

# Intermediate problems in modular circuits satisfiability

Paweł M. Idziak  
Faculty of Mathematics and  
Computer Science,  
Department of Theoretical Computer  
Science  
Jagiellonian University  
Kraków, Poland  
idziak@tcs.uj.edu.pl

Piotr Kawalek  
Faculty of Mathematics and  
Computer Science,  
Department of Theoretical Computer  
Science  
Jagiellonian University  
Kraków, Poland  
piotr.kawalek@doctoral.uj.edu.pl

Jacek Krzaczkowski  
Faculty of Mathematics, Physics and  
Computer Science,  
Institute of Computer Science  
Maria Curie-Skłodowska University  
Lublin, Poland  
krzacz@poczta.umcs.lublin.pl

## Abstract

In [15] a generalization of Boolean circuits to arbitrary finite algebras had been introduced and applied to sketch P versus NP-complete borderline for circuits satisfiability over algebras from congruence modular varieties. However the problem for nilpotent (which had not been shown to be NP-hard) but not supernilpotent algebras (which had been shown to be polynomial time) remained open.

In this paper we provide a broad class of examples, lying in this grey area, and show that, under the Exponential Time Hypothesis and Strong Exponential Size Hypothesis (saying that Boolean circuits need exponentially many modular counting gates to produce boolean conjunctions of any arity), satisfiability over these algebras have intermediate complexity between  $\Omega(2^{c \log^{h-1} n})$  and  $O(2^{c \log^h n})$ , where  $h$  measures how much a nilpotent algebra fails to be supernilpotent. We also sketch how these examples could be used as paradigms to fill the nilpotent versus supernilpotent gap in general.

Our examples are striking in view of the natural strong connections between circuits satisfiability and Constraint Satisfaction Problem for which the dichotomy had been shown by Bulatov [4] and Zhuk [28].

**CCS Concepts:** • Theory of computation → Complexity theory and logic; Problems, reductions and completeness; Circuit complexity; Complexity classes; • Mathematics of computing → Combinatorial algorithms.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*LICS '20, July 8–11, 2020, Saarbrücken, Germany*  
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7104-9/20/07...\$15.00  
<https://doi.org/10.1145/3373718.3394780>

**Keywords:** circuit satisfiability, intermediate problems, solving equations, constraint satisfaction problem

## ACM Reference Format:

Paweł M. Idziak, Piotr Kawalek, and Jacek Krzaczkowski. 2020. Intermediate problems in modular circuits satisfiability. In *Proceedings of the 35th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS '20)*, July 8–11, 2020, Saarbrücken, Germany. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3373718.3394780>

## 1 Introduction

In [15] a generalization of Boolean circuits to multi-valued ones had been introduced. This concept was formalized by defining circuits over arbitrary finite algebra  $A$ . Then the computational complexity of the following problems was considered:

- $\text{CSAT}(A)$  – circuits satisfiability over the algebra  $A$ ,
- $\text{SCSAT}(A)$  – simultaneous satisfiability of a set of circuits over the algebra  $A$ ,
- $\text{CEQV}(A)$  – circuits equivalence over the algebra  $A$ .

This has been done by treating the (basic) gates of the circuits as fundamental operations of the corresponding algebra  $A$ , while the universe  $A$  of this algebra consists of the possible values on inputs and output of the gates. Such translation has been shown to preserve the complexity when passing respectively between

- $\text{CSAT}(A)$  and deciding if an equation over  $A$  has a solution,
- $\text{SCSAT}(A)$  and deciding if a system of equations over  $A$  has a solution,
- $\text{CEQV}(A)$  and deciding if two polynomials determine the same function over  $A$ ,

but with the possibility of endowing algebra  $A$  with finitely many additional operations that are already definable in  $A$ . Such (finite) expansions allows to concentrate on the algebraic structure of the considered algebras in order to classify them with respect to computational complexity of the above problems. Making the algebra independent of its basic operations is crucial, as for example equation solving over the group  $S_3$  can be done in P, while for the same group

endowed with definable operation resembling binary commutator  $[x, y] = x^{-1}y^{-1}xy$  the very same problem became NP-complete. Actually [15] presents an attempt to such classification for a very broad class of algebras covering most of the ones considered in mathematics and computer science, like groups, rings, modules, lattices, Heyting algebras and many other algebras arising from logic. The restriction put for those algebras was that they have to belong to congruence modular varieties. This assumption made it possible to use advanced tools of universal algebras that work in such a setting. Under this additional assumption it has been shown that if an algebra  $A$  fails to decompose nicely, i.e. into a direct product of a nilpotent algebra and an algebra that essentially is a subreduct of a distributive lattice then  $\text{CSAT}$  for  $A$  (or at least one of its quotients) is NP-complete. And, almost conversely, if  $A$  does decompose nicely (in the above sense), but with the additional assumption that the nilpotent factor is actually supernilpotent, then  $\text{CSAT}(A)$  is P. Very similar statements hold for  $\text{CEQV}(A)$ , but the ‘lattice’ factor disappears here as  $\text{CEQV}$  for distributive lattices is NP-complete.

Although the problems  $\text{CSAT}$  or  $\text{SCSAT}$  resemble Constraint Satisfaction Problem, there are some subtle differences here, so that the CSP dichotomy shown by Bulatov [4] and Zhuk [28] can not be used for  $\text{CSAT}$ . As it was noticed in [15] the problems  $\text{SCSAT}$  and  $\text{CSP}$  can be bisimulated in polynomial time, namely each finite algebra  $A$  can be transformed into a finite relational structure  $\mathbb{D}$ , and each finite relational structure  $\mathbb{D}$  can be translated into a finite algebra  $A$  so that the problems  $\text{SCSAT}(A)$  and  $\text{CSP}(\mathbb{D})$  are equivalent. Surprisingly for a single circuit/equation such kind of translation works only when going from relational structures to algebras. In fact, under some complexity hypothesis (like for example ETH) this paper shows that the other way is blocked. As we will see, this is due to the hardness of incorporating arbitrary long conjunction (between constraints) by translating them into relatively short polynomials of an algebra.

As it is easily seen the nice pre-characterization of algebras with  $\text{CSAT}/\text{CEQV}$  solvable in polynomial time leaves the nilpotent but not supernilpotent gap which is unsolved. The essential difference between this two concepts of nilpotency lies in fact that in supernilpotent algebras there is an absolute bound for the arity of expressible (by polynomials) conjunction. In nilpotent (but not supernilpotent) algebras conjunction-like polynomials of arbitrary arity  $n$  do always exist but the known ones are too long to be used to polynomially code NP-complete problems in  $\text{CSAT}$ . In section 2 we split nilpotent algebras into slices that will correspond to the measure how much a nilpotent algebra fails to be supernilpotent. This distance  $h$  is determined by the behavior of a multi-ary commutator operation on congruences of  $A$  which is used to define  $h$ -step supernilpotent algebras. On the other hand we show that it strictly corresponds to the

longest chain of alternating primes hidden in the algebra. Then, we start with any sequence  $p_1 \neq p_2 \neq \dots \neq p_h$  of primes with  $h \geq 2$  and construct an example of the simplest algebra  $\mathbb{D}[p_1, \dots, p_h]$  that is  $h$ -step (but not  $(h-1)$ -step) supernilpotent to demonstrate how to construct an  $n$ -ary conjunction polynomial  $\text{AND}_n$  of size  $O(2^{cn^{1/(h-1)}})$ . This together with the assumption of Exponential Time Hypothesis will be used to show the following theorem

**Theorem 1.1.** *The complexity for  $\text{CSAT}(\mathbb{D}[p_1, \dots, p_h])$  and  $\text{CEQV}(\mathbb{D}[p_1, \dots, p_h])$  is at least  $\Omega(2^{c \cdot \log^{h-1} |\Gamma|})$ , where  $|\Gamma|$  is the size of a circuit  $\Gamma$  on the input (unless ETH fails).*

Obviously these lower bounds would be even higher if one finds shorter conjunction terms. Thus an upper bound for the complexity of  $\text{CSAT}$  relies on the (necessary) length of polynomials that are able to express  $\text{AND}_n$ . A kind of such lower bound had been already introduced as a conjecture by Barrington, Straubing and Thérien [3] in their study of non-uniform automata over groups. To reword their conjecture for our purposes recall that a counting gate  $\text{MOD}_m^R$  (with unbounded fan-in) returns 1 if all the 1’s on the input sum up modulo  $m$  to an element in  $R$ , and 0 otherwise. Moreover recall that  $\text{CC}[m]$ -circuit is build up with  $\text{MOD}_m^R$ -gates only. In this language the conjecture says that:

- the sizes of  $\text{CC}[m]$ -circuits  $(\Gamma_n)_n$  with bounded depth computing  $(\text{AND}_n)_n$  grow exponentially in  $n$ .

Very recently Kompatscher [21] has used this conjecture to show that

- for every nilpotent algebra  $A$  from a congruence modular variety  $\text{CSAT}(A)$  and  $\text{CEQV}(A)$  can be solved in quasi polynomial time  $O(2^{c \log^t m})$ , for some constants  $c, t$  depending on  $A$ .

In our study of nilpotent algebras we had noticed that the exponent  $t$  in Kompatscher’s bound is strongly correlated with  $h$ -step supernilpotency (or in other words the depth of corresponding circuits). This is now confirmed by Theorem 1.1, so that the above hypothesis has to be weakened accordingly.

A promising weaker version of such a hypothesis might be:

- the sizes of  $\text{CC}[m]$ -circuits  $(\Gamma_n)_n$ , of depth bounded by  $h > 1$ , that compute  $(\text{AND}_n)_n$ , grow at least as  $\Omega(2^{cn^{1/(h-1)}})$ .

A dual version of the above hypothesis, namely that to build  $\text{MOD}_m$  gates (of arbitrary large arity  $n$ ) by circuits of bounded depth needs superpolynomial (in  $n$ ) number of classical Boolean gates AND, OR (of unbounded fan in) and NOT, has been used by Furst, Sax and Sipser [8, 24] to separate PSPACE from polynomial hierarchy by oracles. Later on Yao [27] confirmed this dual hypothesis, while Håstad [11] has shown that for  $\text{MOD}_2$ -gates the required sizes are even exponential.

Unfortunately the hypothesis that  $CC[m]$ -circuits of depth  $h > 1$  require  $\Omega(2^{cn^{1/(h-1)}})$  gates to express  $(AND_n)_n$  is blocked by Barrington, Beigel and Rudich in [2]. They use integers  $m$  that have  $r \geq 2$  different prime factors to construct  $CC[m]$ -circuits of depth 3 that compute  $(AND_n)_n$  using only  $2^{O(n^{1/r} \log n)}$  gates. Such relatively small circuits are possible to built by exploring the interaction of  $r$  different primes on the very same level of the circuits. However in our setting of the algebras  $D[p_1, \dots, p_h]$  there is only one prime at each level so that it suffices to use  $CC[p_1, \dots, p_h]$ -circuits, i.e.  $CC$ -circuits where on the  $i$ -th level there are only  $MOD_{p_i}$  gates, with  $p_i$  being prime. Note here that by our definition  $CC[p_1, \dots, p_h]$ -circuits have depth  $h$ . Thus the hypothesis we will build our upper bounds is the following Strong Exponential Size Hypothesis (SESH).

**Conjecture (SESH).** *The sizes of  $CC[p_1, \dots, p_h]$ -circuits  $(\Gamma_n)_n$ , of depth  $h > 1$ , that compute  $(AND_n)_n$ , grow at least as  $\Omega(2^{cn^{1/(h-1)}})$ .*

Now, with the help of SESH we can show the upper bound for our problems that almost matches the lower bound of Theorem 1.1.

**Theorem 1.2.** *There are deterministic algorithms solving  $CSAT(D[p_1, \dots, p_h])$  and  $CEQV(D[p_1, \dots, p_h])$  in  $O(2^{c \log^h |\Gamma|})$  time, where  $|\Gamma|$  is the size of a circuit  $\Gamma$  on the input (unless SESH fails).*

Relaxing deterministic realm to a probabilistic one we can match the lower bound much better.

**Theorem 1.3.** *There are probabilistic algorithms solving  $CSAT(D[p_1, \dots, p_h])$  and  $CEQV(D[p_1, \dots, p_h])$  in time  $O(2^{c \log^{h-1} |\Gamma|})$ , where  $|\Gamma|$  is the size of a circuit  $\Gamma$  on the input (unless SESH fails).*

Note here that all the above theorems give interesting bounds only for  $h \geq 3$ . In fact [16] gives a polynomial upper bound for  $D[p, q]$  with  $p \neq q$ . However we decided to keep  $h = 2$  in our theorems, as their proofs give a nice insight into the structure of the corresponding algebras.

Unfortunately this inside is still not deep enough to be generalized to 2-step supernilpotent algebras. Both  $CSAT$  and  $CEQV$  remain open here.

On the other hand with  $h \geq 3$  our results show that the dichotomy conjecture (similar to the one for Constraint Satisfaction Problem) is unlikely to be confirmed. It may even happen that  $CSAT$  for these algebras would provide a natural example of an intermediate problem.

Our choice of the algebras  $D[p_1, \dots, p_h]$  has been done very carefully so that the proof of the above lower and upper bounds demonstrate the main general idea but are still readable enough. The important ingredient here is that the primes involved do alternate, i.e.  $p_1 \neq p_2 \neq \dots \neq p_h$ . This corresponds to the fact that all the  $AND_n$ 's can be obtained from  $MOD_m$  gates only if  $m$  is not the power of a prime.

We decided to stay with our argument for this particular family of algebras although most of the ideas used here can be generalized to  $h$ -step supernilpotent realm. In fact in Section 2 we show why alternation of primes is important in the study of the expressive power of definable polynomials. Unfortunately the arguments in a general setting have to be terribly involved and make a heavy use of tame congruence theory [12] and modular commutator theory [6]. A reader that is not experienced enough with universal algebraic tools may skip Section 2 and go directly to Section 3 where the main results are shown. In Section 4 we apply our methods to the special case of the symmetric group  $S_4$  (but considered in its pure group language, without a possibility of endowing it by definable operations). We do that as for many years the complexity of equation solving over this group has been unsettled. In view of the fact that this problem is polynomial time for  $S_3$  (while with NP-complete for  $CSAT(S_3)$ ) there has been a hope that the same holds for  $S_4$ . Now, under the assumption of ETH we destroy this hope.

## 2 Stratification of algebras

Our study of  $CSAT$  for nilpotent algebras relies on the observation that there is a very strong connection between the depth  $h$  of the  $CC^0$ -circuits and stratification of such algebras into  $h$ -step supernilpotent slices. To define this stratification we start with recalling the concept of commutator.

If  $\alpha, \beta, \gamma$  are congruences of an algebra then we say that  $\alpha$  centralizes  $\beta$  modulo  $\gamma$ , denoted  $C(\alpha, \beta; \gamma)$ , if for every  $n \geq 1$ , every  $(n+1)$ -ary term  $t$ , every  $(a, b) \in \alpha$ , and every  $(c_1, d_1), \dots, (c_n, d_n) \in \beta$  we have

$$t(a, \bar{c}) \stackrel{\gamma}{\equiv} t(a, \bar{d}) \text{ iff } t(b, \bar{c}) \stackrel{\gamma}{\equiv} t(b, \bar{d}).$$

Obviously among all congruences  $\gamma$  such that  $C(\alpha, \beta; \gamma)$  there is the smallest one and it is denoted by  $[\alpha, \beta]$  and called the commutator of  $\alpha$  and  $\beta$ .

By means of the commutator it is possible to define notions of abelianity, solvability and nilpotency for arbitrary algebras. First, for a congruence  $\theta$  and  $i = 1, 2, \dots$  we put

$$\begin{aligned} \theta^{(0)} &= \theta & \theta^{[0]} &= \theta \\ \theta^{(i+1)} &= [\theta, \theta^{(i)}] & \theta^{[i+1]} &= [\theta^{[i]}, \theta^{[i]}]. \end{aligned}$$

Now, a congruence  $\theta$  of  $\mathbf{A}$  is called  $k$ -nilpotent [or  $k$ -solvable] if  $\theta^{(k)} = 0_{\mathbf{A}}$  [ $\theta^{[k]} = 0_{\mathbf{A}}$ ] and the algebra  $\mathbf{A}$  is nilpotent [solvable] if  $1_{\mathbf{A}}$  is  $k$ -nilpotent [ $k$ -solvable] for some finite  $k$ .

A more detailed discussions of the generalized commutator may be found in [6, 12, 23].

The concept of centrality and of the binary commutator has a natural generalization. Namely, for a bunch of congruences  $\alpha_1, \dots, \alpha_k, \beta, \gamma \in \text{Con } \mathbf{A}$  we say that  $\alpha_1, \dots, \alpha_k$  centralize  $\beta$  modulo  $\gamma$ , and write  $C(\alpha_1, \dots, \alpha_k, \beta; \gamma)$ , if for all polynomials  $f \in \text{Pol } \mathbf{A}$  and all tuples  $\bar{a}_1 \stackrel{\alpha_1}{\equiv} \bar{b}_1, \dots, \bar{a}_k \stackrel{\alpha_k}{\equiv} \bar{b}_k$

and  $\bar{u} \stackrel{\beta}{\equiv} \bar{v}$  such that

$$\mathbf{f}(\bar{x}_1, \dots, \bar{x}_k, \bar{u}) \stackrel{\gamma}{\equiv} \mathbf{f}(\bar{x}_1, \dots, \bar{x}_k, \bar{v})$$

for all possible choices of  $(\bar{x}_1, \dots, \bar{x}_k)$  in  $\{\bar{a}_1, \bar{b}_1\} \times \dots \times \{\bar{a}_k, \bar{b}_k\}$  but  $(\bar{b}_1, \dots, \bar{b}_k)$ , we also have

$$\mathbf{f}(\bar{b}_1, \dots, \bar{b}_k, \bar{u}) \stackrel{\gamma}{\equiv} \mathbf{f}(\bar{b}_1, \dots, \bar{b}_k, \bar{v}).$$

This notion was introduced by A. Bulatov [5] and further developed by E. Aichinger and N. Mudrinski [1]. In particular they have shown that for all  $\alpha_1, \dots, \alpha_k \in \text{Con } \mathbf{A}$  there is the smallest congruence  $\gamma$  with  $C(\alpha_1, \dots, \alpha_k; \gamma)$  called the  $k$ -ary commutator and denoted by  $[\alpha_1, \dots, \alpha_k]$ . Such generalized commutator behaves especially well in algebras from congruence modular varieties. In particular this commutator is fully symmetric, monotone, join-distributive and we have

$$[\alpha_1, [\alpha_2, \dots, \alpha_k]] \leq [\alpha_1, \dots, \alpha_k] \leq [\alpha_2, \dots, \alpha_k]$$

We will often use this generalized commutator when some (or possibly all) of the  $\alpha_i$ 's coincide. Thus to emphasize the arity of this supercommutator we will sometimes write  $[\alpha_1, \dots, \alpha_k]_k$  instead of  $[\alpha_1, \dots, \alpha_k]$ .

We say that an algebra  $\mathbf{A}$  is  $k$ -supernilpotent if  $[1, \dots, 1]_k = 0$ . The first inequality in the above display implies that a  $k$ -supernilpotent is  $k$ -nilpotent.

However, what is more interesting for our purposes, is the going down with supernilpotent powers of the congruences in the fashion of the solvable powers  $\theta^{[i]}$ . Since in a finite algebra the sequence

$$\theta \geq [\theta, \theta] = [\theta, \theta]_2 \geq \dots \geq [\theta, \dots, \theta]_i \geq [\theta, \dots, \theta]_{i+1} \geq \dots$$

has to stabilize, the intersection  $\theta_{[1]} = \bigcap_i [\theta, \dots, \theta]_i$  is actually one of the  $[\theta, \dots, \theta]_j$ 's. Now we simply put

$$\theta_{[0]} = \theta \quad \text{and} \quad \theta_{[k+1]} = \bigcap_i [\theta_{[k]}, \dots, \theta_{[k]}]_i.$$

This allows us to define  $h$ -step supernilpotent algebras, as those in which  $1_{[h]} = 0$ . Note that  $h$ -solvable algebras are  $h$ -step supernilpotent in this sense, so that  $h$ -step supernilpotent algebras need not be nilpotent.

However in our stratification we restrict ourselves to algebras that are nilpotent. First we recall a very nice result (due to [6] and [20]) illustrating the precise difference between nilpotency and supernilpotency for finite algebras  $\mathbf{A}$  from congruence modular variety. It says that the following two conditions are equivalent

- $\mathbf{A}$  is  $k$ -supernilpotent,
- $\mathbf{A}$  is  $k$ -nilpotent, decomposes into a direct product of algebras of prime power order and the clone of all terms of  $\mathbf{A}$  is generated by finitely many operations.

This nice result can be localized. To do that, first we need a concept of a characteristic of a covering pair  $\theta < \delta$  of congruences (of type 2, in the sense of Tame Congruence Theory [12]). The fact that  $\text{typ}(\theta, \delta) = 2$  says in particular that all

traces of  $(\theta, \delta)$ -minimal sets are, modulo  $\theta$ , (and up to polynomial equivalence) one-dimensional vector spaces over the same finite field. The prime number that is the characteristic of this fields is also used to be called the characteristic of the prime quotient  $\theta < \delta$  and denoted by  $\text{char}(\theta, \delta)$ . Now, for  $\alpha \leq \beta$  we put  $\text{char}\{\alpha, \beta\} = \{\text{char}(\theta, \delta) : \alpha \leq \theta < \delta \leq \beta\}$ . Note here that in our setting if the intervals  $I[\alpha, \beta]$  and  $I[\alpha', \beta']$  are projective we have  $\text{char}\{\alpha, \beta\} = \text{char}\{\alpha', \beta'\}$ . In case  $\varphi$  is a meet irreducible congruence, so that it has the unique cover, say  $\varphi^+$ , we will write  $\text{char}(\varphi)$  instead of  $\text{char}(\varphi, \varphi^+)$ .

The second concept needed to localize the characterization of supernilpotent algebras among the nilpotent ones, is a concept of a supernilpotent interval  $I[\alpha, \beta]$  in  $\text{Con } \mathbf{A}$ . We say that a congruence  $\beta \geq \alpha$  is supernilpotent over  $\alpha$  if  $\beta_{[1]} \leq \alpha$ .

The last concept needed is the one of a product interval. We say that  $I[\alpha, \beta]$  is the product interval if there are congruences  $\beta_1, \dots, \beta_s$  (called decomposition congruences) that intersect to  $\alpha$  and for each  $j$  satisfy  $(\bigcap_{i \neq j} \beta_i) \vee \beta_j = \beta$ . In nilpotent algebras, if  $\beta$  is supernilpotent over  $\alpha$  then the interval  $I[\alpha, \beta]$  is not only a product interval, but in fact it is prime uniform over each the  $\beta_j$ 's. To be more precise, by a pupi (or prime uniform product interval) we mean a product interval  $I[\alpha, \beta]$  in which each  $\text{char}\{\beta_j, \beta\}$  consists of a single prime, say  $p_j$ . Finally we say that  $I[\alpha, \beta]$  is prime strongly uniform product interval (psupi, for short) if it is a pupi and moreover the primes  $p_j$ 's are different for different  $\beta_j$ 's.

Note here that modularity of the congruence lattice implies that in a psupi there are no skew congruences between the  $\beta_j$ 's, i.e. for every  $\theta \in I[\alpha, \beta]$  we have  $\theta = \bigcap_{j=1}^s (\theta \vee \beta_j)$ . In particular each such  $\theta$  that is locally meet irreducible (i.e. meet irreducible in the interval  $I[\alpha, \beta]$ ) has to lie over one of the  $\beta_j$ 's.

Now, our localization says that for congruences  $\alpha < \beta$  of nilpotent algebra of finite type (from congruence modular varieties) the following two conditions are equivalent

- $\beta$  is supernilpotent over  $\alpha$ ,
- the interval  $I[\alpha, \beta]$  is prime uniform product interval.

A more detailed study of supernilpotent intervals and  $h$ -step supernilpotent stratification is contained in [17]. Here we only note that the equivalence of the above conditions can be shown by applying the VanderWerf's idea [25] of wreath decomposition. In fact this has been independently done by Mayr and Szendrei in [22].

For a better understanding of  $h$ -step supernilpotent algebras we observe first that in a finite algebra  $\mathbf{A}$  for every congruence  $\alpha$  there is the largest supernilpotent congruence over  $\alpha$ . This is due to the fact that the join of two supernilpotent (over  $\alpha$ ) congruences  $\beta_1, \beta_2$  is supernilpotent over  $\alpha$ . Indeed, this supernilpotency can be witnessed by



$[\beta_i, \dots, \beta_i]_k \leq \alpha$  with the same  $k$  for both the  $\beta_i$ 's. But now taking  $[\beta_1 \vee \beta_2, \dots, \beta_1 \vee \beta_2]_{2k}$  and distributing over the join we get a join of  $2k$ -folds supercommutators of the  $\beta_i$ 's. Since in each such supercommutator one of the  $\beta_i$ 's occurs at least  $k$ -times, this puts each of them, and therefore entire join of  $2^{2k}$  summands, below  $\alpha$ .

This allows us to define the sequence

$$0 = \sigma_0 \leq \sigma_1 \leq \dots \leq \sigma_k \leq \sigma_{k+1} \leq \dots$$

of congruences such that  $\sigma_{k+1}$  is the largest congruence that is supernilpotent over  $\sigma_k$ . This sequence of supernilpotent intervals strongly corresponds to the other one used to define  $h$ -step supernilpotency, namely

$$\dots \leq 1_{[k+1]} \leq 1_{[k]} \leq \dots \leq 1_{[2]} \leq 1_{[1]} \leq 1_{[0]} = 1.$$

Indeed, in  $h$ -step supernilpotent algebra, we induct on  $k = 0, \dots, h$  to show that  $1_{[h-k]} \leq \sigma_k$ . To pass from  $k$  to  $k+1$  we start with distributing over the join in the first supercommutator

$$\begin{aligned} (1_{[h-(k+1)]} \vee \sigma_k)_{[1]} &\leq (1_{[h-(k+1)]})_{[1]} \vee (\sigma_k)_{[1]} \\ &= 1_{[h-k]} \vee (\sigma_k)_{[1]} \leq \sigma_k, \end{aligned}$$

where the last inequality follows by induction hypothesis. But what we get means that  $1_{[h-(k+1)]} \vee \sigma_k$  is supernilpotent over  $\sigma_k$ , so that it has to be below  $\sigma_{k+1}$ , as required.

In particular, with  $k = h$  we get that  $1_{[h]} = 0$  implies  $\sigma_h = 1$ . In a similar fashion one shows that if  $\sigma_h = 1$  then  $1_{[k]} \leq \sigma_{h-k}$ , so that  $1_{[h]} = 0$ .

This gives that a finite algebra is  $h$ -step supernilpotent (i.e.  $1_{[h]} = 0$ ) iff  $\sigma_h = 1$ .

After all this preparation we are ready to show how the alternation of primes (crucial in our study of CSAT and CEQV for nilpotent algebras) is connected with  $h$ -step nilpotency.

**Theorem 2.1.** *For a finite nilpotent algebra  $\mathbf{A}$  from a congruence modular variety the following two conditions are equivalent:*

- $\mathbf{A}$  is  $h$ -step supernilpotent,
- every chain  $\varphi_1 < \varphi_2 < \dots < \varphi_s$  of meet irreducible congruences with alternating characteristics (i.e.  $\text{char}(\varphi_i) \neq \text{char}(\varphi_{i+1})$ , for  $i = 1, \dots, s-1$ ), has its length  $s$  bounded by  $h$ .

*Proof.* Suppose first that  $\varphi_1 < \varphi_2 < \dots < \varphi_s$  is such an alternating chain of meet irreducible congruences in an  $h$ -step supernilpotent algebra. The idea is to project them into the prime strongly uniform product intervals of the form  $\Sigma_i = I[\sigma_i, \sigma_{i+1}]$  by sending  $\theta$  to  $f_i(\theta) = (\theta \wedge \sigma_{i+1}) \vee \sigma_i$ . For meet irreducible  $\theta$  with the unique cover  $\theta^+$  pick  $j$  to be maximal with  $\sigma_j \leq \theta$ . Observe that then  $f_j(\theta) < f_j(\theta^+)$ , as otherwise the congruences  $\theta^+ \wedge \sigma_{j+1}, \theta, \theta^+, \sigma_{j+1}, \theta \vee \sigma_{j+1}$  would form a pentagon. Moreover one can show that for this particular  $j$  the congruence  $f_j(\theta)$  is locally (i.e. in  $\Sigma_j$ ) meet irreducible.

Thus, if  $s > h$  then after projecting the  $\varphi_i$ 's from our alternating chain into the  $\Sigma_i$ 's at least two consecutive ones will fall into the same  $\Sigma_j$ , say  $\Sigma_j$ , without collapsing them with their covers. But after such projection, both of them are locally meet irreducible in  $\Sigma_j$  so that being comparable they have to be over the same decomposition congruence  $\beta_i$ . Consequently they must have the same characteristic, contrary to our alternating assumption. This puts the bound for the alternating chain, as required.

Conversely, first note that since  $\mathbf{A}$  is nilpotent it is  $h$ -step supernilpotent for some  $h$ . But now we will use intervals  $\Pi_j = I[1_{[j]}, 1_{[j-1]}]$ . From the assumption that  $1_{[h-1]} > 0$  (i.e.  $\Pi_h \neq \emptyset$ ) we will construct the required chain of meet irreducible congruences of length  $h$ .

First, starting with an arbitrary prime  $p_h \in \text{char}\{\Pi_h\}$  we can go down with  $j = h-1, \dots, 1$  to isolate the consecutive primes  $p_j \in \text{char}\{\Pi_j\}$  satisfying  $p_j \neq p_{j+1}$ . After fixing  $p_h, \dots, p_{j+1}$  the possibility to properly choose  $p_j$  is equivalent to  $\text{char}\{\Pi_j\} \not\subseteq \{p_{j+1}\}$ . This obviously holds if  $|\text{char}\{\Pi_j\}| \geq 2$  so that we assume, to the contrary, that  $\text{char}\{\Pi_j\} = \{p_{j+1}\}$ . But  $p_{j+1} \in \text{char}\{\Pi_{j+1}\}$  is actually witnessed by one of the decomposition congruences, say  $\beta$ , of the interval  $\Pi_{j+1}$  so that  $\text{char}\{\beta, 1_{[j]}\} = \{p_{j+1}\}$ . But this gives  $\text{char}\{\beta, 1_{[j-1]}\} = \{p_{j+1}\}$ . Indeed, by modularity, every covering pair  $\beta \leq \theta < \delta \leq 1_{[j-1]}$  either projects down to  $1_{[j]} \wedge \theta < 1_{[j]} \wedge \delta$  inside  $I[\beta, 1_{[j]}]$ , or up to  $1_{[j]} \vee \theta < 1_{[j]} \vee \delta$  inside  $\Pi_{j+1}$ . In either case it inherits the characteristic  $p_{j+1}$ . However now,  $\text{char}\{\beta, 1_{[j-1]}\} = \{p_{j+1}\}$  yields that  $1_{[j-1]}$  is supernilpotent over  $\beta$  so that we get a contradiction  $1_{[j]} \leq \beta$ .

Now, knowing that there is an alternating chain of primes from  $\Pi_h \times \dots \times \Pi_1$  we will inductively show that  $\mathbf{A}$  has  $h$ -long chain of meet irreducible congruences  $\psi_h < \psi_{h-1} < \dots < \psi_1$  with alternating characteristics. The characteristics of these meet irreducibles do not necessarily coincide with the one from the starting chain of primes, as during the recursion process we call our procedure for a smaller (quotient) algebra  $\mathbf{A}'$  in which the sets  $\Pi'_j$  may be smaller. Thus, when passing from  $\mathbf{A}$  to  $\mathbf{A}'$  the initial sequence  $p_h, p_{h-1}, \dots, p_1$  may change to  $p_h, p'_{h-1}, \dots, p'_1$ , but the prime  $p_h$  at the lower level remains unchanged. All we need to take care of is that  $p_h \neq p'_{h-1}$ .

The easier case is when  $1_{[h-1]}$  does not cover 0 so that we can pick  $0 < \alpha < 1_{[h-1]}$ . Passing to the quotient algebra  $\mathbf{A}' = \mathbf{A}/\alpha$  we know that its chain of intervals  $\Pi'_j$  coincide with the original one of the  $\Pi_j$ , except  $\Pi'_h = I[\alpha, 1_{[h-1]}]$ . But constructing the chain of the  $p'_j$ 's for  $\mathbf{A}$  we may start with  $p_h = \text{char}(\alpha, 1_{[h-1]})$ . Then the chain of meet irreducibles for  $\mathbf{A}'$  nicely serves also for the original  $\mathbf{A}$ .

Also, if  $1_{[h-1]}$  is the unique atom of  $\mathbf{A}$ , the algebra is subdirectly irreducible, i.e. 0 is a meet irreducible congruence of  $\mathbf{A}$ . Induction hypothesis applied to the quotient  $\mathbf{A} = \mathbf{A}/1_{[h-1]}$ , but this time with  $h$  smaller by 1 and the shorter chain of primes  $p_{h-1}, \dots, p_1$  obtained from the one for  $\mathbf{A}$  by simply

deleting  $p_h$ , equip us with the  $(h - 1)$ -long chain of meet irreducibles, which after adding  $\psi_h = 0_A$  serves pretty well for  $A$ .

In the last case we have two different atoms  $1_{[h-1]}$  and  $\alpha$  in Con  $A$ . Again we will pass to the quotient  $A' = A/\alpha$ , but this time to make sure that this is going to work we need to make sure that the new intervals  $\Pi'_j = I[1_{[j]} \vee \alpha, 1_{[j-1]} \vee \alpha]$ 's are non trivial (so that exactly  $h$  corresponding primes can be chosen at all). Suppose to the contrary that for some  $j < h$  we have  $\alpha \vee 1_{[j]} = \alpha \vee 1_{[j-1]}$ , so that  $1_{[j-1]} \leq \alpha \vee 1_{[j]}$ . Obviously  $\alpha \not\leq 1_{[j]}$ , as otherwise  $1_{[j-1]} \leq 1_{[j]}$ , contrary to our assumption that the sequence of the  $1_{[j]}$ 's is strictly decreasing. On the other hand  $\alpha \leq 1_{[j-1]}$ , as otherwise  $1_{[j-1]}$  and  $\alpha$  would meet to 0 and therefore (together with  $1_{[j]}$ ) would generate a pentagon. In fact  $\alpha \neq 1_{[h]}$  tells us that then  $\alpha < 1_{[j-1]}$ , so that we can pick  $\gamma$  with  $\alpha \leq \gamma < 1_{[j-1]}$ . As every congruence is supernilpotent over each of its subcovers, we get that  $1_{[j]} \leq \gamma$  and consequently we get a contradiction  $\gamma \geq 1_{[j]} \vee \alpha = 1_{[j-1]} \vee \alpha = 1_{[j-1]}$ .

Now note that although the intervals  $\Pi_j$ 's may loose the prime  $\text{char}(0, \alpha)$ , we know that the only candidate for  $p_h$ , namely  $\text{char}(0, 1_{[h-1]})$  still stays in  $\text{char}\{\Pi'_j\}$  as  $p_h = \text{char}(0, 1_{[h-1]}) = \text{char}(\alpha, \alpha \vee 1_{[h-1]})$ .  $\square$

### 3 A paradigm for $h$ -step supernilpotent algebras

We start with an algebra that will serve us as paradigm for our considerations. Fix a positive integer  $h$  and a sequence  $p_1, p_2, \dots, p_h$  of primes. Define an algebra  $D[p_1, \dots, p_h]$  to be the expansion of the product  $Z_{p_1} \times \dots \times Z_{p_h}$  of Abelian groups by the additional unary operations  $e_1, \dots, e_h$  and  $v_1, \dots, v_{h-1}$  defined for  $x = (x_1, \dots, x_h) \in Z_{p_1} \times \dots \times Z_{p_h}$  by

$$\begin{aligned} e_j(x) &= (0, \dots, 0, x_j, 0, \dots, 0), \\ v_j(x) &= (0, \dots, 0, b_j(x_{j+1}), 0, \dots, 0), \end{aligned}$$

where  $b_j : Z_{p_{j+1}} \rightarrow Z_{p_j}$  is a function given by  $b_j(0) = 0$  and  $b_j(a) = 1$  otherwise.

Note here that

- the algebra  $D[p]$  is simply the group  $Z_p$ , so that it is Abelian,
- the algebra  $D[p, q]$ , with  $p \neq q$  had been extensively studied in [16] where a polytime algorithm was presented both for  $\text{CSAT}(D[p, q])$  and  $\text{CEQV}(D[p, q])$ .

Here we will study the algebras of the form  $D[p_1, \dots, p_h]$  with the assumption that the sequence  $p_1, p_2, \dots, p_h$  of primes is alternating, i.e.  $p_i \neq p_{i+1}$ . Then we will show that

- the algebra  $D[p_1, \dots, p_h]$  is  $h$ -nilpotent (actually  $h$ -step supernilpotent).

Since the algebra  $D[p_1, \dots, p_h]$  has an underlying group structure each equation of polynomials  $t = s$  that may be an input to  $\text{CSAT}$  or  $\text{CEQV}$  can be replaced by  $t - s = 0$  so that we restrict ourselves to the equations of this special shape.

#### 3.1 The structure of $D[p_1, \dots, p_h]$

To understand the algebra  $D = D[p_1, \dots, p_h]$  we start with defining a couple of its constants  $0 = (0, \dots, 0)$ ,  $1 = (1, \dots, 1)$  and polynomials  $e^k$  by putting  $e^k(x) = \sum_{j \geq k} e_j(x)$ .

Now it is easy to observe that the relations

$$\theta_k = \{(a, b) \in D^2 : e^k(a) = e^k(b)\}$$

together with the total congruence  $\theta_{h+1}$  form a chain  $0 = \theta_1 < \theta_2 < \dots < \theta_h < \theta_{h+1} = 1$  and that they are actually all congruences of  $D$  – indeed every principal congruence of  $D$  is one of the  $\theta_i$ 's.

Inducting on the complexity of a polynomial  $t(x_1, \dots, x_n)$  of  $D$  we can easily show that  $e_j t(\bar{a}) = e_j t(\bar{b})$ , whenever  $j \geq k$  and  $a_i \theta_k b_i$ . This means that a polynomial having a range contained in  $e_k(D) = \{0\} \times \dots \times \{0\} \times Z_{p_k} \times \{0\} \times \dots \times \{0\}$  does not depend on the values of the first  $k - 1$  summands in  $x = \sum_{j=1}^h e_j(x)$ . Also an inspection of the behavior of the basic operations of  $D$  (in particular noticing that  $e_k(x+y) = e_k(x) + e_k(y)$ ,  $e_k(e_k(x)) = e_k(x)$ ,  $e_k(v_k(x)) = v_k(x) = v_k(e_{k+1}(x))$  and  $e_k(e_\ell(x)) = 0 = e_k(v_\ell(x))$  for  $k \neq \ell$ ), allows us to represent every polynomial  $t(\bar{x})$  with the range contained in  $e_k(D)$ , i.e. a polynomial satisfying  $t = e_k t$ , by a sum of expressions of the form  $e_k x_i$ ,  $e_k c$  or  $v_k s$ , where  $x_i$  is a variable,  $c$  is a constant and  $s$  is some polynomial of  $D$ . In order to have  $v_k s \neq 0$  we may assume that the range of  $s$  is contained in  $e_{k+1}(D)$ , as  $v_k s = v_k e_{k+1} s$ . However, as we have already noticed, polynomials with the range contained in  $e_{k+1}(D)$  depends only on the projections  $e^{k+1}(x_i)$  of its variables. Summing up we know that

$$e_k t(\bar{x}) = c + \sum_{i=1}^n \lambda_i \cdot e_k(x_i) + \sum_{s \in S} \kappa_s \cdot v_k s(e^{k+1}(x_1), \dots, e^{k+1}(x_n)),$$

where  $c \in e_k(D)$  is a constant, the multiplication by the scalars  $\lambda_i$ 's or  $\kappa_s$  (taken from  $Z_{p_k}$ ) is a shortening for adding the appropriate elements appropriate number of times, and  $S$  is a set of polynomials of  $D$  with ranges contained in  $e_{k+1}(D)$ .

To estimate the length of the above representation of  $e_k t$  note that since  $e_k$  distributes over the addition, we know that the number of summands in the above display (including those hidden in the  $\lambda_i$ 's and the  $\kappa_s$ 's) is bounded by the number of additions in  $t$ . Moreover note that each  $s \in S$  is in fact a subterm of  $t$  (and that they are pairwise disjoint subterms of  $t$ ) so that  $\sum_{s \in S} |s| \leq |t|$ . In particular the length of the above representation is bounded by  $O(|t|)$ . We will often refer to this representation as the canonical representation

keeping in mind that  $\mathbf{t}(\bar{x}) = \sum_{j=1}^k e_k \mathbf{t}(\bar{x})$  and that

$$\begin{aligned} e_h \mathbf{t}(\bar{b}) &= c^h + \sum_{i=1}^n \lambda_i^h \cdot e_h(b_i), \\ e_{h-1} \mathbf{t}(\bar{b}) &= c^{h-1} + \sum_{i=1}^n \lambda_i^{h-1} \cdot e_{h-1}(b_i) \\ &\quad + \sum_{s \in S^{h-1}} \kappa_s^{h-1} \cdot v_{h-1} \mathbf{s}(e^h(b_1), \dots, e^h(b_n)), \\ &\vdots \\ e_1 \mathbf{t}(\bar{b}) &= c^1 + \sum_{i=1}^n \lambda_i^1 \cdot e_1(b_i) \\ &\quad + \sum_{s \in S^1} \kappa_s^1 \cdot v_1 \mathbf{s}(e^2(b_1), \dots, e^2(b_n)). \end{aligned} \quad (1)$$

But what is more important for us is that such (relatively short) canonical representation can be obtained not only from a polynomial of  $\mathbf{D}$  but also from a circuit  $\Gamma$  over  $\mathbf{D}$  that computes this polynomial. This is not entirely obvious, as sometimes circuits may have logarithmic size with respect to the length of a polynomial they compute. Each node of the circuit  $\Gamma$  determines a subcircuit  $\Gamma'$  of  $\Gamma$ . With each  $\Gamma'$  we associate a polynomial  $\mathbf{t}_{\Gamma'}$  (possibly too large) in such a way that  $\mathbf{t}_{\Gamma''}$  is a subpolynomial of  $\mathbf{t}_{\Gamma'}$  whenever  $\Gamma''$  is determined by a node in  $\Gamma'$ . Despite the sizes of the  $\mathbf{t}_{\Gamma'}$ 's we go to their canonical representations, as described in (1). All the data we need to store for the  $e_j \mathbf{t}_{\Gamma'}$ 's ( $j = 1, \dots, h$ ) are the constants  $c^j, \lambda_i^j, \kappa_s^j$  and the sets  $S^j$  themselves. There is an easy bound for the constants, once we bound  $|S^j|$ . To unwind this recursive construction note that for each  $\mathbf{s} \in S^j$  we need to keep its data only for one level, namely  $e_{j+1} \mathbf{s}$ , as  $v_j \mathbf{s} = v_j e_{j+1} \mathbf{s}$ . Now, since  $|S^j|$  is bounded from above by the number of subcircuits of  $\Gamma'$  we get  $|S^j| \leq |\Gamma'|$ . So, unwinding this construction for entire  $\Gamma$  we get that our canonical representation of  $\mathbf{t}_{\Gamma}$  is of size  $O(|\Gamma|^{h^2})$ . Thus, in what follows, we will simply put our lower and upper bounds in terms of the size of canonical representation for polynomials rather than for circuits.

Our next observation shows a connection between some polynomials of  $\mathbf{D}[p_1, \dots, p_h]$  and  $\text{CC}[p_1, \dots, p_h]$ -circuits.

**Fact 3.1.** *For an  $n$ -ary polynomial  $g$  of  $\mathbf{D}$ , and  $j < k$  the mapping*

$$v_j e_{j+1} \mathbf{g}(e_k x_1, \dots, e_k x_n) : \{0, e_k 1\}^n \longrightarrow \{0, e_j 1\}$$

*can be simulated by a  $\text{CC}[p_{j+1}, \dots, p_k]$ -circuit  $\Gamma$  of depth  $k - j$  and size  $O(|g|)$  in a way, that*

$$v_j e_{j+1} \mathbf{g}(e_k x_1, \dots, e_k x_n) = \begin{cases} 0, & \text{if } \Gamma(b(x_1), \dots, b(x_n)) = 0, \\ e_j 1, & \text{if } \Gamma(b(x_1), \dots, b(x_n)) = 1, \end{cases}$$

*where the Boolean function  $b(x) : \{0, e_k 1\} \longrightarrow \{0, 1\}$  returns 0 if  $e_k(x) = 0$  and 1 otherwise.*

*Proof.* We induct on  $j = k - 1, \dots, 1$  to build the required circuit from the gates  $\text{MOD}_{p_i}^R$ 's. Since for  $j = k - 1$  the function  $e_k \mathbf{g} e_k$  actually maps  $e_k D^n$  into  $e_k D$ , it has to be an affine function of the form  $c + \sum_{i=1}^n \lambda_i e_k x_i$ . Thus  $v_{k-1} e_k \mathbf{g} e_k$  can be simulated by one gate  $\text{MOD}_{p_k}^{Z_{p_k} - \{-c\}}$  with each of the  $b(x_i)$ 's put to the gate  $\lambda_i$  times on input.

Going down with  $j$  our canonical form gives that

$$\begin{aligned} e_{j+1} \mathbf{g}(e_k \bar{x}) &= c^{j+1} + \sum_{i=1}^n \lambda_i^{j+1} \cdot e_{j+1}(e_k x_i) \\ &\quad + \sum_{s \in S^{j+1}} \kappa_s^{j+1} \cdot v_{j+1} \mathbf{s}(e^{j+1}(e_k x_1), \dots, e^{j+1}(e_k x_n)), \end{aligned}$$

which actually reduces to

$$e_{j+1} \mathbf{g}(e_k \bar{x}) = c^{j+1} + \sum_{s \in S^{j+1}} \kappa_s^{j+1} \cdot v_{j+1} \mathbf{s}(e_k x_1, \dots, e_k x_n),$$

as  $e_{j+1} e_k x = 0$  and  $e^{j+1} e_k x = e_k x$ . Now, given the circuits  $\Gamma_s$  that do the job for all the  $v_{j+1} e_{j+2} s e_k$  with  $\mathbf{s} \in S^{j+1}$ , we feed  $\text{MOD}_{p_{j+1}}^{Z_{p_k} - \{-c^{j+1}\}}$  with each  $\Gamma_s$  repeated  $\kappa_s^{j+1}$  times.  $\square$

Using our understanding of polynomials of  $\mathbf{D}$ , provided by the canonical representation (1), we can now easily determine the behavior of the commutator of congruences of  $\mathbf{D}$ . Namely if  $i \leq j$  then  $[\theta_i, \theta_j] = \theta_{i-1}$ .

We start here with an adaptation of Lemma 3.1 from [16]. The original Lemma has been formulated for the algebra of the form  $\mathbf{D}[p, q]$  with  $p \neq q$ , while we will need it in our more general context of  $\mathbf{D}[p_1, \dots, p_h]$ . Obviously the algebra  $\mathbf{D}[p_k, p_{k+1}]$  can be identified with  $e_k \mathbf{D} + e_{k+1} \mathbf{D} = \{0\} \times \dots \times \{0\} \times Z_{p_k} \times Z_{p_{k+1}} \times \{0\} \times \dots \times \{0\} \subseteq \mathbf{D}$ .

**Lemma 3.2.** *For  $k < \ell$  and all  $m$ , every function of the form  $g : (e_\ell D)^m \longrightarrow e_k D$  can be represented by an  $m$ -ary polynomial  $\mathbf{p}$  of  $\mathbf{D}$ , with both its size and the time needed to actually compute it bounded by  $O(2^{cm})$ , where the constant  $c$  depends only on the algebra  $\mathbf{D}$ .*

*Proof.* We start with observing that for a polynomial  $v'_k(x) = e_k 1 - v_k(e_{k+1} 1 - e_{k+1} x)$  we have  $v'_k(0) = 0, v'_k(e_{k+1} 1) = e_k 1 \neq 0$  and  $v'_k(e_{k+1} a) = 0$  for all  $a \in e_{k+1} D - \{e_{k+1} 1\}$ . Thus  $\sum_{a \in e_{k+1} D - \{0\}} v'_k(a) = v'_k(e_{k+1} 1) \neq 0$ , so that we are in the scope of Lemma 3.1 of [16] which yields a required polynomial  $\mathbf{p}$  representing the function  $g$ . Moreover the shape of this polynomial (provided in that Lemma) allows us to bound its size (and the time to produce it) by  $O(p_{k+1}^{m+1} \cdot p_k \cdot m \cdot (p_{k+1} + 1)) = O(p_{k+1}^m \cdot m)$ , as required.

Now, if  $\ell > k + 1$ , we inflate each variable  $x_i$ , ( $i = 1, \dots, m$ ) into  $p_\ell$  variables  $x_i^0, \dots, x_i^{p_\ell-1}$ . This allows us to project an element  $a \in e_\ell D$  into a tuple  $(a^0, a^1, \dots, a^{p_\ell-1}) \in e_k D^{p_\ell}$  by putting  $a^\lambda = 1 - v_{k+1} \dots v_{\ell-1}(a - \lambda \cdot e_\ell 1)$ . Note that for each such  $a$  exactly one of the  $a^\lambda$ 's is nonzero (actually it is  $e_{k+1} 1$ ), namely the one with  $\lambda$  occurring in the  $\ell$ -th position of  $a = (a_1, \dots, a_h)$ . Take any function  $g' : (e_{k+1} D)^{p_\ell \cdot m} \longrightarrow e_k D$

satisfying

$$g'(x_1^0, \dots, x_1^{p_\ell-1}, x_2^0, \dots, x_m^0, \dots, x_m^{p_\ell-1}) = g(x_1, \dots, x_m)$$

whenever  $x_i^\lambda = 1 - v_{k+1} \dots v_{\ell-1}(x_i - \lambda \cdot e_\ell 1)$ . Using the case  $\ell = k+1$  the function  $g'$  can be represented by a  $p_\ell \cdot m$ -ary polynomial  $\mathbf{p}'$  of  $\mathbf{D}$ . It should be obvious that now substituting  $1 - v_{k+1} \dots v_{\ell-1}(a - \lambda \cdot e_\ell 1)$ 's for the  $x_i^\lambda$ 's we get an  $m$ -ary polynomial  $\mathbf{p}$  of  $\mathbf{D}$  representing  $g$ . Moreover  $|\mathbf{p}| \leq O(|\mathbf{p}'|) \leq O(2^{c'm})$  with  $c' = p_\ell c$ .

An inspection of the proof of Lemma 3.1 in [16] provides a bound for the time needed to actually find the required polynomials, as claimed.  $\square$

We conclude this subsection with mentioning a very nice feature of the algebra  $\mathbf{D}$ . Namely  $\mathbf{D}$  is as rich in polynomials as possible. This means that every function  $g : D^n \rightarrow D$  that preserves congruences of  $\mathbf{D}$  and their commutator is already a polynomial of  $\mathbf{D}$ . As we are not going to use this fact in our future considerations we provide only a brief sketch of its proof.

Starting with  $g$  preserving congruences and their commutator we know that the algebra  $\mathbf{D}$  endowed by  $g$  is still nilpotent. Obviously  $g$  can be represented as the sum  $\sum_{k=1}^h e_k g$ . Since the range of  $e_k g$  is contained in  $e_k D$ , we can recursively apply Proposition 7.1 of [6] to claim that  $e_k g(\bar{x})$  can be represented by  $c^k + \sum_{i=1}^n \lambda_i^k \cdot e_k(x_i) + g'(e^{k+1}(x_1), \dots, e^{k+1}(x_n))$ , where  $g'$  is some  $n$ -ary function mapping  $e^{k+1}D$  into  $e_k D$ . Now, with a little bit more effort we can strengthen Lemma 3.2 to represent every functions mapping simultaneously all upper levels  $e_{k+1}D, \dots, e_h D$  (i.e. the entire  $e^{k+1}D$  not just one level  $e_\ell$  as in that Lemma) into a lower level  $e_k D$  by a polynomial of  $\mathbf{D}$ . This would show that  $g'$  and therefore  $g$  are the polynomials.

### 3.2 Lower bound

We are going to show that under the assumption of the ETH the complexity for both CSAT and CEQV for  $\mathbf{D}[p_1, \dots, p_h]$  is at least  $\Omega(2^{c \cdot \log^{h-1} |\Gamma|})$  where  $|\Gamma|$  is the size of a circuit  $\Gamma$  on the input.

To deal with CSAT, for every formula  $\Phi(x_1, \dots, x_n)$  in 3-CNF we will construct an  $n$ -ary polynomial  $\mathbf{t}_\Phi(x_1, \dots, x_n)$  such that  $\Phi$  is satisfiable iff the equation  $\mathbf{t}_\Phi(\bar{x}) = e_1 1$  has a solution in  $\mathbf{D}$ . We will make sure that the time required to produce  $\mathbf{t}_\Phi$  is bounded by  $O(2^{cm^{1/(h-1)}})$ , where  $m$  is the number of clauses in  $\Phi$ . Now, having algorithms for CSAT( $\mathbf{D}[p_1, \dots, p_h]$ ) working in time  $O(2^{\varepsilon \cdot \log^{h-1} |\Gamma|})$  for arbitrary small  $\varepsilon > 0$  we would be able to solve 3-CNF-SAT in the very same time with  $|\Gamma|$  replaced by  $d \cdot 2^{cm^{1/(h-1)}}$ , i.e. in  $O(2^{\varepsilon c^{h-1} m})$ . This obviously contradicts ETH (after remodelling it with the Sparsification Lemma).

To produce  $\mathbf{t}_\Phi$  we start with the  $s$ -ary functions  $\text{AND}_k^s : e_{k+1}D^s \rightarrow e_k D$  defined by  $\text{AND}_k^s(a_1, \dots, a_s) = 0$  if at least of the  $a_i$ 's is 0, and  $\text{AND}_k^s(a_1, \dots, a_s) = e_k 1$  otherwise. Lemma

3.2 assures us that all those functions can be realized by polynomials of  $\mathbf{D}$  in  $O(2^{cs})$  time, possibly with different constants  $c$  depending on  $k$ .

Although the functions  $\text{AND}_k^s$  are long, the composition of two consecutive ones is shorter (in terms of the variables involved). Indeed the function

$$\text{AND}_{k-1}^s (\text{AND}_k^s(x_1, \dots, x_s), \dots, \text{AND}_k^s(x_{(s-1)s+1}, \dots, x_{s^2}))$$

acts from  $e_{k+1}D^{s^2}$  into  $e_{k-1}D$  and can be produced in  $O(s \cdot 2^{cs^{2k-1}}) = O(2^{cs})$  time. Repeating this procedure we end up with a  $s^{h-2}$ -ary polynomial AND, of size/time  $O(2^{cs})$ , mapping  $e_{h-1}D^{s^{h-2}}$  into  $e_1 D$  and behaving as a conjunction, i.e.  $\text{AND}(\bar{a}) = 0$  if some of the  $a_i$ 's is 0, and  $\text{AND}(\bar{a}) = e_1 1$  otherwise.

The above part of our construction has been independent of  $\Phi$ . We are going to use the action of the level  $e_h D$  onto  $e_{h-1}D$  to code  $\Phi$ . To start with we define a boolean function  $b : e_h D \rightarrow \{\top, \perp\}$  by putting  $b(a) = \top$  for all  $a \neq 0$  and  $b(0) = \perp$ . Now, if  $m$  is the number of clauses in  $\Phi$  we fix  $s$  to be  $\lceil m^{1/(h-1)} \rceil$  and split the clauses into  $s^{h-2}$  parts, say  $\Phi_\ell$ 's, each of which containing at most  $s$  clauses, so that each  $\Phi_\ell$  involves at most  $n_\ell \leq 3s$  variables. Again we refer to Lemma 3.2 to ensure that the function  $\text{CNF}_{\Phi_\ell} : e_h D^{n_\ell} \rightarrow e_{h-1}D$  given by  $\text{CNF}_{\Phi_\ell}(a_1, \dots, a_{n_\ell}) = 0$  if  $\Phi_\ell(b(a_1), \dots, b(a_{n_\ell})) = \perp$  and  $\text{CNF}_{\Phi_\ell}(a_1, \dots, a_{n_\ell}) = e_{h-1} 1$  otherwise, can be realized by a polynomial (again in time bounded by  $O(2^{cs})$ ). For simplicity we make sure that occurrence of each variable  $x$  is replaced by  $e_h(x)$ .

Now, filling up our  $s^{h-2}$ -ary polynomial AND with  $n_\ell$ -ary polynomials  $\text{CNF}_{\Phi_\ell}$ 's we finally arrive at the polynomial  $\mathbf{t}_\Phi$ . Again we produced it in  $O(2^{cs})$  time, for some (possibly new) constant  $c$ . It should be obvious that  $\mathbf{t}_\Phi$  does the required job for us.

To get a similar lower bound for CEQV( $\mathbf{D}[p_1, \dots, p_h]$ ) it suffices to notice that the constructed polynomial  $\mathbf{t}_\Phi$  takes only two values: 0 and  $e_1 1$ . In such a case CEQV and CSAT could be bisimulated.

### 3.3 Deterministic upper bound

This subsection is devoted to analyze a solution space for an equation  $\mathbf{t}(\bar{x}) = 0$  over the algebra  $\mathbf{D} = \mathbf{D}[p_1, \dots, p_h]$ . This analysis is based on SESH (Strong Exponential Size Hypothesis). This will lead to an algorithm that solves the equations (and therefore satisfiability of circuits) over  $\mathbf{D}$  in subexponential time almost matching the lower bound from Subsection 3.2.

Recall here that in a supernilpotent algebra  $\mathbf{A}$  an equation has a solution if it has one which is almost constant, say equal to 0, i.e. the number of non-zero values for  $x_1, \dots, x_n$  is bounded by a constant depending only on the algebra  $\mathbf{A}$ . We are using the algebra  $\mathbf{D}[p_1, \dots, p_h]$  as a paradigm for  $h$ -step supernilpotent algebras to show (under the assumption of



SESH) that in such realm if an equation  $\mathbf{t}(\bar{x}) = 0$  has a solution then it has one which again is *almost* constant, but this time *almost* means that there are at most  $O(\log^{h-1} |\mathbf{t}|)$  non-zero values. Thus to check if  $\mathbf{t}(\bar{x}) = 0$  has a solution it suffices to check if there is one among those *almost* constant tuples. Since there are at most  $O(n^{\log^{h-1} |\mathbf{t}|} \cdot |D|^{\log^{h-1} |\mathbf{t}|}) = O(2^{c \log^h |\mathbf{t}|})$  such candidates, while checking if one is actually a solution takes roughly  $O(|\mathbf{t}|)$ , we have an algorithm working in  $O(2^{c \log^h |\mathbf{t}|})$  time.

For two tuples  $\bar{a} = (a_1, \dots, a_n)$  and  $\bar{b} = (b_1, \dots, b_n)$  from  $D^n$  we put  $\|\bar{a} = \bar{b}\| = \{i : a_i = b_i\}$  and analogously for  $\|\bar{a} \neq \bar{b}\|$ .

Now we will show how a solution  $\bar{a}$  of  $\mathbf{t}(\bar{x}) = 0$  can be successively modified, by to get a sequence of solutions  $\bar{a} = \bar{a}^0 \rightarrow \bar{a}^1 \rightarrow \dots \rightarrow \bar{a}^h$ . When passing from  $\bar{a}^{k-1}$  to  $\bar{a}^k$  we will introduce zeros on the  $k$ -th coordinate, i.e. making  $e_k(a_i) = 0$ , for more and more  $i$ 's, while keeping the other coordinates unchanged. To be more precise we will make sure that  $\#\|e_k(\bar{a}^k) \neq 0\| \leq O(\log^{k-1} |\mathbf{t}|)$  and  $e_j(\bar{a}^k) = e_j(\bar{a}^{k-1})$  for  $j \neq k$ . Thus we will get

$$\#\left\{i : \sum_{j \leq k} e_j(a_i^k) \neq 0\right\} \leq \sum_{j \leq k} O(\log^{j-1} |\mathbf{t}|) = O(\log^{k-1} |\mathbf{t}|),$$

so that finally arriving at  $\bar{a}^h$  we end up with  $\#\|\bar{a}^h \neq 0\| \leq O(\log^{h-1} |\mathbf{t}|)$ , as promised.

To keep our second invariant when passing from  $\bar{a}^{k-1}$  to  $\bar{a}^k$  we need to stay inside the set

$$E^k = \{\bar{b} : e_j(\bar{b}) = e_j(\bar{a}^{k-1}) \text{ for } j \neq k\},$$

in particular we secure  $e^{k+1}(\bar{b}) = e^{k+1}(\bar{a}^{k-1})$  so that for  $j > k$  we have  $e_j \mathbf{t}(\bar{b}) = e_j \mathbf{t}(\bar{a}^{k-1})$ .

In particular, when producing  $\bar{a}^1 \in E^1$  we need to take care only of  $e_1 \mathbf{t}(\bar{a}^1) = c^1 + \sum_i \lambda_i^1 e_1(a_i^1) + \mathbf{t}'(e^2(\bar{a}^1))$ . However  $\mathbf{t}'(e^2(\bar{b}))$  gives the same value for all  $\bar{b} \in E^1$ . Thus our requirement that  $\bar{a}^1$  is still a solution reduces to the equation  $0 = e_1 \mathbf{t}(\bar{a}^1) = c^1 + \sum_i \lambda_i^1 e_1(a_i^1)$ . Therefore one can easily find a solution  $\bar{a}^1$  to this linear equation with at most one of the  $a_i^1$ 's being non-zero.

Also, when passing from  $\bar{a}^{k-1}$  to  $\bar{a}^k$  we choose  $\bar{a}^k \in E^k$  to be a solution to  $\mathbf{t}(\bar{x}) = 0$  that maximizes the number of zeros for  $e_k(a_1^k), \dots, e_k(a_n^k)$ . If  $\bar{a}^k$  would still have too many non-zeros we will construct a relatively short polynomial (of the arity corresponding to the number of those nonzeros) that behaves as conjunction and refer to SESH to get a contradiction.

We start this argument with a better understanding of solutions  $\bar{b} \in E^k$  to the equation  $\mathbf{t}(\bar{x}) = 0$ . For such  $\bar{b}$  to be a

solution reduces to the system of equations:

$$\begin{aligned} 0 &= e_k \mathbf{t}(\bar{b}), \\ 0 &= e_{k-1} \mathbf{t}(\bar{b}), \\ &\vdots \\ 0 &= e_1 \mathbf{t}(\bar{b}), \end{aligned}$$

where each  $e_j \mathbf{t}(\bar{b})$  is represented in its canonical form as in (1). The last sum in the representation of  $e_k \mathbf{t}$  occurs only if  $k \neq h$ . Actually this sum disappears independently of how big is  $k$ . This is because this sum is constant on the set  $E^k$ . Also the linear parts in all equations with  $j < k$  are constant as  $e_j(b_i) = e_j(a_i^{k-1})$  for  $\bar{b} \in E^k$ . This also allows to replace  $e^j(b_i)$  by  $e_k(b_i)$ . By possibly modifying the constants  $c^1, \dots, c^k$  (and the sets  $S^1, \dots, S^{k-1}$  of polynomials) we are left with finding  $\bar{b} \in E^k$  satisfying

$$\begin{aligned} 0 &= c^k + \sum_{i=1}^n \lambda_i^k \cdot e_k(b_i), \\ 0 &= c^{k-1} + \sum_{s \in S^{k-1}} \kappa_s^{k-1} \cdot v_{k-1} s(e_k(b_1), \dots, e_k(b_n)), \\ &\vdots \\ 0 &= c^1 + \sum_{s \in S^1} \kappa_s^1 \cdot v_1 s(e_k(b_1), \dots, e_k(b_n)). \end{aligned} \tag{2}$$

We want to replace this system of equations by a single equation (of about the same size). We will do it with the help of the  $(|S^1| + k - 1)$ -ary function  $V : e_2 D^{|S^1|+k-1} \rightarrow e_1 D$  defined (on the variables  $z_s$  indexed by  $s \in S^1$  and  $z_2, \dots, z_k$ ) by

$$V(\dots, z_s, \dots, z_2, \dots, z_k) = \left( e_1 1 - \left( c^1 + \sum_{s \in S^1} \kappa_s^1 \cdot v_1(z_s) \right)^{p_1-1} \right) \cdot \prod_{j=2}^k (e_1 1 - v_1(z_j)).$$

Note that  $V(\dots, z_s, \dots, z_2, \dots, z_k) = e_1 1$  iff all the  $z_2, \dots, z_k$  as well as  $c^1 + \sum_{s \in S^1} \kappa_s^1 \cdot v_1(z_s)$  are zeros. Now, denoting by  $r_j(\bar{b})$  the right hand side of the  $j$ -th equation (counting from the bottom) and substituting  $v_2 \dots v_{j-1} r_j(\bar{b})$  for  $z_j$  and  $s(e_k(b_1), \dots, e_k(b_n))$  for the  $z_s$ 's, we reduced our system of equations to just one equation of the form  $V(\dots) = e_1 1$ , where inside  $V$  there are polynomials of  $D$  with total length bounded by  $O(|\mathbf{t}|)$ .

Obviously, by Lemma 3.2,  $V$  can be represented by a polynomial of  $D$ . However to have a control of its size we need a little bit more subtle argument. First we distribute all the multiplications in  $V$  to end up with a sum of a constant and expressions of the form  $v_1(y_1) \dots v_1(y_\ell)$ , with  $y_i$ 's being the variables  $z_j$ 's or  $z_s$ 's. It should be obvious that this sum has at most  $(1 + |\mathbf{t}|^{p_1-1} \cdot 2^{k-1}) \leq O(|\mathbf{t}|^{p_1})$  summands. Moreover  $\ell$  is bounded by a constant  $(p_1 - 1) + (k - 1)$  independent of  $\mathbf{t}$ . This allows us to call Lemma 3.2 to represent all the

$\ell$ -ary functions  $v_1(y_1) \cdot \dots \cdot v_1(y_\ell)$  by polynomials of  $\mathbf{D}$  with lengths bounded by a constant independent of  $\mathbf{t}$ .

Up to now, we end up with a polynomial  $\mathbf{t}^*(\bar{x})$  of size  $O(|\mathbf{t}|^c)$  (for some constant  $c$ ) such that inside  $E^k$  the equations  $\mathbf{t}(\bar{x}) = 0$  and  $\mathbf{t}^*(\bar{x}) = e_1 1$  have the same solutions. Moreover, the shape of  $\mathbf{V}$  tells us that  $\mathbf{V}$  (and therefore  $\mathbf{t}^*$ ) takes only two values, namely 0 and  $e_1 1$  and therefore we will modify it to simulate the operation of conjunction with entries from  $\{0, e_k 1\}$  and values 0,  $e_1 1$ . The fact that in the polynomials  $\mathbf{r}_j$  (and therefore in  $\mathbf{t}^*$ ) all variables  $x_i$  are in the scope of  $e_k$  will be helpful in our further analysis.

By our choice  $\bar{a}^k \in E^k$  is a solutions to  $\mathbf{t}^*(\bar{x}) = e_1 1$  minimizing the cardinality of the set  $\|e_k(\bar{a}^k) \neq 0\|$ . Now we modify  $\mathbf{t}^*$  to  $\mathbf{t}^{**}$ , first by fixing each variable  $x_i$  to be  $a_i^k$  whenever  $e_k(a_i^k) = 0$  and then by replacing each of the remaining variables  $x_i$  by  $\lambda_i \cdot x_i$  where  $\lambda_i$  is the unique nonzero coordinate of  $e_k(a_i^k)$  (and as previously  $\lambda \cdot x$  is the sum  $x + \dots + x$  with  $\lambda$  summands). Let  $\ell$  be the arity of  $\mathbf{t}^{**}$  so that without loss of generality we may assume that the first  $\ell$  variables of  $\mathbf{t}^*$  survived. We claim that  $\mathbf{t}^{**}$  is the required conjunction. Indeed,  $\mathbf{t}^{**}(e_k 1, \dots, e_k 1) = \mathbf{t}^*(e_k(\bar{a}^k)) = \mathbf{t}^*(\bar{a}^k) = e_1 1$ , while, by maximality of  $\|e_k(\bar{a}^k) \neq 0\|$ , a tuple  $\bar{b} \in D^\ell$  with  $b_i = 0$  for  $i \leq \ell$  cannot be a solution to  $\mathbf{t}(\bar{x}) = 0$  so that  $\mathbf{t}^{**}(\bar{b}) = 0$ .

Now Fact 3.1 allows us to create a circuit of size  $O(|\mathbf{t}^{**}|) = O(|\mathbf{t}|^d)$  and of depth  $k$  that computes the  $\ell$ -ary conjunction. However SESH tells us that the size of this circuit has to be at least  $\Omega(2^{c\ell^{1/(k-1)}})$ . This gives  $\# \|e_k(\bar{a}^k) \neq 0\| = \ell \leq O(\log^{k-1} |\mathbf{t}|)$ , as required.

To see that  $\text{CEQV}(\mathbf{D}[p_1, \dots, p_h])$  can be solved roughly in the very same time, note that determining if the identity  $\mathbf{t}(\bar{x}) = 0$  holds we need to check that none of the equations of the form  $\mathbf{t}(\bar{x}) - d = 0$ , with  $d \in D - \{0\}$  has a solution.

### 3.4 Probabilistic upper bound

We present a randomized algorithm for checking whether an equation  $\mathbf{t}(\bar{x}) = 0$  has a solution over  $\mathbf{D}$ . This time, again using SESH, we will show that if a polynomial  $\mathbf{t}$  returns some value  $d \in D$ , i.e.  $\mathbf{t}^{-1}(d) \neq \emptyset$  then it actually returns this value many times, namely  $|\mathbf{t}^{-1}(d)| \geq \Omega\left(\frac{|D|^n}{2^{c \log^{h-1} |\mathbf{t}|}}\right)$ . Thus, randomly choosing sufficiently many tuples from  $D^n$ , say  $\Omega\left(2^{c \log^{h-1} |\mathbf{t}|}\right)$  many of them, with probability at least  $1/2$  we will find a solution to  $\mathbf{t}(\bar{x}) = d$ , if there is at least one. This algorithm works then in time  $O\left(2^{c \log^{h-1} |\mathbf{t}|}\right)$  which matches the complexity  $\Omega\left(2^{c \log^{h-1} |\mathbf{t}|}\right)$  of the lower bound provided in Subsection 3.2, but possibly with a different constant  $c$ .

We start with observing that replacing  $\mathbf{t}(\bar{x})$  by the polynomial  $\mathbf{t}(\bar{x}) - d$  we may assume that  $d = 0$ . Now starting with a single solution for the equation  $\mathbf{t}(\bar{x}) = 0$  we inductively create the sets  $T^1, \dots, T^h$  of solutions such that

$|T^k| \geq \Omega\left(\frac{p_1^n \dots p_k^n}{2^{c_k \log^{k-1} |\mathbf{t}|}}\right)$ . It should be obvious that our final set  $|T^h| \subseteq \mathbf{t}^{-1}(0)$  witnesses that the size of  $\mathbf{t}^{-1}(0)$  is big enough.

We parameterise the sets  $E^k$  defined in section 3.3 by tuples  $\bar{u} \in D^n$  simply putting

$$E^k(\bar{u}) = \left\{ \bar{b} \in D^n : e_j(\bar{b}) = e_j(\bar{u}) \text{ for all } j \neq k \right\}.$$

Then, inside  $E^k(\bar{u})$  we distinguish the subset

$$E_0^k(\bar{a}) = \left\{ \bar{b} \in E^k(\bar{u}) : \mathbf{t}(\bar{b}) = 0 \right\}$$

of solutions to our equation. Then we fix one solution tuple  $\bar{a} \in \mathbf{t}^{-1}(0)$  from which we will produce many other ones. To do that we put  $T^1 = E_0^1(\bar{a})$  and  $T^k = \bigcup_{\bar{u} \in T^{k-1}} E_0^k(\bar{u})$ . It should be clear that any tuple in all  $T^k$ 's is a solution to our equation. Thus, after showing that  $|E_0^k(\bar{u})| \geq \Omega\left(\frac{p_k^n}{2^{c_k \log^{k-1} |\mathbf{t}|}}\right)$  we get that the size of  $T^k$  is as big as promised, so that we can conclude our proof.

Despite of our relativization of the  $E^k$ 's to the  $E^k(\bar{u})$ 's (but keeping  $\bar{u}$  in the solution set  $\mathbf{t}^{-1}(0)$ ) we still know that as long as  $\bar{b} \in E^k(\bar{u})$  the fact that  $\mathbf{t}(\bar{b}) = 0$  can be replaced (as previously) by the system of only  $k$  equations  $e_1 \mathbf{t}(\bar{b}) = 0, \dots, e_k \mathbf{t}(\bar{b}) = 0$ , where the normal forms  $\mathbf{f}_j$  for  $e_j \mathbf{t}$  reduce accordingly as in (2). Thus to see that  $|E_0^1(\bar{a})| \geq p_1^{n-1}$  note only that  $E_0^1(\bar{a})$  consists of solutions to the linear equation  $0 = c^1 + \sum_{i=1}^n \lambda_i^1 e_1(b_i)$ .

Establishing the lower bound for  $E_0^k(\bar{u})$  is more laborious. We fix  $\bar{u}$  in  $T^{k-1}$  (or more generally in  $\mathbf{t}^{-1}(0)$ ) we repeat the procedure of section 3.3 to produce a relatively short (i.e. of size  $O(|\mathbf{t}|^{p_1})$ ) polynomial  $\mathbf{t}_u^*(\bar{x})$  of  $\mathbf{D}$  that maps everything to only two values 0,  $e_1 1$  and that depends only on  $e_k(\bar{x})$ , and – what is the most important – has the property that over the set  $E^k(\bar{u})$  the equations  $\mathbf{t}_u^*(\bar{b}) = e_1 1$  and  $\mathbf{t}(\bar{b}) = 0$  have exactly the same solutions.

As previously (in section 3.3) our goal is to rearrange polynomial  $\mathbf{t}_u^*$  to a polynomial  $\mathbf{t}_u^{**}$  that behaves on the set  $\{0, e_k 1\}$  like a conjunction and then apply SESH to the size of  $\mathbf{t}_u^{**}$  to bound its arity. On the way from  $\mathbf{t}_u^*$  to  $\mathbf{t}_u^{**}$  we create a polynomial  $\mathbf{t}_u^\dagger$ . To do that we refer to Lemma 3.3 (which is shown at the end of this section) with  $q = p_k$  and  $Z = (\mathbf{t}_u^*)^{-1}(e_1 1) \cap e_k D^n$  to get a hyperplane  $H \subseteq e_k D^n$  of codimension  $d \leq \log_{p_k} Z + p_k \log p_k$ . By Gauss elimination the set  $\{1, \dots, n\}$  can be split into two disjoint subsets  $I, J$  with  $|J| = d$  such that the hyperplane  $H$  can be described by  $d$  equations of the form  $x_j = \sum_{i \in I} \alpha_i^j x_i + \beta^j$ , with the  $\alpha$ 's taken from  $Z_{p_k}$ , while the  $\beta$ 's originally living in  $Z_{p_k}$  are modified so that they are put into  $e_k D$ . Now  $\mathbf{t}_u^\dagger$  is obtained from  $\mathbf{t}_u^*$  by replacing  $x_j$  with  $\sum_{i \in I} \alpha_i^j x_i + \beta^j$ . This slightly reduces the arity of  $\mathbf{t}_u^\dagger$  to be at least  $n - \log_{p_k} Z - p_k \log p_k$  but  $|\mathbf{t}_u^\dagger| \leq O(n \cdot |\mathbf{t}|) \leq O(|\mathbf{t}|^2)$ . However now the equation  $\mathbf{t}_u^\dagger(\bar{x}) = e_1 1$  has exactly one solution  $\bar{b} = (b_1, \dots, b_{n-d})$ , namely the

one corresponding to the unique point in the intersection  $Z \cap e_k D^n$ . To make sure that  $\mathbf{t}_u^{\star\star}(x_1, \dots, x_{n-d})$  behaves like a conjunction we put  $\mathbf{t}_u^{\star\star}(x_1, \dots, x_{n-d}) = \mathbf{t}_u^{\star\star}(x_1 - e_k 1 + b_1, \dots, x_{n-d} - e_k 1 + b_{n-d})$  and then turn  $\mathbf{t}_u^{\star\star}$  into a Boolean circuit of  $(n-d)$ -ary conjunction of size  $O(|t|^c)$  for some constant  $c$ . This, by SESH gives that  $\Omega\left(2^{c' \cdot (n-d)^{1/(k-1)}}\right) \leq O(|t|^c)$ , or in other words  $n - c \log^{k-1} |t| \leq d$ . To conclude with our lower bound for  $E^k(\bar{u})$  first note that this set fully corresponds to  $Z = (\mathbf{t}_u^{\star})^{-1}(e_1 1) \cap e_k D^n$  so that  $|E_0^k(\bar{u})| = |Z|$ . Summing up we get

$$n - c \log^{k-1} |t| \leq d \leq \log_{p_k} |Z| + c' = \log_{p_k} |E_0^k(\bar{u})| + c',$$

and consequently  $|E_0^k(\bar{u})| \geq \Omega\left(\frac{p_k^n}{2^{c \log^{k-1} |t|}}\right)$ , as required.

**Lemma 3.3.** *For a non-empty subset  $Z$  of the  $n$ -dimensional vector space  $GF(q)^n$  there is an affine subspace  $H$  of codimension at most  $\log_q |Z| + q \log_2 q$  such that  $|Z \cap H| = 1$ .*

*Proof.* We will successfully replace  $Z$  by  $Z \cap H$  where  $H$  at the start is  $GF(q)^n$ . As long as  $|Z| > q^{q-1}$  the set  $Z$  has to contain at least  $q$  linearly independent vectors, say  $w_1, \dots, w_q$ . Now for a  $q \times n$ -matrix  $W$  with rows  $w_1, \dots, w_q$  and the vector  $a = (a_1, \dots, a_q) \in GF(q)^q$  listing all elements of the field the system of equations  $W \cdot x = a$  has solutions, so that we pick one, say  $[\alpha_1, \dots, \alpha_n]$ . Consider  $q$  hyperplanes determined by the equations of the form  $\sum_i \alpha_i x_i = a_j$ . Note that each such hyperplane intersects  $Z$ , as  $w_j$  belongs to such intersection. Pick the one that leads to the intersection of the smallest size, and replace  $H$  by its intersection with this particular hyperplane. Note that  $Z \cap H$  has now at most  $\frac{|Z|}{q}$  elements.

At some point we will arrive with  $Z$  being too small to repeat this procedure. So, if  $|Z| \leq q^{q-1}$  but still  $|Z| \geq 2$  we pick a coordinate  $i_0$  such that  $Z$  contains at least two vectors that differ at this coordinate. This time we consider all  $q$  hyperplanes given by the equations  $x_{i_0} = a_j$  and pick one that non-empty intersects  $Z$  but this intersection is the smallest possible. Replace  $H$  with its intersection with this hyperplane. Since that are at least two hyperplanes non-empty intersecting  $Z$  we know that this time  $|Z \cap H| \leq \frac{|Z|}{2}$ .  $\square$

## 4 The group case

Both CSAT and CEQV are fully solved for groups. The problems are polynomial time solvable for nilpotent groups and NP/co-NP-complete otherwise. This is because a nilpotent groups are already supernilpotent. However, as we have already mentioned, equations solving (not compressed by circuits) may be still poly-time solvable – this in fact is the case of the non-nilpotent group  $S_3$ . Actually, there are much more such examples [7]. The smallest group for which the complexity is not known is the group  $S_4$ . The method used in Section 3 can be almost directly applied to provide an  $\Omega(m^{c \log m})$  lower bound for time complexity of solving

equations (POLSAT) and polynomials equivalence (POLEQV), where  $m$  is the size of the equation on input.

**Fact 4.1.** *The complexity of both POLSAT( $S_4$ ) and POLEQV( $S_4$ ) is  $\Omega(m^{c \log m})$ , where  $m$  is the size of input (unless ETH fails).*

*Proof.* Before we start with the proof we note that  $\{1\} < V < A_4 < S_4$  is the full sequence of normal subgroup of  $S_4$ , where  $V \simeq Z_2 \times Z_2$  is the Klein group and  $A_4$  is the alternating group. They correspond to the levels  $e_1(D)$ ,  $e_2(D)$  and  $e_3(D)$  of the algebra  $D$  from Section 3.

Below we summarize a few simple observations about the structure of  $S_4$  and its normal subgroups:

- $S_4/V \simeq S_3$ ,
- $[S_4, S_4] = A_4$ ,
- $[A_4, A_4] = V$
- $[V, V] = 1$ .
- for every  $a \in A_4 \setminus V$  we have that  $[V, a] = V$ ,

We will show lower bound for POLSAT( $S_4$ ). The proof for POLEQV is nearly the same. Let  $c \in V \setminus \{1\}$ . Analogously as in our construction in Section 3.2 we start with a 3-CNF formula  $\Phi$  (with  $m$  clauses) we construct  $\mathbf{t}_\Phi$  such that  $\Phi$  is satisfiable iff  $\mathbf{t}_\Phi(y_1, y_2, y_3, y_4, x_1, \dots, x_n) = c$  has a solution.

The construction of  $\mathbf{t}_\Phi$  is split into two steps. To imitate  $\text{AND}_s^1(x_1, \dots, x_s)$  we will use the  $(s+4)$ -ary terms

$$\alpha_s(y_1, y_2, y_3, y_4, x_1, \dots, x_s) = \alpha^\circ([y_1, y_2], [y_3, y_4], x_1, \dots, x_s),$$

where  $\alpha^\circ(y, x_1, \dots, x_s) = [[\dots [y, x_1], \dots], x_s]$ . Note that, independently of how  $y_1, y_2, y_3, y_4 \in S_4$  are chosen the value  $[[y_1, y_2], [y_3, y_4]]$  is in  $V$ . Moreover, any  $d \in V$  can be realized as  $[[y_1, y_2], [y_3, y_4]]$  for some  $y_1, y_2, y_3, y_4 \in S_4$ .

Now we divide the clauses of  $\Phi$  into  $s$  parts, each of which consist of at most  $s$  clauses, say  $\Phi_\ell$ 's, where  $s \leq \lceil \sqrt{m} \rceil$ . We will imitate  $\text{CNF}(\Phi_\ell)$  (with  $n_\ell$  variables) to code 3-CNF formula  $\Phi_\ell$ . To do that we borrow (e.g. from [10] or from [15]) the polynomial  $p_{\Phi_\ell}(x_1, \dots, x_{n_\ell})$  (of exponential size in  $n_\ell$ ) with range contained in  $A_4$  whose behavior on each tuple  $(x_1, \dots, x_{n_\ell}) \in S_4^{n_\ell}$  is, modulo  $V$ , fully determined by the behavior of the  $x_i$ 's modulo  $A_4$ . Namely  $p_{\Phi_\ell}(x_1, \dots, x_{n_\ell}) \in V$  iff  $\Phi(b(x_1), \dots, b(x_{n_\ell})) = \perp$ , where  $b : S_4 \rightarrow \{\perp, \top\}$  is given by  $b(x) = \top$  if  $x \in A_4$  and  $b(x) = \perp$  otherwise. Now we put  $\mathbf{t}_\Phi(\bar{y}, \bar{x})$  to be  $\alpha_k(y_1, y_2, y_3, y_4, p_{\Phi_1}(\bar{x}), \dots, p_{\Phi_s}(\bar{x}))$ .

Suppose  $\mathbf{t}_\Phi(\bar{y}, \bar{x}) = c$  for some  $y$ 's and  $x$ 's. Indeed the fact that  $[[y_1, y_2], [y_3, y_4]] \in V$  ensure us that none of the  $p_{\Phi_\ell}(\bar{x})$ 's might be  $V$ . Consequently for all the  $\ell$ 's we have  $\Phi_\ell(b(\bar{x})) = \top$  so that  $\Phi$  itself is satisfied while evaluated by  $b(\bar{x})$ .

Conversely, we translate a Boolean evaluation of the variables in  $\Phi$  by the  $z_i$ 's, to a corresponding evaluation of the  $x_i$ 's by elements of  $S_4$  so that we chose  $x_i \in A_4$  whenever  $z_i = \top$ , and all the other  $x_i$ 's are chosen from outside  $A_4$ . Obviously all the  $p_{\Phi_\ell}$ 's are then put inside  $A_4$  but outside  $V$ . We are left with finding values for the  $y_i$ 's. But, using the

fact that for  $[V, a] = V$  for any  $a \in A_4 - V$  and knowing that the  $p_{\Phi_r}(\bar{x})$ 's are in this difference, we find  $u \in V$  so that

$$\alpha^\circ(u, p_{\Phi_1}(\bar{x}), \dots, p_{\Phi_s}(\bar{x})) = c.$$

Now, this  $u$  can be decomposed into  $u = [[y_1, y_2], [y_3, y_4]]$  for some  $y_1, \dots, y_4 \in S_4$ .

Finally we refer to ETH and argue like at the beginning of Section 3.2 to get the promised lower bound for equation solution in  $S_4$ .  $\square$

Solvable but not non-nilpotent gap in equation solving for groups is open for about 20 years since Goldmann's and Russel's paper [9]. Since then, a lot of effort has been put into finding new classes of solvable but non-nilpotent groups for which PolSAT and PolEQV are in P (e.g. [14], [13], [7]). The group  $S_4$  is now the first known example of a solvable but non-nilpotent group for which probably do not exist polynomial time algorithms solving these problems. Moreover, our method used in the proof of Fact 4.1 is quite general and can be used for showing lower bounds for other groups or even for other solvable but non-nilpotent algebras from congruence modular varieties.

In fact, very recently Armin Weiß presented a proof [26] that under ETH neither PolSAT nor PolEQV can be in P for the solvable groups that are not 3-step supernilpotent (or even not 2-step supernilpotent, but with an additional technical assumption). Note here that the concept of the  $h$ -step supernilpotency in groups coincides with the one of the Fitting length  $h$ . Combining his and our efforts now we can remove this artificial technical assumption and actually strengthen the lower bound to be read:

*If  $G$  is a finite solvable nonnilpotent group of Fitting length  $h > 2$  then both PolSAT( $G$ ) and PolEQV( $G$ ) require at least  $O(2^{c \log^{h-1} m})$  steps, where  $m$  is the length of the polynomial(s) on input (unless ETH fails).*

## 5 Conclusions

We propose a couple of methods that are highly effective in filling the nilpotent versus supernilpotent gap for the problems CSAT and CEQV, but with the help of two strong complexity hypothesis. Our methods are particularly effective for  $h$ -step supernilpotent algebras for  $h \geq 3$ . However these methods do not fully solve the problems for 2-step supernilpotent algebras (as they lead only a probabilistic upper bound, and this bound relies on SESH).

Since supernilpotent algebras do already have polynomial time algorithms for both CSAT and CEQV, it seems that the 2-step supernilpotent ones form the natural next step to be attacked. All the known to us examples of such algebras, including the  $D[p, q]$ 's, lie on the polynomial side (without any additional complexity hypothesis). Moreover [18] contains a proof that CEQV for 2-nilpotent algebras is in P. As 2-nilpotent algebras are 2-step supernilpotent this still leaves the hope that the last ones also lie on the polynomial side.

Also, Theorem 1.3 provides a polynomial randomized upper bound for  $h = 2$ . This makes our hope even stronger.

On the other hand we do hope that the boundary between tractable and hard algebras is determined by this new measure of failure of the supernilpotency, as there are examples [17] of 3-nilpotent but not 2-nilpotent algebras with polynomially solvable CSAT and CEQV. They are 2-step supernilpotent.

In our second remark we note that both our algorithms can be parameterized by the (lower) bound for the conjunction-like polynomials or  $CC^0$ -circuits. If the lower bound provided by SESH is replaced by a computable but slower growing function  $f(n)$  Then our method gives

- a deterministic algorithm of complexity  $O(n^{c \cdot f^{-1}(|t|^d)})$ ,
- a randomized algorithm of complexity  $O(2^{c \cdot f^{-1}(|t|^d)})$ ,

where  $|t|$  is the size of polynomial or a circuit,  $n$  is the number of variables (or input gates), and  $c, d$  are some constants. This shows a very strong connections between the complexity of CSAT and CEQV and the size in which conjunctions can be expressed by  $CC^0$ -circuits (or polynomials).

In particular if  $f(n) \geq 2^{cn}$  for some  $c > 0$  (which is true in  $D[p, q]$ , i.e. in the case of  $h = 2$ ) then from what we said above the proof of Theorem 1.3 supplies us with a polynomial time randomized algorithms. In the case  $h = 1$ , i.e. for supernilpotent algebras, there is even no such function  $f$ , as there is a bound for the arity of polynomials that expresses conjunction-like behavior. In this case we can slightly modify the method used in the proof of Theorem 1.3 to get linear algorithms for CSAT and CEQV. In particular (as nilpotent groups are supernilpotent) we get a striking division between untractable (NP/co-NP-complete) non-nilpotent groups and the nilpotent ones that can be treated in probabilistic linear time [19].

The other feature provided by our proof of Theorem 1.3 tells us that a short polynomial splits its domain into rather large subsets on which it is constant. In particular it is not possible to separate, by polynomials, not only single points (what is usually done by a conjunction-like function) but even larger subsets in the big powers of the algebra.

## Acknowledgments

The project is partially supported by Polish NCN Grant # 2014/14/A/ST6/00138

## References

- [1] Erhard Aichinger and Nebojša Mudrinski, Some applications of higher commutators in Malcev algebras, *Algebra Universalis*, **63**(2010), 367–403.
- [2] David A. Mix Barrington, Richard Beigel and Steven Rudich, Representing Boolean functions as polynomials modulo composite numbers, *Computational Complexity*, **4**(1994), 367–382.
- [3] David A. Mix Barrington, Howard Straubing and Denis Thérien, Non-uniform automata over groups, *Information and Computation*, **89**(1990), 109–132.



- [4] Andrei Bulatov, A dichotomy theorem for nonuniform CSPs, *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, 2017.
- [5] Andrei Bulatov, On the number of finite Mal'tsev algebras, *Contributions to general algebra*, **13**(2000) 41–54.
- [6] Ralph Freese and Ralph McKenzie, *Commutator Theory for Congruence Modular Varieties*, London Math. Soc. Lecture Notes, No. 125, Cambridge U. Press, Cambridge, 1987.
- [7] Attila Földvári and Gábor Horváth, The complexity of the equation solvability and equivalence problems over finite groups *International Journal of Algebra and Computation*, to appear.
- [8] Merrick Furst, James B. Saxe and Michael Sipser, Parity circuits and the polynomial time hierarchy, *Proceedings of 22nd Annual IEEE Symposium on Foundations of Computer Science*, 1981, pp. 260–270.
- [9] Mikael Goldmann and Alexander Russell, The complexity of solving equations over finite groups, *Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity*, 1999, pp. 80–86.
- [10] Tomasz Gorazd and Jacek Krzaczkowski, Term equation satisfiability over finite algebras, *International Journal of Algebra and Computation*, **20**(2010), 1001–1020.
- [11] Johan Torkel Hästad, Computational limitations of small-depth circuits, *PhD thesis*, MIT, 1987.
- [12] David Hobby and Ralph McKenzie, *The Structure of Finite Algebras*, Contemporary Mathematics vol. 76, Amer. Math. Soc., Providence, RI, 1988.
- [13] Gábor Horváth, The complexity of the equivalence and equation solvability problems over meta-Abelian groups, *Journal of Algebra*, **433**(2015), 208–230.
- [14] Gábor Horváth and Csaba Szabó, The complexity of checking identities over finite groups, *International Journal of Algebra and Computation*, **16**(2006), 931–940.
- [15] Paweł M. Idziak and Jacek Krzaczkowski, Satisfiability in multi-valued circuits, *LICS '18: Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, 2018, pp. 550–558.
- [16] Paweł M. Idziak, Piotr Kawałek and Jacek Krzaczkowski, Expressive power, satisfiability and equivalence of circuits over nilpotent algebras, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- [17] Paweł M. Idziak, Piotr Kawałek and Jacek Krzaczkowski, Stratifying algebras by supernilpotent intervals, manuscript 2020.
- [18] Piotr Kawałek, Michael Kompatscher and Jacek Krzaczkowski, Circuit equivalence in 2-nilpotent algebras, preprint arXiv:1909.12256, (2019).
- [19] Piotr Kawałek and Jacek Krzaczkowski, Even faster algorithms for circuit satisfiability over supenilpotent algebras, preprint arXiv:2002.08634, (2020).
- [20] Keith A. Kearnes, Congruence modular varieties with small free spectra, *Algebra Universalis*, **42**(1999), 165–181.
- [21] Michael Kompatscher, CC-circuits and the expressive power of nilpotent algebras, preprint arXiv:1911.01479, (2019).
- [22] Peter Mayr and Agnes Szendrei, Algebras from Congruences, preprint arXiv:1910.00689 (2019).
- [23] Ralph McKenzie, George McNulty and Walter Taylor, *Algebras, Lattices, Varieties*, Wadsworth/Brooks Cole, Monterrey, CA, 1987.
- [24] Michael Sipser, Borel Sets and circuit complexity, *Proceedings of 15th Annual ACM Symposium on Theory of Computing* 1983, pp. 61–69.
- [25] Joel WanderWerf, *Wreath Decomposition of Algebra*, PhD thesis, University of California, Berkeley 1995.
- [26] Armin Weiß, Hardness of equations over finite solvable groups under the exponential time hypothesis, *The 47th International Colloquium on Automata, Languages and Programming (ICALP 2020)*, to appear.
- [27] Andrew Yao, Separating the polynomial-time hierarchy by oracles, *1985 IEEE 26th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, 1985.
- [28] Dmitriy Zhuk, A proof of CSP dichotomy conjecture, *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, 2017.