F}_q$.

The key property of the formula $\mathbf{h}$ is that it is invariant under permutations of the variables $x_1, \ldots, x_d$, which is not the case for original Degree Decreasing Lemma of [22]. The next step in the proof is to apply the Symmetric Degree Decreasing Lemma (Lemma 12) to expressions generated by highly-symmetric hypergraphs in order to obtain an even nicer representation defined as follows:

▶ **Definition 13.** *We call an $\mathbb{F}_p$-labeled d-hypergraph $G = (V, \lambda)$ symmetry-purified with respect to $C \subseteq V$ if*

1. *C is fully symmetric in G,*

2. *if $\lambda(e) \neq 0$, then e is completely contained either in C or in $\overline{C}$ (i.e., every edge e is fully contained either in C or in $\overline{C}$,*
3. *if $\lambda(e) \neq 0$ and $e \subseteq \overline{C}$, then $|e| = 1$ (i.e., every edge e with $e \cap C = \emptyset$ satisfies $|e| = 1$).*

Moreover, if the graph satisfies only conditions 1 and 2, we will call it partially symmetry purified*. We write* $\mathrm{sp}(V, C)$ *for the set of all symmetry-purified d-hypergraphs with respect to $C \subseteq V$ and* $\mathrm{psp}(V, C)$ *for the set of partially symmetry-purified d-hypergraphs with respect to $C \subseteq V$ (note that d and p are implicitly defined from the context for* $\mathrm{sp}(V, C)$ *and* $\mathrm{psp}(V, C)$*).*

The next crucial lemma allows us to restrict our attention only to expressions $\mathbf{s}(G; u)$ over symmetry-purified graphs, which have a very regular and much easier to analyze structure. This enables a later combinatorial analysis of the periodic behaviour of such $\mathbf{s}(G; u)$. The proof of the lemma relies on carefully applying both Lemma 11 and Lemma 12 to alter the graphs while preserving the symmetry of the corresponding expression.

▶ **Lemma 14.** *Let $p \neq q$ be prime numbers, let $u \in \mathbb{F}_p$, and let $G = (V, \lambda)$ be an $\mathbb{F}_p$-labeled d-hypergraph. Moreover, let $C \subseteq V$ be a maximal fully symmetric subset with $|C| > |V|/2$. Then there are constants $\beta_{H,r} \in \mathbb{F}_q$ such that*

$$\mathbf{s}(G; u) \equiv \sum_{H \in \mathrm{sp}(V,C)} \sum_{r \in \mathbb{F}_p} \beta_{H,r}\, \mathbf{s}(H, r).$$

Proof sketch. We will first represent a function computed by $\mathbf{s}(G; u)$ by a sum of expressions $\mathbf{s}(H, r)$ for H being partially symmetry purified. Let e be an edge in G, such that $e^C = e \cap C \neq \emptyset$ and $e^{\overline{C}} = e \cap \overline{C} \neq \emptyset$. We would like to remove the edge e from G. To do so, we apply Lemma 11 to replace e by a linear combination of e^C and $e^{\overline{C}}$. However, if we do this only for e we may destroy the symmetry of the resulting expression. For this reason, we pick not only e but its entire orbit $O = \mathrm{Aut}(G) \cdot e = \{e_1, \ldots, e_\ell\}$ and apply Lemma 11 simultaneously to it. Indeed, setting $P = \mathrm{Aut}(G) \cdot e^C$ and $Q = \mathrm{Aut}(G) \cdot e^{\overline{C}}$, since C is fully symmetric, we have

$$e_1 + \cdots + e_\ell = (\sum_{w \in P} w)(\sum_{v \in Q} v).$$

Since all the e_i must have the same label γ, we have

$$\mathbf{s}(G; u) = \mathbf{s}(\gamma(e_1 + \cdots + e_\ell) + G'; u) = \mathbf{s}(\gamma \cdot z_1 \cdot z_2 + G'; u)$$

where $z_1 = \sum_{w \in P} w$, $z_2 = \sum_{v \in Q} v$ and $G' = G - \gamma \cdot (e_1 + \cdots + e_\ell)$. Thus, after applying Lemma 11 to each summand of $\mathbf{s}$ (as in the formula (4)), we get a symmetric expression which is a linear combination of subexpressions of the form $\mathbf{b}^{\pi_i}(j_1 z_1 + j_2 z_2 + j_3 G'; r)$. Hence, we have replaced $\mathbf{s}(G, u)$ with a sum $\sum_{H,r} \mathbf{s}(H, r)$, where each graph H does not contain the edge e anymore and has no new edges intersecting both C and $\overline{C}$ non-trivially. We apply this reasoning recursively for $\mathbf{s}(H, r)$ as long as there is any edge in any graph in the sum that intersects non-trivially with C and $\overline{C}$. At the very end there are no such edges left, so we managed to represent $\mathbf{s}(G, u)$ as a sum of expressions $\mathbf{s}(H, r)$ over partially symmetry purified graphs H.

Next, assume that G is already partially symmetry purified with respect to C but not yet symmetry purified. Pick an edge $e \subseteq \overline{C}$ of size at least 2 and denote its orbit as $\mathrm{Aut}(G) \cdot e = \{e_1, e_2, \ldots, e_\ell\}$. We have

$$\mathbf{s}(G, u) = \sum_{i=1}^{k} \mathbf{b}^{\pi_i}(\gamma \cdot e_1 + \cdots + \gamma \cdot e_\ell + G'; u).$$

Now by applying Lemma 12 subsequently to the edges $e_1, \dots, e_\ell$, we can replace all these edges with linear combinations of its vertices. Because the formula in Lemma 12 is symmetric, we end up with a symmetric expression at the end. So we have replaced $\mathbf{s}(G, u)$ by a sum of $\mathbf{s}(H, r)$, but now each H has less bad edges. We apply this reasoning recursively to write $\mathbf{s}(G, u)$ as a sum of $\mathbf{s}(H, r)$ where each H is symmetry purified. ◀

4.3 Period of Symmetry-Purified Expressions

For a fixed input $\bar{b} \in \{0,1\}^n$ and an n-ary symmetric expression $\mathbf{f}$, we can compute the value $\mathbf{f}(\bar{b})$ only knowing the Hamming weight of the input, i.e. the number of 1s in $\bar{b}$. This means that $\mathbf{f}$ represents not only a function $\{0,1\}^n \to D$, but we can also view it as a function $\{0, 1, \dots, n\} \to D$. It turns out that a relatively small $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit can compute only functions with a relatively small period. Here, by a period of $\mathbf{f}$ we mean an integer $r \in \mathbb{N} \setminus \{0\}$ that satisfies $\mathbf{f}(m + r) = \mathbf{f}(m)$ for all m in the range $[0, n - r]$. Note that all functions have periods $> n$, so we are mainly interested in finding periods in the range $[1, n]$. Note that the AND_n function, which is of our particular interest, does not have any period (less than $n + 1$). Thus, proving an upper bound for a period of a function computed by a relatively short expression will give us a lower bound for the length of representation of AND_n. This is in line with some of the previous research [3, 22, 37]. As any $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ can be transformed into a symmetric expression over symmetry-purified graphs, we only need to concentrate on these special graphs. Indeed, any common period among all the elements of the sum, transfers to the sum itself.

For the following theorem, we need a careful analysis how an expression $\mathbf{s}(G; u)$ for some symmetry purified graph G is computed. We rely on the fact that, for fixed $s \in \mathbb{N}$, the function $m \mapsto \binom{m}{s} \bmod p$ is periodic with period p^k for each $k \in \mathbb{N}$ such that $p^k > s$ (see for instance [29, Proof of Fact 3.4]). Recall that a multinomial coefficient $\binom{n}{s_1,\dots,s_l}$ counts the number of ordered partition of n elements set into sequences of disjoint subsets of sizes $s_1, \dots, s_l$. More formally, for $s_1 + \cdots + s_l = n$ we have

$$\binom{n}{s_1, \dots, s_l} = \binom{s_1}{s_1} \cdot \binom{s_1 + s_2}{s_2} \cdot \ldots \cdot \binom{s_1 + \cdots + s_l}{s_l}. \tag{5}$$

▶ **Lemma 15.** *Let p be a prime. Let $s_1, \dots s_{l-1}$ be a sequence of integers. Let k be such that $p^k > s_1 + \cdots + s_{l-1}$. Then the function*

$$f(n) = \binom{n}{s_1, \dots, s_l(n)} \mod p$$

is periodic with period p^k where $s_l(n) = n - (s_1 + s_2 + \cdots + s_{l-1})$.

Proof. We use the formula (5) for computing multinomial coefficient. Note that the first $l - 1$ factors of the above product are constants, while the last one satisfies

$$\binom{s_1 + \cdots + s_l}{s_l(n)} = \binom{n}{s_l(n)} = \binom{n}{n - s_l(n)} = \binom{n}{s_1 + \cdots + s_{l-1}}.$$

So, periodicity of multinomial coefficient comes from periodicity of the binomial coefficients. ◀

▶ **Theorem 16.** *Let $p \neq q$ be prime numbers, $r \in \mathbb{F}_p$ and let G be an $\mathbb{F}_p$-labeled d-hypergraph on n vertices that is symmetry purified with respect to a maximal fully symmetric subset of vertices C of size $|C| > n/2$. Then $p^{k_p} \cdot q^{k_q}$ is a period of $\mathbf{s}(G; r)$ where k_p is the smallest integer satisfying $p^{k_p} > d$ and k_q is the smallest integer satisfying $q^{k_q} > n - |C|$.*

Note that, if $p^{k_p} \cdot q^{k_q} > n$, Theorem 16 establishes no non-trivial periods.

Proof. Note that the induced subgraph on the set C is a pseudo-clique. Moreover, as $|C| > \frac{|V|}{2}$, we can reconstruct the graph G up to isomorphism having the following information

1. the size l_C of the largest pseudo-clique C in G,
2. the type $\bar{t} = (t_1, \ldots, t_d) \in \mathbb{F}_p^d$ of the pseudo-clique, where t_i is the label of every i-ary edge in the pseudo-clique
3. sizes $l_0, l_1, \ldots, l_{p-1}$, where l_i is the number of vertices j that are not in the pseudo-clique C and have a unary edge with label $i = \lambda(\{j\})$. Vertices corresponding to label i in G will be denoted L_i (thus, $l_i = |L_i|$).

Using this characterization of a symmetry purified hypergraph $G = (V, \lambda)$, we obtain

▶ **Fact 17.** $\mathrm{Aut}(G) = \mathrm{Sym}(C) \times \mathrm{Sym}(L_0) \times \cdots \times \mathrm{Sym}(L_{p-1})$

In particular, we have at most $p+1$ orbits under the automorphism groups (note that we have less than $p+1$ orbits if some of the l_i are 0)

Now we evaluate $\mathbf{s}(G;r)$ on some integer m (and denote this by $\mathbf{s}(G;r)(m)$). In order to do it, pick $\bar{b}$ which has ones on the first m coordinates, i.e $\bar{b} = 1^m \cdot 0^{n-m}$. Hence,

$$\mathbf{s}(G;r)(m) = \mathbf{s}(G;r)(\bar{b}) = \sum_{i=1}^{k} \mathbf{b}(\pi_i(G);r)(\bar{b})$$

where, as before, $\pi_1, \ldots, \pi_k$ is a transversal of $\mathrm{Sym}(V)/\mathrm{Aut}(G)$. Each summand $\mathbf{b}(\pi_i(G);r)(\bar{b})$ evaluates to 1 or 0, depending on the mapping π_i. Being more precise, if π_i maps s_0 elements of L_0 to $[1..m]$, …, and s_{p-1} elements of L_{p-1} to $[1..m]$, then $\pi_i(G)(\bar{b})$ evaluates to

$$\sum_{j=1}^{d} t_j \cdot \binom{s_C(m)}{j} + \sum_{j=1}^{q-1} j \cdot s_j \pmod{p}. \tag{6}$$

where $s_C(m)$ denotes the number of elements of C mapped to $[1..m]$, which is computed according to formula $s_C(m) = m - (s_0 + \cdots + s_{p-1})$. Recall that $\mathbf{b}(\pi_i(G);r)(\bar{b}) = 1$ if and only if $\pi_i(G)(\bar{b}) = r$.

Let $\chi[G;r](m)$ denote the set of $\bar{s} = (s_0, \ldots, s_{p-1})$ that make the sum (6) evaluate to r. Observe that we have natural inequalities $0 \le s_i \le l_i$ for all $i \in \{0, \ldots, p-1\}$ and $0 \le s_C(m) \le l_C$. Hence, the feasible $\bar{s} \in \mathbb{N}^p$ can be described by

$$\bar{s} \in \chi[G;r](m) \iff \begin{cases} \sum_{j=1}^{d} t_j \cdot \binom{s_C(m)}{j} + \sum_{j=1}^{p-1} j \cdot s_j \pmod{p} = r \\ 0 \le s_i \le l_i \text{ for } i \in \{0, \ldots, p-1\} \\ 0 \le s_C(m) \le l_C. \end{cases} \tag{7}$$

Moreover, let $\#[\bar{s}](m)$ denote the number of permutations in $\{\pi_1, \ldots, \pi_k\}$ that map $s_C(m)$ elements of C to $[m]$ and s_i elements of L_i to $[m]$ (for $i = 0, \ldots, p-1$). Hence, we have

$$\mathbf{s}[G;r](m) = \left(\sum_{(\bar{s}) \in \chi[G;r](m)} \#[\bar{s}](m) \right) \pmod{q}.$$

Next, let us determine $\#[\bar{s}](m)$. Note that each permutation π_i in $\{\pi_1, \ldots, \pi_k\}$ maps each of the sets $L_0, \ldots, L_{p-1}, C$ to some subsets $L'_0, \ldots, L'_{p-1}, C' \subseteq [n]$ with $|L'_0| = |L_0|, \ldots, |L'_{p-1}| = |L_{p-1}|$, and $|C'| = |C|$. Now, if two mappings π_i, π_j output the same image $\pi_i L_0 = \pi_j L_0, \ldots, \pi_i L_{p-1} = \pi_j L_{p-1}, \pi_i C = \pi_j C$, then $\pi_i G = \pi_j G$ and hence π_i and

π_j must belong to the same coset of $\mathrm{Aut}(G)$ in $\mathrm{Sym}(V)$. Since $\{\pi_1, \ldots, \pi_k\}$ were chosen to be a transversal of $\mathrm{Sym}(V)/\mathrm{Aut}(G)$, this is only possible when $\pi_i = \pi_j$. So, the mapping $\pi_i \mapsto (L'_0, \ldots, L'_{p-1}, C')$ is injective.

In fact, the mapping is also surjective as for any particular $(L'_0, \ldots, L'_{p-1}, C')$ with $|L'_0| = |L_0|, \ldots, |L'_{p-1}| = |L_{p-1}|, |C'| = |C|$ we can find some $\pi \in \{\pi_1, \ldots, \pi_k\}$ which maps $\pi(L_i) = L'_i$ and $\pi(C) = C'$. Indeed, just pick any $\sigma \in \mathrm{Sym}(V)$ satisfying $\sigma(L_i) = L'_i$ for all i and $\sigma(C) = C'$ and take its representative in the equivalence class modulo $\mathrm{Aut}(G)$ as π_i. So we have a bijective mapping between permutations $(\pi_i)_{i=1..p-1}$ and ordered partitions $(L'_0, \ldots, L'_{p-1}, C')$ of $[n]$ satisfying $|L'_0| = |L_0|, \ldots, |L'_{p-1}| = |L_{p-1}|$, and $|C'| = |C|$.

Thus, if we want to count $\#(\bar{s})$, we need to count the number of proper partitions satisfying $|L'_1 \cap [1..m]| = s_1, \ldots, |L'_{p-1} \cap [1..m]| = s_{p-1}$, and $|C' \cap [1..m]| = s_C(m)$. In other words, we partition an m-element set into disjoint subsets of sizes $s_0, s_1, \ldots, s_{p-1}, s_C(m)$ and an $n-m$-element set into disjoint subsets of sizes $l_0 - s_0, \ldots, l_{p-1} - s_{p-1}, l_C - s_C(m)$; this leads to the following formula

$$\#[\bar{s}](m) = \binom{m}{s_0, \ldots, s_{p-1}, s_C(m)} \cdot \binom{n-m}{l_0 - s_0, \ldots, l_{p-1} - s_{p-1}, l_C - s_C(m)}. \tag{8}$$

Now, when fixing $s_0, \ldots, s_{p-1}$, we will use Lemma 15 to show that the function $m \mapsto \#[\bar{s}](m) \bmod q$ is periodic with period q^{k_q} in the interval $\{0, \ldots, n\}$ where k_q is the smallest integer such that $q^{k_q} > l_0 + \cdots + l_{p-1} = n - |C|$. The periodicity is immediate for the first element of the product by Lemma 15. Let $s'_i = l_0 - s_0$. One can see that the values of the second element of the product produce, in the interval $[0..n]$, a reversed sequence compared to the one produced by

$$\binom{m}{s'_0, \ldots, s'_{p-1}, m - (s'_0 + \cdots + s'_{p-1})}$$

Indeed, $l_C - s_C(m) = n - (l_0 + \cdots + l_{p-1}) - (m - (s_0 + \cdots + s_{p-1})) = (n-m) - (s'_0 + \cdots + s'_{p-1})$ and $n - m$ plays the role of m in the reversed sequence. As periodicity of any given sequence on the interval $[0..n]$ is preserved under reversing, and as $q^{k_q} > l_0 + \cdots + l_{p-1} \geq s'_0 + \cdots + s'_{p-1}$, we get the desired periodicity of $\#[\bar{s}](m)$ (as the period transfers to the product).

Now, one could argue that then the sum

$$\mathbf{s}[G;r](m) = \left(\sum_{(\bar{s}) \in \chi[G;r](m)} \#[\bar{s}](m) \right) \pmod{q}$$

must be periodic, as it is just a sum of elements that are periodic. Unfortunately, $\chi[G;r](m)$ selects elements of the sum depending on m (i.e. the s_i depend on m). We need to address this issue.

Note that in (7) we can drop the condition $0 \leq s_C(m) \leq l_C$ because whenever $s_C(m) < 0$ or $s_C(m) > l_C$ the formula (8) returns value 0 anyway (as multinomial coefficient takes value 0 whenever $s_0 + \cdots + s_{p-1} > m$ or $s'_0 + \cdots + s'_{p-1} > m$). So we can effectively get rid of $s_C(m)$ in χ to get an updated definition

$$\bar{s} \in \chi'[G;r](m) \iff \begin{cases} \sum_{j=1}^{d} t_j \cdot \binom{s_C(m)}{j} + \sum_{j=1}^{q-1} j \cdot s_j \pmod{p} = r \\ 0 \leq s_i \leq l_i \text{ for } i \in \{0, \ldots, p-1\} \end{cases} \tag{9}$$

and maintain the value of the sum, i.e.

$$\left(\sum_{(\bar{s}) \in \chi'[G;r](m)} \#[\bar{s}](m) \right) = \left(\sum_{(\bar{s}) \in \chi[G;r](m)} \#[\bar{s}](m) \right).$$

Let K be the smallest integer satisfying $p^K > \max(l_0 + l_1 + \cdots + l_{p-1}, d)$. We further modify the definition of χ' to create χ^* in the following way

$$\bar{s} \in \chi^*[G;r](m) \iff \begin{cases} \sum_{j=1}^{d} t_j \cdot \binom{s_C(m)+p^K}{j} + \sum_{j=1}^{q-1} j \cdot s_j \pmod{p} = r \\ 0 \le s_i \le l_i \text{ for } i \in \{0, \dots, p-1\} \end{cases} \tag{10}$$

We claim that for all m

$$\left(\sum_{(\bar{s}) \in \chi'[G;r](m)} \#[\bar{s}](m) \right) \pmod{q} = \left(\sum_{(\bar{s}) \in \chi*[G;r](m)} \#[\bar{s}](m) \right) \pmod{q}$$

There are 2 cases we need to consider.

1. When some fixed $\bar{s}$ belongs to both $\chi'[G;r](m)$ and $\chi^*[G;r](m)$, then $\#[s](m)$ cancels out from both sides of the equation. Similarly if $\bar{s}$ does not belong to either of the sets, we do not have $\#[s](m)$ on either of sides of the equation.
2. If $\bar{s} \in \chi'[G;r](m)$ and $\bar{s} \notin \chi^*[G;r](m)$ or $\bar{s} \notin \chi'[G;r](m)$ and $\bar{s} \in \chi^*[G;r](m)$, there must be some j such that $\binom{s_C(m)}{j} \neq \binom{s_C(m)+p^K}{j}$. This can only be the case if $s_C(m)$ is negative: otherwise, because $p^K > d \ge j$, from the periodicity of function $a \mapsto \binom{a}{j} \bmod p$ (for natural numbers $a \ge 0$), we would get that $\binom{s_C(m)}{j} = \binom{s_C(m)+p^K}{j} \bmod p$ and, hence, the conditions for $\chi'[G;r](m)$ and $\chi^*[G;r](m)$ would be identical from the perspective of $\bar{s}$. But when $s_C(m) < 0$, then $\#[s](m)$ is zero due to definition of multinomial coefficient, so it does not contribute to any of the sides anyway.

So we obtain that

$$\mathbf{s}(G;r)(m) = \left(\sum_{(\bar{s}) \in \chi^*[G;r](m)} \#[\bar{s}](m) \right) \pmod{q}.$$

But now the formula (10) gives ranges $0 \le s_j \le l_j$ for $j \in \{0, \dots, p-1\}$; thus, we conclude that $s_C(m) + p^K$ is always positive (since $s_C(m) + p^K = m + p^K - (s_0 + \cdots + s_{p-1}) \ge m \ge 0$). So $m \mapsto \binom{s_C(m)+p^K}{j} \bmod p$ is periodic with period $p^{k_p} > d$ where k_p is the smallest integer with $p^{k_p} > d$. Looking at the definition of χ^* we conclude:

► **Fact 18.** *Let $\bar{s} \in \mathbb{N}^p$. For $m \ge 0$ we have*

$$\bar{s} \in \chi^*[G;r](m) \iff \bar{s} \in \chi^*[G;r](m + p^{k_p})$$

Hence, the condition $\bar{s} \in \chi^*(G;r)$ depends not really on m, but on the remainder of m modulo p^{k_p}. So now, for all integers $j \in \{0, \dots, p^{k_p} - 1\}$ we define $\chi_j^*[G;r]$ as $\chi^*[G;r](j)$ in order to obtain that $\chi^*[G;r](m) = \chi_j^*[G;r]$ for $j = m \bmod p^{k_p}$. Now we can see that $\mathbf{s}(G;r)(m)$ is periodic with period $p^{k_p} \cdot q^{k_q}$:

$$\begin{aligned} \mathbf{s}(G;r)(m + p^{k_p} \cdot q^{k_q}) &= \sum_{(\bar{s}) \in \chi_j^*[G;r]} \#[\bar{s}](m + p^{k_p} \cdot q^{k_q}) \\ &= \sum_{(\bar{s}) \in \chi_j^*[G;r]} \#[\bar{s}](m) \\ &= \mathbf{s}(G;r)(m) \pmod{q}. \end{aligned}$$

The first and the third equality comes from periodicity of the condition χ^* and the fact that m and $m + p^{k_p} \cdot q^{k_q}$ give the same rest modulo p^{k_p} and the second one comes from the equality of rests modulo q^{k_q} and periodicity of $\#(s)(m)$. ◄

4.4 Main Theorems

Now we have all the necessary components to prove our main theorems.

Proof of Theorem 1. As discussed in Section 3, any $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit has a corresponding symmetric $\Sigma_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ expression $\mathbf{f}$ with no periods smaller than the circuit we started with (note that there can happen some blow-up in size, but this does not matter as we argue below). By Fact 7, $\mathbf{f}$ can be written as a sum of expressions of the form $\mathbf{s}(G; r)$. Hence, from now on let us consider one of these expressions $\mathbf{s}(G; r)$.

We choose ε such that $2s = 2^{\varepsilon \cdot n}$ (meaning that $\varepsilon \cdot n = \log s + 1$ and $\varepsilon < 1/8$). If G does not contain a fully symmetric set $|C|$ of size at least $n - \lfloor \varepsilon n \rfloor$, by Lemma 10, it satisfies $|\mathrm{Sym}([n])/\operatorname{Aut}(G)| \geq 2^{\lfloor \varepsilon n \rfloor}$. Thus, writing $\mathbf{s}(G; r) = \sum_{i=0}^{k} \mathbf{b}^{\pi_i}(G, r)$ as in Equation (4), it follows that $k \geq 2^{\lfloor \varepsilon n \rfloor}$. As all the different terms in this sum get their inputs from different graphs $\pi_i(G)$, also for each term in the sum there must have been a different gate in the original circuit we started with. This is a contradiction as $2^{\lfloor \varepsilon n \rfloor} > s$.

Hence, all the subexpressions $\mathbf{s}(G; r)$ contain a fully symmetric set C of size at least $n - \lfloor \varepsilon n \rfloor$. Now, Lemma 14 tells us that we can write $\mathbf{f}$ as a sum of expressions of the form $\mathbf{s}(G; r)$ where G is symmetry-purified with respect to C. Then, Theorem 16 implies that each such $\mathbf{s}(G; r)$ has a period $p^{k_p} \cdot q^{k_q}$, where k_p is the smallest integer such that $p^{k_p} > d$ and k_q is the smallest integer such that $q^{k_q} > n - |C| = \lfloor \varepsilon n \rfloor$. As $\mathbf{f}$ is a sum of different $\mathbf{s}(G; r)$, which all share the period $p^{k_p} \cdot q^{k_q}$, it itself has period $p^{k_p} \cdot q^{k_q}$. ◀

The following result shows that the estimate of size which can be derived from Theorem 1 is asymptotically (almost) precise.

▶ **Proposition 19.** *Let $p \neq q$ be primes. Let k_p, k_q be natural numbers. For every symmetric function f with a period $p^{k_p} \cdot q^{k_q}$ there is a symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit of size $O(d \cdot \binom{n}{d} + d \cdot l^2 \cdot \binom{n}{l})$ which computes f, where $d = p^{k_p} - 1$ and $l = q^{k_q} - 1$ and $d, l < \frac{n}{2}$.*

The proof, which is a rather straightforward application of known facts, can be found in the full version on arXiv.

▶ **Corollary 20.** *Let $p \neq q$ be primes and $d : \mathbb{N} \to \mathbb{N}$ with $d(n) \leq n/2$ for all n. A function $f = (f_n)_{n \in \mathbb{N}}$ (with $f_n : \{0,1\}^n \to \{0,1\}$) can be computed by a family of symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_{d(n)}$ circuits of quasipolynomial size if and only if, for each n, f_n has a period $p^{k_p(n)} q^{k_q(n)} \in \log^{O(1)}(n)$ for some functions $k_p, k_q : \mathbb{N} \to \mathbb{N}$.*

Moreover, if $d = p^{k_p} - 1$ is a constant, then f can be computed by symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits of quasipolynomial size if and only if, for each n, f_n has a period $p^{k_p} q^{k_q(n)} \in \log^{O(1)}(n)$ for some function $k_q : \mathbb{N} \to \mathbb{N}$.

Proof. If f_n has period $p^{k_p(n)} q^{k_q(n)} \in \log^{O(1)}(n)$, by Proposition 19, f_n is computed by a $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit of size $O(d \cdot \binom{n}{d} + d \cdot l^2 \cdot \binom{n}{l})$ where $l = q^{k_q}$. As $\binom{n}{\log^{O(1)}(n)} \subseteq 2^{\log^{O(1)}(n)}$, it follows that f can be computed by a family of quasipolynomial-size $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_{d(n)}$ circuits. In the case that $d \in \{p^k - 1 \mid k \in \mathbb{N}\}$ is a constant, Proposition 19 still can be applied with the same outcome.

On the other hand, if f is computed by a family of quasipolynomial-size $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_{d(n)}$ circuits, we first observe that without loss of generality $d(n) \in \log^{O(1)}(n)$. Indeed, if for some n the circuits use an AND_d gate for some $d \geq \log^k(n)$, then the size of the circuit, because of the symmetry property, is at least $\binom{n}{d} \geq (\frac{n}{d})^d \geq 2^d \geq 2^{\log^k(n)}$. If this holds for every $k \in \mathbb{N}$ and infinitely many n, then the circuit family is not of quasipolynomially bounded size.

Now, it remains to apply Theorem 1, which tells us that f_n has period $p^{k_p}q^{k_q}$ where k_p is the smallest integer with $p^{k_p} > d(n)$ and k_q is the smallest integer with $q^{k_q} > \log s(n) + 1$ where $s(n) \in 2^{\log^{O(1)}(n)}$ is the size of the n-input circuit. As $p^{k_p}q^{k_q} \in \log^{O(1)}(n)$, both parts of the corollary follow (for the second part, observe that p^{k_p} is the smallest p-power greater than d). ◀

Instead of directly proving Theorem 3, let us derive the following slightly more explicit and general variant of the theorem:

▶ **Theorem 21.** *Let $p \neq q$ be primes and let $n \geq \max\{13, 4p^2q^2\}$ and $d \leq n - \sqrt{n}$. Then every symmetric $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit computing the AND_n function has size at least $2^{\max\{n/(2dpq), \sqrt{n}\}}$.*

Proof. Let us write $V = \{x_1, \ldots, x_n\}$. First consider the case that $d \geq \sqrt{n}$ and there is actually an AND_k gate v with $n - \sqrt{n} \geq k \geq \sqrt{n}$ inputs. Since for any $\pi \in \mathrm{Sym}(V)$ also $\pi(v)$ must be a gate in the circuit, we obtain different AND_k gates for each k-element subset of V. As there are $\binom{n}{k} \geq \max\{(n/k)^k, (n/(n-k))^{n-k}\} \geq 2^{\sqrt{n}} \geq 2^{n/d}$ many such subsets, the theorem holds in this case (almost trivially).

Therefore, in the following, we assume $d < \sqrt{n} \leq n/(2pq)$ and consider an arbitrary n-input $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuit $\mathcal{C}$ of size $s \leq 2^{n/(2pqd)}$. By Theorem 1, the function computed by $\mathcal{C}$ has period $p^{k_p}q^{k_q}$ where k_p is the smallest integer with $p^{k_p} > d$ and k_q is the smallest integer with $q^{k_q} > \log s + 1 \geq n/(2pqd) + 1$. Notice that $p^{k_p} \leq d \cdot p$ and $q^{k_q} \leq (n/(2pqd) + 1) \cdot q$ and, hence, we have

$$p^{k_p} \cdot q^{k_q} \leq d \cdot p \cdot (n/(2pqd) + 1) \cdot q = n/2 + dpq < n.$$

Thus, $\mathcal{C}$ does not compute the AND_n function as AND_n does not have any non-trivial period. ◀

5 Further Perspectives

Arguably one of the strongest applications of the Degree Decreasing Lemma is Theorem 4 in [22]. It implies that, if all the polynomials over $\mathbb{F}_p$ that compose the top levels of a $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$-circuit can be written with a sublinear number of (binary) multiplications, then the circuit can be replaced with a $\mathrm{MOD}_q \circ \mathrm{MOD}_p$ circuit with only a subexponential blow-up in size. We argue that this kind of theorem cannot be applied in the context of our proof.

Note that a large pseudo-clique in the symmetry-purified expressions are (arbitrary) symmetric polynomials. Most of symmetric polynomials over $\mathbb{F}_p$ require at least a linear number of multiplications in any formula (circuit) defining them. To see it, consider the example $p(\overline{x}) = \sum_{i<j} x_i \cdot x_j$ as a polynomial over $\mathbb{F}_2$. One can easily check that it represents a function with smallest period 4. But now, if it could be written with a sub-linear number of multiplications, by Theorem 4 in [22], it could be represented by a sub-exponential size $\mathrm{MOD}_3 \circ \mathrm{MOD}_2$ circuit. However, this contradicts [23, Theorem 2.4] as subexponential size $\mathrm{MOD}_3 \circ \mathrm{MOD}_2$ circuits can only represent periodic functions with period of the form $2 \cdot 3^k$. This shows that that the Degree Decreasing Lemma cannot be used in this context, as it puts the limitations on its own applicability, by providing arithmetic circuit lower bounds. Such lower bounds can be proved for all non-trivial symmetric polynomials over $\mathbb{F}_p$ with $d \geq p$ using a similar period analysis. Thus, our symmetry purification technique as well as combinatorial analysis contained in the proof of Theorem 16 constitute a substantial improvement over the Degree Decreasing Lemma and its accompanying techniques.

The results of the present paper indicate what type of symmetric functions might be computable by small, but not necessarily symmetric, $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits. It is natural to believe that the optimal (or nearly optimal) representation of the symmetric function should also be symmetric. Thus, we state the following

▶ **Conjecture 22.** *For fixed d, p, q, the only symmetric functions that can be represented by $\mathrm{MOD}_q \circ \mathrm{MOD}_p \circ \mathrm{AND}_d$ circuits of subexponential size have to be periodic with some period of the form $p^{k_p} \cdot q^{k_q}$, for k_p being the smallest integer with $p^{k_p} \geq d$ and k_q be such that $p^{k_p} \cdot q^{k_q} \leq n$.*

References

1 Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983. doi:10.1016/0168-0072(83)90038-6.

2 László Babai. Primitive coherent configurations and the order of uniprimitive permutation groups, 2018. URL: https://people.cs.uchicago.edu/~laci/papers/uni-update.pdf.

3 David A. Mix Barrington, Richard Beigel, and Steven Rudich. Representing Boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4:367–382, 1994. doi:10.1007/BF01263424.

4 David A. Mix Barrington and Howard Straubing. Complex polynomials and circuit lower bounds for modular counting. *Computational Complexity*, 4:325–338, 1994. doi:10.1007/BF01263421.

5 David A. Mix Barrington, Howard Straubing, and Denis Thérien. Non-uniform automata over groups. *Information and Computation*, 89(2):109–132, 1990. doi:10.1016/0890-5401(90)90007-5.

6 Abhishek Bhrushundi, Kaave Hosseini, Shachar Lovett, and Sankeerth Rao. Torus polynomials: An algebraic approach to ACC lower bounds. In *10th Innovations in Theoretical Computer Science Conference, ITCS 2019*, volume 124 of *LIPIcs*, pages 13:1–13:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ITCS.2019.13.

7 Alfred Bochert. Ueber die Zahl der verschiedenen Werthe, die eine Function gegebener Buchstaben durch Vertauschung derselben erlangen kann. *Mathematische Annalen*, 33(4):584–590, 1889. doi:10.1007/BF01444035.

8 Joshua Brakensiek, Sivakanth Gopi, and Venkatesan Guruswami. Constraint satisfaction problems with global modular constraints: Algorithms and hardness via polynomial representations. *SIAM Journal on Computing*, 51(3):577–626, 2022. doi:10.1137/19m1291054.

9 Bettina Brustmann and Ingo Wegener. The complexity of symmetric functions in bounded-depth circuits. *Information Processing Letters*, 25(4):217–219, 1987. doi:10.1016/0020-0190(87)90163-3.

10 Peter J. Cameron. *Permutation Groups.* London Mathematical Society Student Texts. Cambridge University Press, 1999. doi:10.1017/CBO9780511623677.

11 Brynmor Chapman and Ryan Williams. Smaller ACC^0 circuits for symmetric functions. In *13th Innovations in Theoretical Computer Science Conference, ITCS 2022*, volume 215 of *LIPIcs*, pages 38:1–38:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.ITCS.2022.38.

12 Arkadev Chattopadhyay, Navin Goyal, Pavel Pudlak, and Denis Therien. Lower bounds for circuits with MOD_m gates. In *47th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2006*, pages 709–718, 2006. doi:10.1109/FOCS.2006.46.

13 Anuj Dawar and Gregory Wilsenach. Symmetric arithmetic circuits. In *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020*, volume 168 of *LIPIcs*, pages 36:1–36:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICALP.2020.36.

14 Larry Denenberg, Yuri Gurevich, and Saharon Shelah. Definability by constant-depth polynomial-size circuits. *Information and Control*, 70(2/3):216–240, 1986. doi:10.1016/S0019-9958(86)80006-7.

15 Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011. doi:10.1137/100804322.

16 Zeev Dvir and Sivakanth Gopi. 2-server PIR with sub-polynomial communication. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*, pages 577–584. ACM, 2015. doi:10.1145/2746539.2746546.

17 Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM Journal on Computing*, 41(6):1694–1703, 2012. doi:10.1137/090772721.

18 Ronald Fagin, Maria M. Klawe, Nicholas Pippenger, and Larry J. Stockmeyer. Bounded-depth, polynomial-size circuits for symmetric functions. *Theoretical Computer Science*, 36:239–250, 1985. doi:10.1016/0304-3975(85)90045-3.

19 Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical systems theory*, 17:13–27, 1984. doi:10.1007/BF01744431.

20 Parikshit Gopalan. Constructing Ramsey graphs from Boolean function representations. *Combinatorica*, 34:173–206, 2014. doi:10.1007/s00493-014-2367-1.

21 Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20(1):71–86, 2000. doi:10.1007/s004930070032.

22 Vince Grolmusz. A degree-decreasing lemma for (MOD_p-MOD_m) circuits. *Discrete Mathematics and Theoretical Computer Science*, 4(2):247–254, 2001. doi:10.46298/dmtcs.289.

23 Vince Grolmusz and Gábor Tardos. Lower bounds for (MOD_p-MOD_m) circuits. *SIAM Journal on Computing*, 29(4):1209–1222, 2000. doi:10.1137/S0097539798340850.

24 Kristoffer Arnsfelt Hansen. Computing symmetric boolean functions by circuits with few exact threshold gates. In *Computing and Combinatorics, 13th Annual International Conference, COCOON 2007, Proceedings*, volume 4598 of *Lecture Notes in Computer Science*, pages 448–458. Springer, 2007. doi:10.1007/978-3-540-73545-8_44.

25 Kristoffer Arnsfelt Hansen and Michal Koucký. A new characterization of ACC^0 and probabilistic CC^0. *Computational Complexity*, 19(2):211–234, 2010. doi:10.1007/s00037-010-0287-z.

26 Johan Håstad. *Computational limitations for small depth circuits.* PhD thesis, Massachusetts Institute of Technology, 1986.

27 William He and Benjamin Rossman. Symmetric formulas for products of permutations. In *14th Innovations in Theoretical Computer Science Conference, ITCS 2023*, volume 251 of *LIPIcs*, pages 68:1–68:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ITCS.2023.68.

28 Paweł M. Idziak, Piotr Kawałek, and Jacek Krzaczkowski. Intermediate problems in modular circuits satisfiability. In *Proceedings of 35th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2020*, pages 578–590, 2020. doi:10.1145/3373718.3394780.

29 Pawel M. Idziak, Piotr Kawalek, and Jacek Krzaczkowski. Complexity of modular circuits. In *Proceedings of 37th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2022*, pages 32:1–32:11, 2022. doi:10.1145/3531130.3533350.

30 Paweł M. Idziak, Piotr Kawałek, Jacek Krzaczkowski, and Armin Weiß. Satisfiability Problems for Finite Groups. In *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022*, volume 229 of *LIPIcs*, pages 127:1–127:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.ICALP.2022.127.

31 Pawel M. Idziak and Jacek Krzaczkowski. Satisfiability in multivalued circuits. *SIAM Journal on Computing*, 51(3):337–378, 2022. doi:10.1137/18m1220194.

32 Piotr Kawalek and Armin Weiß. Violating constant degree hypothesis requires breaking symmetry. *CoRR*, abs/2311.17440, 2023. doi:10.48550/arXiv.2311.17440.

33 Tianren Liu and Vinod Vaikuntanathan. Breaking the circuit-size barrier in secret sharing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018*, pages 699–708, 2018. doi:10.1145/3188745.3188936.

34 Chi-Jen Lu. An exact characterization of symmetric functions in qAC^0[2]. *Theoretical Computer Science*, 261(2):297–303, 2001. doi:10.1016/S0304-3975(00)00145-6.

35 Cheryl E. Praeger and Jan Saxl. On the orders of primitive permutation groups. *The Bulletin of the London Mathematical Society*, 12(4):303–307, 1980. doi:10.1112/blms/12.4.303.

36 Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, STOC 1987*, pages 77–82. ACM, 1987. doi:10.1145/28395.28404.

37 Howard Straubing and Denis Thérien. A note on MOD$_p$-MOD$_m$ circuits. *Theory of Computing Systems*, 39(5):699–706, 2006.

38 Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles (preliminary version). In *26th Annual Symposium on Foundations of Computer Science, FOCS 1985*, pages 1–10. IEEE Computer Society, 1985. doi:10.1109/SFCS.1985.49.

39 Zhi-Li Zhang, David A. Mix Barrington, and Jun Tarui. Computing symmetric functions with AND/OR circuits and a single MAJORITY gate. In *Proceedings of 10th Annual Symposium on Theoretical Aspects of Computer Science, STACS 1993*, volume 665 of *Lecture Notes in Computer Science*, pages 535–544. Springer, 1993. doi:10.1007/3-540-56503-5_53.