Journal of Mathematical Sciences, Vol. 193, No. 4, September, 2013

SUBEXPONENTIAL ESTIMATES IN THE HEIGHT THEOREM AND ESTIMATES ON NUMBERS OF PERIODIC PARTS OF SMALL PERIODS

A. Ya. Belov and M. I. Kharitonov UDC 512.5+512.64+519.1

Abstract. This paper is devoted to subexponential estimates in Shirshov's height theorem. A word W is *n-divisible* if it can be represented in the form $W = W_0W_1 \cdots W_n$, where $W_1 \prec W_2 \prec \cdots \prec W_n$. If an affine algebra A satisfies a polynomial identity of degree n, then A is spanned by non-n-divisible words of generators $a_1 \prec \cdots \prec a_l$. A. I. Shirshov proved that the set of non-n-divisible words over an alphabet of cardinality l has bounded height h over the set Y consisting of all words of degree $\leq n-1$. We show that $h < \Phi(n,l)$, where $\Phi(n,l) = 2^{87} l \cdot n^{12\log_3 n+48}$. Let l, n, and $d \geq n$ be positive integers. Then all words over an alphabet of cardinality l whose length is greater than $\Psi(n,d,l)$ are either n-divisible or contain the dth power of a subword, where $\Psi(n,d,l) = 2^{18} l (nd)^{3\log_3(nd)+13} d^2$. In 1993, E. I. Zelmanov asked the following question in the Dniester Notebook: Suppose that $F_{2,m}$ is a 2-generated associative ring with the identity $x^m = 0$. Is it true that the nilpotency degree of $F_{2,m}$ has exponential growth? We give the definitive answer to E. I. Zelmanov by this result. We show that the nilpotency degree of the l-generated associative algebra with the identity $x^d = 0$ is smaller than $\Psi(d,d,l)$. This implies subexponential estimates on the nilpotency index of nil-algebras of arbitrary characteristic. Shirshov's original estimate was just recursive; in 1982 a double exponent was obtained, and an exponential estimate was obtained in 1992. Our proof uses Latyshev's idea of an application of the Dilworth theorem. We think that Shirshov's height theorem is deeply connected to problems of modern combinatorics. In particular, this theorem is related to the Ramsey theory. We obtain lower and upper estimates of the number of periods of length 2, 3, $n-1$ in some non-n-divisible word. These estimates differ only by a constant.

1. Introduction

1.1. Shirshov's Theorem on Height. In 1958, A. I. Shirshov proved his famous theorem on height [36,37].

Definition 1.1. A word W is called *n-divisible* if W can be represented in the form $W = vu_1u_2 \cdots u_n$ such that $u_1 \succ u_2 \succ \cdots \succ u_n$.

In this case, any nonidentical permutation σ of subwords u_i produces a word $W_\sigma = vu_{\sigma(1)}u_{\sigma(2)} \cdots u_{\sigma(n)}$ that is lexicographically smaller than W. Some authors take this feature as the definition of n-divisibility.

Definition 1.2. A PI-algebra A is called an algebra *of bounded height $h = \mathrm{Ht}_Y(A)$ over a set of words $Y = \{u_1, u_2, \ldots\}$* if h is the minimal integer such that any word x from A can be represented in the form

$$x = \sum_i \alpha_i u_{j_{(i,1)}}^{k_{(i,1)}} u_{j_{(i,2)}}^{k_{(i,2)}} \cdots u_{j_{(i,r_i)}}^{k_{(i,r_i)}},$$

where $\{r_i\}$ do not exceed h. The set Y is called a *Shirshov basis* for A. If no misunderstanding can occur, then we use h instead of $\mathrm{Ht}_Y(A)$.

Shirshov's theorem on height ([36,37]). *The set of non-n-divisible words in a finitely generated algebra with an admissible polynomial identity has bounded height H over the set of words of degree not exceeding $n-1$.*

Translated from Fundamentalnaya i Prikladnaya Matematika, Vol. 17, No. 5, pp. 21–54, 2011/12.

The Burnside-type problems related to the height theorem are considered in [41]. The authors believe that Shirshov's theorem on height is a fundamental fact in word combinatorics independently of its applications to PI-theory. (All our proofs are elementary and fit in the framework of word combinatorics.) Unfortunately, the experts in combinatorics have not sufficiently appraised this fact yet. As regards the notion of *n-divisibility* itself, it seems to be fundamental as well. V. N. Latyshev's estimates on $\xi_n(k)$, the number of non-n-divisible polylinear words in k symbols, have led to fundamental results in PI-theory. At the same time, this number is nothing but the number of arrangements of integers from 1 to k such that no n integers (not necessarily consecutive) are placed in decreasing order. Furthermore, it is the number of permutationally ordered sets of diameter n consisting of k elements. (A set is called *permutationally ordered* if its ordering is the intersection of two linear orderings; the *diameter* of an ordered set is the length of its maximal antichain.)

The height theorem implies the solution of a set of problems in ring theory. Suppose an associative algebra over a field satisfies a polynomial identity $f(x_1,\dots,x_n)=0$. It is possible to prove that then it satisfies an admissible polynomial identity (that is, a polynomial identity with coefficient 1 at some term of the highest degree)

$$x_1x_2\cdots x_n=\sum_{\sigma}\alpha_\sigma x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(n)},$$

where α_σ belongs to the ground field. In this case, if $W=vu_1u_2\cdots u_n$ is n-divisible, then for any permutation σ the word $W_\sigma=vu_{\sigma(1)}u_{\sigma(2)}\cdots u_{\sigma(n)}$ is lexicographically smaller than W, and thus an n-divisible word can be represented as a linear combination of lexicographically smaller words. Hence a PI-algebra has a basis consisting of non-n-divisible words. By Shirshov's theorem on height, a PI-algebra has bounded height. In particular, if a PI-algebra satisfies $x^n=0$, then it is nilpotent, that is, any of its words of length exceeding some N is identically zero. Surveys on the height theorem can be found in [6, 18–20, 39].

This theorem implies the positive solution of the Kurosh problem and of other Burnside-type problems for PI-rings. Indeed, if Y is a Shirshov basis and all its elements are algebraic, then the algebra A is finite-dimensional. Thus, Shirshov's theorem explicitly indicates a set of elements whose algebraicity makes the whole algebra finite-dimensional. This theorem implies the following corollary.

Corollary 1.3 (A. Berele)**.** *Let A be a finitely generated PI-algebra. Then*

$$\mathrm{GK}(A)<\infty.$$

$\mathrm{GK}(A)$ is the *Gelfand–Kirillov dimension of the algebra A*, that is,

$$\mathrm{GK}(A)=\lim_{n\to\infty}\frac{\ln V_A(n)}{\ln(n)},$$

where $V_A(n)$ is the *growth function of A*, the dimension of the vector space generated by words of degree not greater than n in the generators of A.

Indeed, it suffices to observe that the number of solutions for the inequality

$$k_1|v_1|+\dots+k_h|v_h|\le n$$

with $h\le H$ exceeds N^H, so that $\mathrm{GK}(A)\le\mathrm{Ht}(A)$.

The number $m=\deg(A)$ will mean the *degree of the algebra*, or the minimal degree of an identity valid in it. The number $n=\mathrm{Pid}(A)$ is the *complexity* of A, or the maximal k such that $\mathbb{M}_k$, the algebra of matrices of size k, belongs to the variety $\mathrm{Var}(A)$ generated by A.

Instead of the notion of *height*, it is more suitable to use the close notion of *essential height*.

Definition 1.4. An algebra A has *essential height* $h=H_{\mathrm{Ess}}(A)$ over a finite set Y called an *s-basis for A* if there exists a finite set $D\subset A$ such that A is linearly representable by elements of the form $t_1\cdot\ldots\cdot t_l$, where $l\le 2h+1$, for all i, $t_i\in D$ or $t_i=y_i^{k_i}$, $y_i\in Y$, and the set of i such that $t_i\notin D$ contains at most h elements. The *essential height* of a set of words is defined similarly.

Informally speaking, any long word is a product of periodic parts and "gaskets" of restricted length. The essential height is the number of periodic parts, and the ordinary height accounts "gaskets" as well.

The height theorem suggests the following questions.

(1) To which classes of rings can the height theorem be extended?
(2) Over which Y has the algebra A bounded height? In particular, what sets of words can be taken for $\{v_i\}$?
(3) What is the structure of the degree vector $(k_1, \ldots, k_h)$? First of all, what sets of its components are essential, that is, what sets of k_i can be unbounded simultaneously? What is the value of essential height? Is it true that the set of degree vectors has some regularity properties?
(4) What estimates for the height are possible?

Let us discuss the above questions.

1.2. Nonassociative Generalizations. The height theorem was extended to some classes of near-associative rings. S. V Pchelintsev [33] has proved it for the alternative and the $(-1, 1)$ cases, and S. P. Mishchenko [32] has obtained an analogue of the height theorem for Lie algebras with a sparse identity. In [1], the height theorem was proved for some class of rings asymptotically close to associative rings. In particular, this class contains alternative and Jordan PI-algebras.

1.3. Shirshov Bases. Suppose that A is a PI-algebra and a subset $M \subseteq A$ is its s-basis. Then if all elements of M are algebraic over K, A is finite-dimensional (the Kurosh problem). Boundedness of essential height over Y implies "the positive solution of the Kurosh problem over Y." The converse is less trivial.

Theorem 1.5 (A. Ya. Belov)**.**

(1) *Suppose that A is a graded PI-algebra and Y is a finite set of homogeneous elements. Then if for all n the algebra $A/Y^{(n)}$ is nilpotent, then Y is an s-basis for A. If, moreover, Y generates A as an algebra, then Y is a Shirshov basis for A.*
(2) *Suppose that A is a PI-algebra and $M \subseteq A$ is a Kurosh subset in A. Then M is an s-basis for A.*

Let $Y^{(n)}$ denote the ideal generated by nth powers of elements from Y. A set $M \subset A$ is called a *Kurosh set* if any projection $\pi\colon A \otimes K[X] \to A'$ such that the image $\pi(M)$ is entire over $\pi(K[X])$ is finite-dimensional over $\pi(K[X])$. The following example motivates this definition. Suppose that $A = \mathbb{Q}[x, 1/x]$. Any projection π such that $\pi(x)$ is algebraic has a finite-dimensional image. However, the set $\{x\}$ is not an s-basis for $\mathbb{Q}[x, 1/x]$. Thus, boundedness of essential height is a noncommutative generalization of the property of *entireness.*

The Shirshov bases consisting of words are described by the following theorem.

Theorem 1.6 ([6,8])**.** *A set Y of words is a Shirshov basis for an algebra A if and only if for any word u of length not exceeding $m = \mathrm{Pid}(A)$, the complexity of A, the set Y contains a word cyclically conjugate to some power of u.*

A similar result was obtained independently by G. P. Chekanu and V. Drensky. Problems related to local finiteness of algebras and to algebraic sets of words of degree not exceeding the complexity of the algebra were investigated in [11–13, 15, 38–40]. Questions related to generalization of the independence theorem were considered in these papers as well.

1.4. Essential Height. Clearly, the Gelfand–Kirillov dimension is estimated by the essential height. Furthermore, an s-basis is a Shirshov basis if and only if it generates A as an algebra. In the representable case, the converse is also true.

Theorem 1.7 (A. Ya. Belov [6])**.** *Suppose that A is a finitely generated representable algebra and $H_{\mathrm{Ess}Y}(A) < \infty$. Then $H_{\mathrm{Ess}Y}(A) = \mathrm{GK}(A)$.*

Corollary 1.8 (V. T. Markov)**.** *The Gelfand–Kirillov dimension of a finitely generated representable algebra is an integer.*

Corollary 1.9. *If $H_{\mathrm{Ess}Y}(A) < \infty$ and A is representable, then $H_{\mathrm{Ess}Y}(A)$ is independent of choice of the s-basis Y.*

In this case, the Gelfand–Kirillov dimension also is equal to the essential height by virtue of local representability of relatively free algebras.

Although in the representable case the Gelfand–Kirillov dimension and the essential height behave well, even in this case the set of degree vectors may have a bad structure, namely, it can be the complement to the set of solutions of a system of exponential-polynomial Diophantine equations [6]. That is why there exists an instance of a representable algebra with the transcendent Hilbert series. However, for a relatively free algebra, the Hilbert series is rational [3].

1.5. n-Divisibility and Dilworth Theorem. The significance of the notion of *n-divisibility* exceeds the limits of Burnside-type problems. This notion is also actual in investigation of polylinear words and estimation of their number (a word is *polylinear* if each letter occurs in it at most one time). V. N. Latyshev applied the Dilworth theorem for estimation of the number of non-m-divisible polylinear words of degree n over the alphabet $\{a_1, \dots, a_n\}$ (see [28]). The estimate is $(m-1)^{2n}$ and is rather sharp. Let us recall this theorem.

Dilworth theorem. *Let n be the maximal number of elements of an antichain in a given finite partially ordered set M. Then M can be divided into n disjoint chains.*

Consider a polylinear word W consisting of n letters. Put $a_i \succ a_j$ if $i > j$ and the letter a_i is located in W to the right from a_j. The condition of non-k-divisibility means the absence of an antichain consisting of n elements. Then by the Dilworth theorem all positions (and the letters a_i as well) split into $n-1$ chains. Attach a specific color to each chain. Then the coloring of positions and of letters uniquely determines the word W. Furthermore, the number of these colorings does not exceed $(n-1)^k \times (n-1)^k = (n-1)^{2k}$.

The above estimate implies the validity of polylinear identities corresponding to an irreducible module whose Young diagram includes the square of size n^4. This, in turn, enables one, firstly, to obtain a transparent proof of the Regev theorem, which asserts that a tensor product of PI-algebras is a PI-algebra as well; secondly, to establish the existence of a sparse identity in the general case and of a Capelli identity in the finitely generated case (and, thus, to prove the theorem on nilpotency of the radical); and, thirdly, to realize A. R. Kemer's "supertrick" that reduces the study of identities in general algebras to that of super-identities in finitely generated superalgebras of zero characteristic. Close questions are considered in [9, 31].

Problems related to the enumeration of polylinear words that are not n-divisible are interesting on their own. (For example, there exists a bijection between non-3-divisible words and Catalan numbers.) On the one hand, this is a purely combinatorial problem, but on the other hand, it is related to the set of codimensions for the general matrix algebra. The study of polylinear words seems to be of great importance. V. N. Latyshev (see, for instance, [29]) has stated the problem of finite-basedness of the set of leading polylinear words for a T-ideal with respect to taking overwords and to isotonous substitutions. This problem implies the Specht problem for polylinear polynomials and is closely related to the problem of weak Noetherian property for the group algebra of an infinite finitary symmetric group over a field of positive characteristic (for zero characteristic this was established by A. Zalessky). To solve the Latyshev problem, it is necessary to translate properties of T-ideals to the language of polylinear words. In [1, 6], an attempt was made to realize a project of translation of structure properties of algebras to the language of word combinatorics. Translation to the language of polylinear words is simpler and enables one to get some information on words of a general form.

In this paper, we transfer V. N. Latyshev's technique to the nonpolylinear case, and this enables us to obtain a subexponential estimate in Shirshov's height theorem. G. R. Chelnokov suggested the idea of this transfer in 1996.

1.6. Estimates for height. The original A. I. Shirshov's proof, being purely combinatorial (it was based on the technique of elimination developed by him for Lie algebras, in particular, in the proof of the theorem on freeness), nevertheless implied only primitively recursive estimates. Later A. T. Kolotov [26] obtained an estimate $\mathrm{Ht}(A) \leq l^{l^n}$ ($n = \deg(A)$, l is the number of generators). A. Ya. Belov in [2] has shown that $\mathrm{Ht}(n,l) < 2nl^{n+1}$. The exponential estimate in Shirshov's height theorem was also presented in [8,17,23]. The above estimates were sharpened in the papers by A. Klein [24,25]. In 2001, Ye. S. Chibrikov proved in [14] that $\mathrm{Ht}(4,l) \geq (7k^2 - 2k)$. M. I. Kharitonov in [21–23] obtained lower and upper estimates for the structure of piecewise periodicity. In 2011, A. A. Lopatin [30] obtained the following result.

Theorem 1.10. *Let $C_{n,l}$ be the nilpotency degree of a free l-generated algebra satisfying $x^n = 0$, and let p be the characteristic of the ground field of the algebra, greater than $n/2$. Then*

$$C_{n,l} < 4 \cdot 2^{n/2} l. \tag{1}$$

By definition, $C_{n,l} \leq \Psi(n,n,l)$. Observe that for small n estimate (1) is smaller than the estimate $\Psi(n,n,l)$ established in this paper, but for growing n the estimate $\Psi(n,n,l)$ is asymptotically better than (1).

Ye. I. Zelmanov has put the following question in the Dniester Notebook [16] in 1993.

Question 1.11. Let $F_{2,m}$ be the free 2-generated associative ring with identity $x^m = 0$. Is it true that the nilpotency class of $F_{2,m}$ grows exponentially in m?

Our paper answers Ye. I. Zelmanov's question as follows: the nilpotency class in question grows subexponentially.

1.7. The Results Obtained. The main result of the paper is as follows.

Theorem 1.12. *The height of the set of non-n-divisible words over an alphabet of cardinality l relative to the set of words of length less than n does not exceed $\Phi(n,l)$, where*

$$\Phi(n,l) = E_1 l \cdot n^{E_2 + 12\log_3 n},$$

$E_1 = 4^{21\log_3 4 + 17}$, $E_2 = 30\log_3 4 + 10$.

This theorem after some coarsening and simplification of the estimate implies that for fixed l and $n \to \infty$ we have

$$\Phi(n,l) < 2^{87} l \cdot n^{12\log_3 n + 48} = n^{12(1+o(1))\log_3 n},$$

and for fixed n and $l \to \infty$ we have

$$\Phi(n,l) < C(n) l.$$

The reader can also find the proof of Theorem 1.12 in [7].

Corollary 1.13. *The height of an l-generated PI-algebra with an admissible polynomial identity of degree n over the set of words of length less than n does not exceed $\Phi(n,l)$.*

Moreover, we prove a subexponential estimate that is better for small n.

Theorem 1.14. *The height of the set of non-n-divisible words over an alphabet of cardinality l relative to the set of words of length less than n does not exceed $\Phi(n,l)$, where*

$$\Phi(n,l) = 2^{40} l \cdot n^{38 + 8\log_2 n}.$$

In particular, we obtain subexponential estimates for the nilpotency index of l-generated nil-algebras of degree n of arbitrary characteristic.

The second main result of our paper is the following.

Theorem 1.15. *Let l, n, and $d \geq n$ be positive integers. Then all l-generated words of length not less than $\Psi(n,d,l)$ either contain x^d or are n-divisible. Here*

$$\Psi(n,d,l) = 4^{5+3\log_3 4} l (nd)^{3\log_3(nd) + (5 + 6\log_3 4)} d^2.$$

This theorem after some coarsening and simplification of the estimate implies that for fixed l and $nd \to \infty$ we have that

$$\Psi(n,d,l) < 2^{18} l (nd)^{3\log_3(nd)+13} d^2 = (nd)^{3(1+o(1))\log_3(nd)},$$

and for fixed n and $l \to \infty$ we have

$$\Psi(n,d,l) < C(n,d)l.$$

Corollary 1.16. *Let l and d be positive integers, and let an associative l-generated algebra A satisfy the identity $x^d = 0$. Then its nilpotency index is less than $\Psi(d,d,l)$.*

Moreover, we prove a subexponential estimate that is better for small n and d.

Theorem 1.17. *Let l, n, and $d \geq n$ be positive integers. Then all l-generated words of length not less than $\Psi(n,d,l)$ either contain x^d or are n-divisible. Here*

$$\Psi(n,d,l) = 256 l (nd)^{2\log_2(nd)+10} d^2.$$

Notation 1.18. For a real number x put $\ulcorner x \urcorner := -[-x]$. Thus, we replace noninteger numbers by the closest greater integers.

Proving Theorem 1.12, we also prove the following theorem on estimation of the essential height.

Theorem 1.19. *The essential height of an l-generated PI-algebra with an admissible polynomial identity of degree n over the set of words of length less than n is less than $\Upsilon(n,l)$, where*

$$\Upsilon(n,l) = 2n^{3\ulcorner\log_3 n\urcorner+4} l.$$

It is necessary to introduce the following definitions.

Definition 1.20. A word u is *acyclic* or *noncyclic* if u cannot be adduced as v^k for any word v and $k > 1$.

Definition 1.21. A word u and all of its cycle shifts is a *word-cycle* u.

Definition 1.22. A word is *strongly n-divisible* or *n-s-divisible over a set of words Z* if it can be adduced as $W = W_0 W_1 \cdots W_n$, where the subwords $W_1, \ldots, W_n$ are in decreasing lexicographical order and every word W_i, $i = 1, 2, \ldots, n$, begins from a word z_i^k. The words z_i are distinct and belong Z.

Then we prove the following lower and upper bounds of particular periodicity.

Theorem 1.23. *Let M be the set of non-n-divisible words with finite essential height over words of degree 2. Then the number of different lexicographically comparable acyclic words with period 2 in any word from M is less than $\beth(2,l,n)$, where*

$$\beth(2,l,n) = \frac{(2l-1)(n-1)(n-2)}{2}.$$

Theorem 1.24. *Let M be the set of non-n-s-divisible words over Z, the set of acyclic words of degree 2. Then the essential height of M over Z is bigger than $\beth'(2,n,l)$, where*

$$\beth'(2,n,l) = \frac{n^2 l}{2}\big(1 - o(l)\big).$$

More precisely,

$$\beth'(2,n,l) = \frac{(l-2^{n-1})(n-2)(n-3)}{2}.$$

Theorem 1.25. *Let M be the set of non-n-divisible words with finite essential height over the words of degree 3. Then the number of different lexicographically comparable acyclic words with period 3 in any word from M is less than $\beth(3,l,n)$, where*

$$\beth(3,l,n) = (2l-1)(n-1)(n-2).$$

One can find the proofs of Theorems 1.23–1.25 in [22].

Theorem 1.26. *Let M be the set of non-n-divisible words with finite essential height over the words of degree $n-1$. Then the number of different lexicographically comparable acyclic words with period $n-1$ in any word from M is less than $\beth(n-1,l,n)$, where*

$$\beth(n-1,l,n)=(l-2)(n-1).$$

We can get an exponential estimate of the essential height using Theorem 1.23 and the idea of encoding.

Theorem 1.27. *Denote by $\Upsilon(n,l)$ the essential height over the set of words of degree less than n of a relatively free algebra generated by l elements and satisfying the polynomial identity of degree n. Then*

$$\Upsilon(n,l)=8(l+1)^n n^5(n-1).$$

One can find the proofs of Theorems 1.26 and 1.27 in [23].

The condition of difference of the subwords in Theorems 1.23, 1.25, and 1.26 is removed in the following theorem.

Theorem 1.28. *Let M be the set of non-n-divisible words with finite essential height over the words of degree k. Then the number of lexicographically comparable acyclic words with period k in any word from M is less than $2(n-1)\beth(k,l,n)$, where*

$$\beth(n-1,l,n)=(l-2)(n-1),$$

where $\beth(k,n,l)$ is the maximal number of different lexicographically comparable acyclic words with period k in any word from M.

In [10], it is established that the nilpotency index of an l-generated nil-semiring of degree n equals the nilpotency index of an l-generated nilring of degree n, where addition is not supposed to be commutative. (The paper also contains examples of nonnilpotent nil-nearrings of index 2.) Thus, our results extend to the case of semirings as well.

1.8. On Estimates from Below.

Let us compare the results obtained with the estimate for height from below. The height of an algebra A is not less than its Gelfand–Kirillov dimension $\mathrm{GK}(A)$. For the algebra of l-generated general matrices of order n, this dimension equals $(l-1)n^2+1$ (see [4, 34]). At the same time, the minimal degree of an identity in this algebra is $2n$ by the Amitsur–Levitsky theorem.

Proposition 1.29. *The height of an l-generated PI-algebra of degree n and of the set of non-n-divisible words over an alphabet of cardinality l is not less than $(l-1)n^2/4+1$.*

Estimates from below for the nilpotency index were established by Ye. N. Kuzmin in [27]. He gave an example of a 2-generated algebra with identity $x^n=0$ such that its nilpotency index exceeds $(n^2+n-2)/2$. The problem of finding estimates from below is considered in [22].

At the same time, for zero characteristic and a countable set of generators, Yu. P. Razmyslov (see, for instance, [35]) obtained an upper estimate for the nilpotency index, namely n^2.

First, we will prove Theorem 1.15, and in the following section we will deal with estimates for the essential height, that is, for the number of distinct periodical pieces in a non-n-divisible word. By the end we evaluate the ways to upgrade the obtained estimates in the future.

Acknowledgements. The authors are grateful to V. N. Latyshev, A. V. Mikhalev and all participants of the seminar "Ring theory" for their attention to this work, as well as to the participants of the seminar under the guidance of A. M. Raigorodsky.

2. Estimates on Occurrence of Degrees of Subwords

2.1. The Outline of the Proof of Theorem 1.15. Lemmas 2.3, 2.4, and 2.6 describe sufficient conditions for the presence of a period of length d in a non-n-divisible word W. Lemma 2.7 connects n-divisibility of a word W with the set of its tails. Further we choose some specific subset in the set of tails of W such that we can apply the Dilworth theorem. After that we color the tails and their first letters according to their location in chains obtained by application of the Dilworth theorem.

We must know the position in any chain where neighboring tails begin to differ. It is of interest what is the "frequency" of this position in a p-tail for some $p \leq n$. Further we somewhat generalize our reasoning dividing tails into segments consisting of several letters each and determining the segment containing the position where neighboring tails begin to differ. Lemma 3.4 connects the "frequencies" in question for p-tails and kp-tails for $k = 3$.

To complete the proof, we construct a hierarchical structure based on Lemma 3.4, that is, we consecutively consider segments of n-tails, subsegments of these segments, and so on. Furthermore, we consider the greatest possible number of tails in the subset to which the Dilworth theorem is applied, and then we estimate from above the total number of tails and hence of letters in the word W.

2.2. Periodicity and n-Divisibility Properties. Let $a_1, a_2, \ldots, a_l$ be the alphabet used for constructing words. The ordering $a_1 \prec a_2 \prec \cdots \prec a_l$ induces lexicographical ordering for words over the alphabet. For convenience, we introduce the following definitions.

Definition 2.1.

(1) If a word v includes a subword of the form u^t, then we say that v includes a *period of length* t.
(2) If a word u is a prefix of a word v, then these words are called *incomparable*.
(3) A word v is a *tail* of a word u if there exists a word w such that $u = wv$.
(4) A word v is a k*-tail* of a word u if v consists of k first letters of some tail u.
(4′) A k*-beginning* is the same as a k-tail.
(5) A word u *is to the left from* a word v if u begins to the left from the beginning of v.

Notation 2.2.

(1) For a real number x put $\ulcorner x \urcorner := -[-x]$.
(2) Let $|u|$ denote the length of a word u.

The proof uses the following sufficient conditions for the presence of a period.

Lemma 2.3. *In a word W of length x, either the first $[x/d]$ tails are pairwise comparable or W includes a period of length d.*

Proof. Suppose that W includes no word of the form u^d. Consider the first $[x/d]$ tails. Suppose some two of them, say v_1 and v_2, are incomparable and $v_1 = u \cdot v_2$. Then $v_2 = u \cdot v_3$ for some v_3. Furthermore, $v_1 = u^2 \cdot v_3$. Arguing in this way, we obtain that $v_1 = u^d \cdot v_{d+1}$, since $|u| < x/d$, $|v_2| \geq (d-1)x/d$. A contradiction. □

Lemma 2.4. *If a word V of length $k \cdot t$ includes at most k different subwords of length k, then V includes a period of length t.*

Proof. We use induction in k. The base $k = 1$ is obvious. If there are at most $k-1$ different subwords of length $k-1$, then we apply the induction assumption. If there exist k different subwords of length $k-1$, then every subword of length k is uniquely determined by its first $k-1$ letters. Thus, $V = v^t$, where v is a k-tail of V. □

Definition 2.5.

(1) A word W is n*-divisible in the ordinary sense* if there exist $u_1, u_2, \ldots, u_n$ such that $W = v \cdot u_1 \cdots u_n$ and $u_1 \succ \cdots \succ u_n$.

(2) In our proof, we will call a word W *n-divisible in the tail sense* if there exist tails $u_1, \dots, u_n$ such that $u_1 \succ u_2 \succ \cdots \succ u_n$ and for any $i = 1, 2, \dots, n-1$ the beginning of u_i is to the left from the beginning of u_{i+1}. If the contrary is not specified, an *n-divisible* word means a word that is n-divisible in the tail sense.

(3) A word W is *n-cancellable* if either it is n-divisible in the ordinary sense or there exists a word of the form $u^d \subseteq W$.

Now we describe a sufficient condition for n-cancellability and its connection with n-divisibility.

Lemma 2.6. *If a word W includes n identical disjoint subwords u of length $n \cdot d$, then W is n-cancellable.*

Proof. Suppose the contrary. Consider the tails $u_1, u_2, \dots, u_n$ of the word u that begin from each of the first n letters of u. Renumerate the tails to provide the inequalities $u_1 \succ \cdots \succ u_n$. By Lemma 2.3, the tails are incomparable. Consider the subword u_1 in the leftmost copy of u, the subword u_2 in the second copy from the left,..., u_n in the nth copy from the left. We get an n-division of W. A contradiction. $\square$

Lemma 2.7. *If a word W is $4nd$-divisible, then it is n-cancellable.*

Proof. Suppose the contrary. Consider the numbers of positions of letters a_i, $a_1 < a_2 < \cdots < a_{4nd}$, that begin the tails u_i dividing W. Set $a_{4nd+1} = |W|$. If W is not n-cancellable, then there exists i, $1 \le i \le 4(n-1)d + 1$, such that for any $i \le b < c \le d < e \le i + 4d$ the $(a_c - a_b)$-tail u_b is incomparable with the $(a_e - a_d)$-tail u_d. Compare $a_{i+2d} - a_i$ and $a_{i+4d} - a_{i+2d}$. We may assume that $a_{i+4d} - a_{i+2d} \ge a_{i+2d} - a_i$. Let $a_{j+1} - a_j = \inf_k (a_{k+1} - a_k)$, $0 \le j < 2d$. We may assume that $j < d$. By assumption, the $(a_{2d} - a_j)$-tail u_j and the $(a_{2d} - a_{j+1})$-tail u_{j+1} are incomparable with the $(a_{4d} - a_{2d})$-tail u_{2d}. Since $a_{4d} - a_{2d} \ge a_{2d} - a_j > a_{2d} - a_{j+1}$, the $(a_2d - a_j)$-tail u_j and the $(a_{2d} - a_{j+1})$-tail u_{j+1} are mutually incomparable. Since $(a_{2d} - a_j)/(a_{2d} - a_{j+1}) \le (d+1)/d$, the $(a_{j+1} - a_j)$-tail u_j in degree d is included in the $(a_2d - a_j)$-tail u_j. A contradiction. $\square$

Corollary 2.8. *If a word W is not n-divisible in the ordinary sense, then W is not $4nd$-divisible* (*in the tail sense*).

Notation 2.9. Set $p_{n,d} := 4nd - 1$.

Let W be a non-n-cancellable word. Consider U, the $[|W|/d]$-tail of W. Then W is not $(p_{n,d} + 1)$-divisible. Let Ω be the set of tails of W that begin in U. Then by Lemma 2.3 any two elements of Ω are comparable. There is a natural bijection between Ω, the letters of U, and positive integers from 1 to $|\Omega| = |U|$.

Let us introduce a word θ that is lexicographically less than any other word.

Remark 2.10. In the proof of Theorem 1.15, all tails are assumed to belong to Ω.

3. Estimates on Occurrence of Periodical Fragments

Consider an application of the Dilworth theorem. For tails u and v put $u < v$ if $u \prec v$ and u is to the left from v. Then by the Dilworth theorem, Ω can be divided into $p_{n,d}$ chains such that in each chain $u \prec v$ if u is to the left from v. Paint the initial positions of the tails in $p_{n,d}$ colors according to their occurrence in chains. Fix a positive integer p. To each positive integer i from 1 to $|\Omega|$, attach $B^p(i)$, an ordered set of $p_{n,d}$ words $\{f(i,j)\}$ constructed as follows: for each $j = 1, 2, \dots, p_{n,d}$ put

$$f(i,j) = \{\max f \le i \colon f \text{ is painted in color } j\}.$$

If there is no such f, then the word from $B^p(i)$ at position j is assumed to be equal to θ, otherwise to the p-tail that begins from the $f(i,j)$th letter.

Informally speaking, we observe the speed of “evolution” of tails in their chains when the sequence of positions in W is considered as the time axis.

3.1. The Sets $B^p(i)$ and the Process at Positions.

Lemma 3.1 (on the process)**.** *Given a sequence S of length $|S|$ consisting of words of length $k-1$. Each word consists of $k-2$ symbols 0 and a single symbol 1. Let S satisfy the following condition: if for some $0 < s \le k-1$ there exist $p_{n,d}$ words such that 1 occupies the sth position, then between the first and the $p_{n,d}$th of these words there exists a word in which 1 occupies a position with number strictly less than s; $L(k-1) = \sup\limits_S |S|$. Then $L(k-1) \le p_{n,d}^{k-1} - 1$.*

Proof. We have that $L(1) \le p_{n,d}-1$. Let $L(k-1) \le p_{n,d}^{k-1}-1$. We will show that $L(k) \le p_{n,d}^k-1$. Consider the words such that 1 occupies the first position. Their number does not exceed $p_{n,d}-1$. Between any two of them, as well as before the first one and after the last one, the number of words does not exceed $L(k-1) \le p_{n,d}^{k-1} - 1$. Hence

$$L(k) \le p_{n,d} - 1 + p_{n,d}\big((p_{n,d})^{k-1} - 1\big) = p_{n,d}^k - 1. \qquad \square$$

We need a quantity that estimates the speed of "evolution" of sets $B^p(i)$.

Definition 3.2. Set

$$\psi(p) := \{\max k \colon B^p(i) = B^p(i+k-1)\}.$$

In particular, by Lemma 2.4 we have that $\psi(p_{n,d}) \le p_{n,d}d$.

For a given α we divide the sequence of the first $|\Omega|$ positions i of W into equivalence classes $\sim_\alpha$ as follows: $i \sim_\alpha j$ if $B^\alpha(i) = B^\alpha(j)$.

Proposition 3.3. *For any positive integers $a < b$ we have that $\psi(a) \le \psi(b)$.*

Lemma 3.4 (basic)**.** *For any positive integers a and k we have that*

$$\psi(a) \le p_{n,d}^k \psi(k\cdot a) + k\cdot a.$$

Proof. Consider the least representative in each class of $\sim_{k\cdot a}$. We get a sequence of positions $\{i_j\}$. Now consider all i_j and $B^{k\cdot a}(i_j)$ from the same equivalence class of $\sim_a$. Suppose it consists of $B^{k\cdot a}(i_j)$ for $i_j \in [b,c)$. Let $\{i_j\}'$ denote the segment of the sequence $\{i_j\}$ such that $i_j \in [b, c-k\cdot a)$.

Fix a positive integer r, $1 \le r \le p_{n,d}$. All $k\cdot a$-beginnings of color r that begin from positions of the word W in $\{i_j\}'$ will be called representatives of type r. All representatives of type r are pairwise distinct because they begin from the least positions in equivalence classes of $\sim_{k\cdot a}$. Divide each representative of type r into k segments of length a. Enumerate segments inside each representative of type r from left to right by integers from zero to $k-1$. If there exist $p_{n,d}+1$ representatives of type r with the same first $t-1$ segments but with pairwise different tth segments where $1 \le t \le k-1$, then there are two tth segments such that their first letters are of the same color. Then the initial positions of these segments belong to different equivalence classes of $\sim_a$.

Now apply Lemma 3.1 as follows: in all representatives of type r except the rightmost one we consider a segment as a *unit segment* if it contains the least position where this representative of type r differs from the preceding one. All other segments are considered as *zero segments*.

Now we apply the process lemma for the values of parameters as given in the condition of the lemma. We obtain that the sequence $\{i_j\}'$ contains at most $p_{n,d}^{k-1}$ representatives of type r. Then the sequence $\{i_j\}'$ contains at most $p_{n,d}^k$ terms. Thus, $c - b \le p_{n,d}^k\psi(k\cdot a) + k\cdot a$. $\square$

3.2. Completion of the Proofs for Theorems 1.15 and 1.17.

Let

$$a_0 = 3^{\ulcorner \log_3 p_{n,d} \urcorner}, \quad a_1 = 3^{\ulcorner \log_3 p_{n,d} \urcorner - 1}, \quad \dots, \quad a_{\ulcorner \log_3 p_{n,d} \urcorner} = 1.$$

Then $|W| \le d|\Omega| + d$ by Lemma 2.3.

Since for the set $B^1(i)$ at most $1 + p_{n,d}l$ different values are possible, we have that

$$|W| \le d(1 + p_{n,d}l)\psi(1) + d.$$

By Lemma 3.4,

$$\psi(1) < (p_{n,d}^3 + p_{n,d})\psi(3) < (p_{n,d}^3 + p_{n,d})^2\psi(9) < \cdots < (p_{n,d}^3 + p_{n,d})^{\ulcorner \log_3 p_{n,d} \urcorner}\psi(p_{n,d})$$
$$\leq (p_{n,d}^3 + p_{n,d})^{\ulcorner \log_3 p_{n,d} \urcorner} p_{n,d} d.$$

Take $p_{n,d} = 4nd - 1$ to get

$$|W| < 4^{5+3\log_3 4} l (nd)^{3\log_3(nd)+(5+6\log_3 4)} d^2.$$

This implies the assertion of Theorem 1.15.

The proof of Theorem 1.17 is completed similarly but instead of the sequence

$$a_0 = 3^{\ulcorner \log_3 p_{n,d} \urcorner}, \quad a_1 = 3^{\ulcorner \log_3 p_{n,d} \urcorner - 1}, \quad \ldots, \quad a_{\ulcorner \log_3 p_{n,d} \urcorner} = 1$$

we must consider the sequence

$$a_0 = 2^{\ulcorner \log_2 p_{n,d} \urcorner}, \quad a_1 = 2^{\ulcorner \log_2 p_{n,d} \urcorner - 1}, \quad \ldots, \quad a_{\ulcorner \log_2 p_{n,d} \urcorner} = 1.$$

4. An Estimate for the Essential Height

In this section, we proceed with the proof of the main theorem 1.12. In passing, we prove Theorem 1.19. We consider positions of letters in the word W as the time axis, that is, a subword u occurs before a subword v if u is entirely to the left from v in W.

4.1. Isolation of Distinct Periodical Fragments in the Word W. Let s denote the number of subwords in W such that each of them includes a period of length less than n more than $2n$ times and each pair of them is separated by subwords of length greater than n, comparable with the preceding period. Enumerate these from the beginning to the end of the word: $x_1^{2n}, x_2^{2n}, \ldots, x_s^{2n}$. Thus, $W = y_0 x_1^{2n} y_1 x_2^{2n} \cdots x_s^{2n} y_s$.

If there is i such that the word x_i has length not less than n, then the word x_i^2 includes n pairwise comparable tails, hence the word x_i^{2n} is n-divisible. Then s is not less than the essential height of W over the set of words of length less than n.

Definition 4.1. A word u will be called *noncyclic* if u is not representable in the form v^k, where $k > 1$.

Definition 4.2. A *word cycle* u is the set consisting of the word u and all its cyclic shifts.

Definition 4.3. A *cycle word* u is the cycle of the letters of the word u, where we mean that the first letter of u is after the last letter of u.

Definition 4.4. If any cyclic shifts of words u and v are comparable, then these words are called *strongly comparable* or *s-comparable*. *Strong comparability* of word cycles and cycle words is defined similarly.

Later we will use a bijection between cycle words and word cycles.

Definition 4.5. A word is *strongly n-divisible* or *n-s-divisible over a set of words Z* if it can be adduced as $W = W_0 W_1 \cdots W_n$, where the subwords $W_1, \ldots, W_n$ are in decreasing lexicographical order and every word W_i, $i = 1, 2, \ldots, n$, begins from a word z_i^k. The words z_i are distinct and belong Z.

Lemma 4.6. *If there is an integer m, $1 \leq m < n$, such that there exist $2n - 1$ pairwise incomparable words of length m: $x_{i_1}, \ldots, x_{i_{2n-1}}$, then W is n-divisible.*

Proof. Put $x := x_{i_1}$. Then W includes disjoint subwords $x^{p_1} v_1', \ldots, x^{p_{2n-1}} v_{2n-1}'$, where $p_1, \ldots, p_{2n-1}$ are positive integers greater than n, and $v_1', \ldots, v_{2n-1}'$ are words of length m comparable with x, $v_1' = v_{i_1}$. Hence among the words $v_1', \ldots, v_{2n-1}'$ either there are n words lexicographically greater than x or there are n words lexicographically smaller than x. We may assume that $v_1', \ldots, v_n'$ are lexicographically greater than x. Then W includes subwords $v_1', x v_2', \ldots, x^{n-1} v_n'$, which lexicographically decrease from left to right. $\square$

Consider an integer m, $1 \le m < n$. Divide all x_i of length m into equivalence classes relative to strong incomparability and choose a single representative from each class. Let these be $x_{i_1}, \ldots, x_{i'_s}$, where s' is a positive integer. Since the subwords x_i are periods, we consider them as word cycles.

Notation 4.7. $v_k := x_{i_k}$

Let $v(k,i)$, where i is a positive integer, $1 \le i \le m$, be a cyclic shift of a word v_k by $k-1$ positions to the right, that is, $v(k,1) = v_k$ and the first letter of $v(k,2)$ is the second letter of v_k. Thus, $\{v(k,i)\}_{i=1}^{m}$ is a word cycle of v_k. Note that for any $1 \le i_1$, $i_2 \le p$, $1 \le j_1$, $j_2 \le m$ the word $v(i_1, j_1)$ is strongly incomparable with $v(i_2, j_2)$.

Remark 4.8. The cases $m = 2, 3, n-1$ were considered in [22, 23].

4.2. An Application of Dilworth Theorem. Consider a set $\Omega' = \{v(i,j)\}$, where $1 \le i \le p$, $1 \le j \le m$. Order the words $v(i,j)$ as follows: $v(i_1, j_1) \succ v(i_2, j_2)$ if

(1) $v(i_1, j_1) > v(i_2, j_2)$;
(2) $i_1 > i_2$.

Lemma 4.9. *If in the set Ω' with ordering $\succ$ there exists an antichain of length n, then W is n-divisible.*

Proof. Suppose that there exists an antichain of length n consisting of words $v(i_1, j_1), v(i_2, j_2), \ldots, v(i_n, j_n)$; here $i_1 \le i_2 \le \cdots \le i_n$. If all inequalities between i_k are strict, then W is n-divisible by definition.

Suppose that for some r there exist $i_{r+1} = \cdots = i_{r+k}$ such that either $r = 0$ or $i_r < i_{r+1}$. Moreover, the positive integer k is such that either $k = n - r$ or $i_{r+k} < i_{r+k+1}$.

The word $s_{i_{r+1}}$ is periodical, hence it is representable as a product of n copies of $v^2_{i_{r+1}}$. The word $v^2_{i_{r+1}}$ includes a word cycle $v_{i_{r+1}}$. Hence in $s_{i_{r+1}}$ there exist disjoint subwords placed in lexicographically decreasing order and equal to $v(i_{r+1}, j_{r+1}), \ldots, v(i_{r+k}, j_{r+k})$, respectively. Similarly we deal with all sets of equal indices in the sequence $\{i_r\}_{r=1}^{n}$. The result is the n-divisibility of W. A contradiction. $\square$

Thus, Ω' can be divided into $n-1$ chains.

Notation 4.10. Put $q_n = (n-1)$.

4.3. The Sets $C^{\alpha}(i)$, the Process at Positions. Paint the first letters of the words from Ω' in q_n colors according to their occurrence in chains. Paint also the integers from 1 to $|\Omega'|$ in the corresponding colors. Fix a positive integer $\alpha \le m$. To each integer i from 1 to $|\Omega'|$, attach an ordered set $C^{\alpha}(i)$ of q_n words in the following way. For each $j = 1, 2, \ldots, q_n$ put

$$\begin{aligned} f(i,j) = \{\max f \le i \colon & \text{there exists } k \text{ such that } v(f,k) \text{ is painted in color } j \\ & \text{and the } \alpha\text{-tail beginning from } f \text{ consists only of letters initial in some tails from } \Omega'\}. \end{aligned}$$

If there is no such f, then a word from $C^{\alpha}(i)$ is assumed to be equal to θ, otherwise we assume it to be equal to the α-tail of $v(f,k)$.

Notation 4.11. Set

$$\phi(a) = \{\max k \colon \text{for some } i \text{ we have } C^a(i) = C^a(i+k-1)\}.$$

For a given $a \le m$ define a division of the sequence of word cycles $\{i\}$ in W into equivalence classes as follows: $i \sim_a j$ if $C^a(i) = C^a(j)$.

Note that the above construction is rather similar to the construction from the proof of Theorem 1.15. Observe that $B^a(i)$ and $C^a(i)$ are rather similar as well as $\psi(a)$ and $\phi(a)$.

Lemma 4.12. *$\phi(m) \le q_n/m$.*

Proof. In Notation 4.7, we have enumerated word cycles. Consider the word cycles with numbers $i, i+1, \dots, i+[q_n/m]$. We have shown that each word cycle consists of m distinct words. Now consider words in the word cycles $i, i+1, \dots, i+[q_n/m]$ as elements of the set Ω'. Then the first letter in each word cycle gets some position. The total number of the positions in question is not less than n. Hence at least two of these positions are of the same color. Now strong incomparability of word cycles implies the assertion of the lemma. $\square$

Proposition 4.13. *For any positive integers $a < b$ we have $\phi(a) \le \phi(b)$.*

Lemma 4.14 (basic)**.** *For positive integers a and k such that $ak \le m$, we have*

$$\phi(a) \le p_{n,d}^k \phi(k \cdot a).$$

Proof. Consider the minimal representative in each class of $\sim_{k\cdot a}$. We get a sequence of positions $\{i_j\}$. Now consider all i_j and $C^{k\cdot a}(i_j)$ from the same equivalence class of $\sim_a$. Suppose it consists of $C^{k\cdot a}(i_j)$ for $i_j \in [b, c)$. Let $\{i_j\}'$ denote the segment of the sequence $\{i_j\}$ such that $i_j \in [b, c)$.

Fix a positive integer r, $1 \le r \le q_n$. All $k \cdot a$-beginnings of color r that begin from positions of W in $\{i_j\}'$ will be called representatives of type r. All representatives of type r are distinct because they begin at the least positions in equivalence classes of $\sim_{k\cdot a}$. Divide each representative of type r into k segments of length a. Enumerate the segments of each representative of type r from left to right by integers from 0 to $k-1$. If there exist $q_n + 1$ representatives of type r with the same first $t-1$ segments but pairwise different tth segments where $1 \le t \le k-1$, then there are two tth segments such that their first letters are of the same color. Then the initial positions of these segments belong to different equivalence classes of $\sim_a$.

Now apply Lemma 3.1 in the following way: in all representatives of type r except the rightmost one, we consider a segment as a *unit segment* if it contains the least position where this representative of type r differs from the preceding one. All other segments are considered as *zero segments*.

Now we can apply the process lemma for the values of parameters as given in the condition of the lemma. We obtain that the sequence $\{i_j\}'$ contains not more than q_n^{k-1} representatives of type r. Then the sequence $\{i_j\}'$ contains at most q_n^k terms. Thus, $c - b \le q_n^k \phi(k \cdot a)$. $\square$

4.4. Completion of the Proof of Theorem 1.19. Suppose that

$$a_0 = 3^{\ulcorner \log_3 p_{n,d} \urcorner}, \quad a_1 = 3^{\ulcorner \log_3 p_{n,d} \urcorner - 1}, \quad \dots, \quad a_{\ulcorner \log_3 p_{n,d} \urcorner} = 1.$$

Substitute these a_i into Lemmas 4.14 and 4.12 to obtain

$$\phi(1) \le q_n^3 \phi(3) \le q_n^9 \phi(9) \le \dots \le q_n^{3^{\ulcorner \log_3 m \urcorner}} \phi(m) \le q_n^{3^{\ulcorner \log_3 m \urcorner + 1}}.$$

Since C_i^1 takes at most $1 + q_n l$ distinct values, we have

$$|\Omega'| < q_n^{3^{\ulcorner \log_3 m \urcorner + 1}} (1 + q_n l) < n^{3^{\ulcorner \log_3 n \urcorner + 2}} l.$$

By virtue of Lemma 4.6, the number of subwords x_i of length m is less than $2n^{3^{\ulcorner \log_3 n \urcorner + 3}} l$. Thus, the total number of subwords x_i is less than $2n^{3^{\ulcorner \log_3 n \urcorner + 4}} l$. So $s < 2n^{3^{\ulcorner \log_3 n \urcorner + 4}} l$ and Theorem 1.19 is proved.

5. Proof of the Main Theorem 1.12 and of Theorem 1.14

5.1. Outline of the Proof. Now an *n-divisible word* will mean a word that is n-divisible in the ordinary sense. To start with, we find the necessary number of fragments in W with length of the period not less than $2n$. For this, it suffices to divide W into subwords of large length and to apply Theorem 1.15 to them. However, the estimate can be improved. For this, we find a periodical fragment u_1 in W with period length not less than $4n$. Removing u_1, we obtain a word W_1. In W_1, we find a fragment u_2 with period length not less than $4n$ and remove it to get a word W_2. Now we again remove a periodical fragment and proceed in this way, as is described in Algorithm 5.2 in more detail. Then we restore the original word W, using the removed fragments. Further we show that a subword u_i in W usually is not a product of a big number of not neighboring subwords. In Lemma 5.4, we prove that application of Algorithm 5.2 enables one to find the necessary number of removed subwords of W with period length not less than $2n$.

5.2. Summing up of Essential Heights and Nilpotency Degrees.

Notation 5.1. Let $\mathrm{Ht}(w)$ denote the height of a word w over the set of words of degree not exceeding n.

Consider a word W of height $\mathrm{Ht}(W) > \Phi(n,l)$. Apply the following algorithm to it.

Algorithm 5.2.

Step 1. By Theorem 1.15, the word W includes a subword with period length $4n$. Suppose that $W_0 = W = u_1' x_{1'}^{4n} y_1'$, where the word $x_{1'}$ is not cyclic. Represent y_1' in the form $y_1' = x_{1'}^{r_2} y_1$, where r_2 is maximal possible. Represent u_1' as $u_1' = u_1 x_{1'}^{r_1}$, where r_1 is maximal possible. Denote by f_1 the word

$$W_0 = u_1 x_{1'}^{4n+r_1+r_2} y_1 = u_1 f_1 y_1.$$

In the sequel, the positions contained in f_1 are called *tedious*, the last position of u_1 is called *tedious of type* 1, the second position from the end in u_1 is called *tedious of type* 2, etc., the nth position from the end in u_1 is called *tedious of type* n. Put $W_1 = u_1 y_1$.

Step k. Consider the words u_{k-1}, y_{k-1}, $W_{k-1} = u_{k-1} y_{k-1}$ constructed at the preceding step. If $|W_{k-1}| \geq \Phi(n,l)$, then we apply Theorem 1.15 to W with the restriction that the process in the main lemma 3.4 is applied only to nontedious positions and to tedious positions of type greater than ka, where k and a are the parameters from Lemma 3.4. Thus, W_{k-1} includes a noncyclic subword with period length $4n$ such that

$$W_{k-1} = u_k' x_{k'}^{4n} y_k'.$$

Then put

$$r_1 := \sup\{r \colon u_k' = u_k x_{k'}^r\}, \quad r_2 := \sup\{r \colon y_k' = x_{k'}^r y_k\}.$$

(Note that the words involved may be empty.) Define f_k by the equation

$$W_{k-1} = u_k x_{k'}^{4n+r_1+r_2} y_k = u_k f_k y_k.$$

In the sequel, the positions contained in f_k are called *tedious*, the last position in u_k is *tedious of type* 1, the second position from the end in u_k is *tedious of type* 2, etc., the nth position from the end in u_k is *tedious of type* n. If a position happens to be tedious of two types, then the lesser type is chosen for it. Put $W_k = u_k y_k$.

Notation 5.3. Perform $4t+1$ steps of Algorithm 5.2 and consider the original word W. For each integer i from the segment $[1, 4t]$ we have

$$W = w_0 f_i^{(1)} w_1 f_i^{(2)} \cdots f_i^{(n_i)} w_{n_i}$$

for some subwords w_j. Here $f_i = f_i^{(1)} \cdots f_i^{(n_i)}$. Moreover, we assume that for $1 \leq j \leq n_i - 1$ the subword w_j is not empty. Let $s(k)$ be the number of indices $i \in [1, 4t]$ such that $n_i = k$.

To prove Theorem 1.15, we must find as many long periodic fragments as possible. For this, we can use the following lemma.

Lemma 5.4. *$s = s(1) + s(2) \geq 2t$.*

Proof. A subword U of the word W will be called *monolithic* if

(1) U is a product of words of the form $f_i^{(j)}$;
(2) U is not a proper subword of a word that satisfies the above condition.

Suppose that after the $(i-1)$th step of Algorithm 5.2 the word W includes k_{i-1} monolithic subwords. Note that $k_i \leq k_{i-1} - n_i + 2$. Thus, if $n_i \geq 3$, then $k_i \leq k_{i-1} - 1$. If $n_i \leq 2$, then $k_i \leq k_{i-1} + 1$. Furthermore, $k_1 = 1$ and $k_t \geq 1 = k_1$. The lemma is proved. $\square$

Corollary 5.5.

$$\sum_{k=1}^{\infty} k \cdot s(k) \leq 10t \leq 5s.$$

Proof. From the proof of Lemma 5.4 we obtain that $\sum\limits_{n_i \geq 3} (n_i - 2) \leq 2t$. By definition, $\sum\limits_{k=1}^{\infty} s(k) = 4t$, i.e., $\sum\limits_{k=1}^{\infty} 2s(k) = 8t$. Summing up these two inequalities and applying Lemma 5.4, we obtain the required inequality. □

Proposition 5.6. *The height of W does not exceed*

$$\Psi(n, 4n, l) + \sum_{k=1}^{\infty} k \cdot s(k) \leq \Psi(n, 4n, l) + 5s.$$

In the sequel, we consider only f_i with $n_i \leq 2$.

Notation 5.7. If $n_i = 1$, then put $f_i' := f_i$.

If $n_i = 2$, then put $f_i' := f_i^{(j)}$, where $f_i^{(j)}$ is the word of maximal length between $f_i^{(1)}$ and $f_i^{(2)}$.

Order the words f_i' according to their distance from the beginning of W. We get a sequence $f'_{m_1}, \dots, f'_{m_s}$, where $s' = s(1) + s(2)$. Put $f_i'' := f'_{m_i}$. Suppose that $f_i'' = w_i' x_{i''}^{p_{i''}} w_i''$, where at least one of the words w_i', w_i'' is empty.

Remark 5.8. We may assume that at starting steps of Algorithm 5.2 we have chosen all f_i such that $n_i = 1$.

Now consider z_j', the subwords in W of the following form:

$$z_j' = x_{(2j-1)''}^{p_{(2j-1)''} + \gimel} v_j, \quad \gimel \geq 0, \quad |v_j| = |x_{(2j-1)''}|.$$

Here v_j is not equal to $x_{(2j-1)''}$ and the beginning of z_j' coincides with the beginning of a periodic subword in f''_{2j-1}. We will show that z_j' are disjoint.

Indeed, if $f''_{2j-1} = f_{m_{2j-1}}$, then $z_j' = f_{m_{2j-1}} v_j$. If $f''_{2j-1} = f^{(k)}_{m_{2j-1}}$, $k = 1, 2$, and z_j' intersects z'_{j+1}, then $f''_{2j} \subset z_i'$. Since $x_{(2j)''}$ and $x_{(2j-1)''}$ are not cyclic, we have $|x_{(2j)''}| = |x_{(2j-1)''}|$. But then the period length in z_j' is not less than $4n$, a contradiction with Remark 5.8.

Thus, we have proved the following lemma.

Lemma 5.9. *In a word W with height not greater than $\Psi(n, 4n, l) + 5s'$, there exist at least s' disjoint periodic subwords such that the period occurs in each of them at least $2n$ times. Furthermore, between any two elements of this set of periodic subwords there is a subword with the same period length as the leftmost of these two elements.*

5.3. Completion of the Proof of the Main Theorem 1.12 and for Theorem 1.14. Replace s' in Lemma 5.9 by s from the proof of Theorem 1.19 to obtain that the height of W does not exceed

$$\Psi(n, 4n, l) + 5s < E_1 l \cdot n^{E_2 + 12 \log_3 n},$$

where $E_1 = 4^{21 \log_3 4 + 17}$, $E_2 = 30 \log_3 4 + 10$.

Thus, we have obtained the assertion of the main theorem 1.12.

Proof of Theorem 1.14 is completed similarly but we must replace in part 4.4 the sequence

$$a_0 = 3^{\ulcorner \log_3 p_{n,d} \urcorner}, \quad a_1 = 3^{\ulcorner \log_3 p_{n,d} \urcorner - 1}, \quad \dots, \quad a_{\ulcorner \log_3 p_{n,d} \urcorner} = 1$$

by the sequence

$$a_0 = 2^{\ulcorner \log_2 p_{n,d} \urcorner}, \quad a_1 = 2^{\ulcorner \log_2 p_{n,d} \urcorner - 1}, \quad \dots, \quad a_{\ulcorner \log_2 p_{n,d} \urcorner} = 1,$$

and to take the value of $\Psi(n, 4n, l)$ from Theorem 1.17.

6. Partial Periodicity Estimates

The represented technique suggests that the given estimate on the height is a little rough and probably could be improved by a bit more sophisticated combinatorial arguments. Anyway it remains subexponential. We need to use new ideas and methods to get or refute a polynomial estimate.

At the beginning of the main theorem's proof we deal with subwords as we deal with independent objects. After that we assume that positions in subwords are colored. If we only use the colors of the first positions of the subwords, then we obtain only an exponential estimate. But on the other hand, if we keep in mind the colors of all positions in the word, then we obtain only an exponential estimate again. This fact is a consequence of the hierarchic structure of the subwords' system. Perhaps detailed studying of relations between subwords lets us upgrade our estimate to polynomial.

It is also of interest to obtain estimates for height of an algebra over a set of words whose degrees do not exceed the complexity of the algebra (PI-degree in literature in English). The paper [6] presents exponential estimates, and for words that are not a linear combination of lexicographically smaller words, overexponential estimates were obtained in [5].

Next we obtain estimates of the number of subwords with period of length 2, 3, $n-1$ of an arbitrary non-n-divisible word W. The case of words of length 2 and 3 is generalized to the proof of the essential height's limitation. Moreover, we get the lower estimate of the number of 2-periodical words. This estimate is less than the upper estimate in four times for $l \to \infty$.

As previously, hereinafter words are constructed over an alphabet $\mathbb{A} = \{a_1, a_2, \dots, a_l\}$.

6.1. The Proof of Theorem 1.23. Suppose that a word W is not strongly n-divisible. Let an arbitrary positive integer m be bigger than $2n$. Let the set Ω'' be the set of nonoverlapping cyclic comparable subwords of the form z^m, where z is an acyclic two-letter word and we deal with subwords of the word W. Elements of the set Ω'' are called *exemplars*. We mean that the elements of Ω'' are the exemplars of different equivalence classes of strong n-divisibility. Let t be the cardinality of the set Ω''. Let us number all exemplars from left to right so the leftmost exemplar has number 1 and the rightmost one has number t. There are exactly two different two-letter subwords in any exemplar. Order these words as follows: $u \prec v$ if

(1) u is lexicographically less than v;
(2) the exemplar that comprises u is to the left from the one that comprises v.

The number of pairwise incomparable cyclic subwords of W is less than $n-1$ because W is not strongly n-divisible. By the Dilworth theorem, there exists a partition of the considered two-letter words into $n-1$ chains. Let us paint the words in the color of the chains these words belong to.

We have a bijection between the following four objects:

(1) positive integers from 1 to t;
(2) equivalence classes of strong comparability;
(3) two-letter word cycles from the equivalence classes of strong comparability;
(4) pairs of colors such that the words of this word cycles are colored in them.

Let us color the first positions of the words of word cycles in the colors of the corresponding words.

Consider a graph Γ with vertices of the form (k, i), where $0 < k < n$ and $0 < i \le l$. The first coordinate complies with a color and the second one complies with a letter. Two vertices (k_1, i_1), (k_2, i_2) are connected by an *edge of weight* j if

(1) there are letters i_1 and i_2 in the jth exemplar;
(2) these letters are colored in the colors k_1 and k_2, respectively.

Let us deal with the set E_{k_1,k_2} of edges between the vertices of the forms (k_1, i_1) and (k_2, i_2), where the numbers k_1 and k_2 are fixed and the numbers i_1 and i_2 are arbitrary. Consider two edges l_1 and l_2 of weight $j_1 < j_2$, respectively, which connect vertices $A = (k_1, i_{1_1})$, $B = (k_2, i_{2_1})$ and $C = (k_1, i_{1_2})$, $D = (k_2, i_{2_2})$, respectively. Then it is true that $i_{1_1} \le i_{1_2}$ and $i_{2_1} \le i_{2_2}$. Since we deal with exemplars of different equivalence classes of the strong n-divisibility, one of these inequalities is strict. That is why

$i_{1_1} + i_{2_1} < i_{1_2} + i_{2_2}$. Since the second coordinates of vertices are less than $l+1$, the cardinality of the set E_{k_1,k_2} is less than $2l$.

Since the first coordinate of the vertices is less than n, the number of edges in the graph is no more than $(2l-1)(n-1)(n-2)/2$. Thus, we have obtained Theorem 1.23.

6.2. The Proof of Theorem 1.25. Suppose that a word W is not strongly n-divisible. Let an arbitrary positive integer m be bigger than $2n$. Let the set Ω'' be a set of nonoverlapping cyclic comparable subwords of the form z^m, where z is an acyclic three-letter word. Elements of the set Ω'' are called *exemplars*. We mean that elements of Ω'' are exemplars of different equivalence classes of strong n-divisibility. Let t be the cardinality of the set Ω''. Let us number all exemplars from left to right so that the leftmost exemplar has number 1 and the rightmost one has number t. There are exactly three different three-letter subwords in any exemplar.

Order these words as follows: $u \prec v$ if

(1) u is lexicographically less than v;

(2) the exemplar that comprises u is to the left from the one that comprises v.

The number of pairwise incomparable cyclic subwords of W is less than $n-1$ because the word W is not strongly n-divisible. By the Dilworth theorem, there exists a partition into $n-1$ chains of the considered three-letter words. Let us paint the words in the color of the chains these words belong to.

One can see that the proof of Theorem 1.25 almost coincides with the proof of Theorem 1.23. But it is necessary to introduce an oriented analog of the graph Γ.

Consider an oriented graph G with vertices of the form (k,i), where $0 < k < n$ and $0 < i \leq l$. The first coordinate complies with a color and the second one complies with a letter. The *edge of weight j* goes from vertex (k_1, i_1) to the vertex (k_2, i_2) if for some i_3 and k_3

(1) there is a cycle word $i_1 i_2 i_3$ in the jth exemplar;

(2) the letters i_1, i_2, and i_3 are colored in the colors k_1, k_2, and k_3, respectively.

So the graph G consists of oriented triangles with edges of the same weight. In contradistinction to the graph Γ, there may be multiple edges in the graph G. For the further exposition we need the following lemma.

Lemma 6.1 (basic)**.** *Let j be some positive integer. Suppose that A, B, and C are vertices of the graph G, and $A \to B \to C \to A$ is an oriented triangle with edges of the same weight j. Moreover, suppose that there exist other edges $A \to B$, $B \to C$, and $C \to A$ of weight a, b, and c, respectively. Then one of the numbers a, b, c is bigger than j.*

Proof. Suppose the contrary. Suppose that $a = b = c$. Then there exist two triangles $A \to B \to C \to A$ such that the edges are of the same weight in any one of them. Then there are two non-s-comparable words in Ω''. A contradiction. If two numbers from the set $\{a, b, c\}$ are equal to each other, then $a = b = c$. That is why a, b, and c are different numbers. Let a be the biggest one of them. Consider the triangle with edges of weight a. This triangle contains vertices A and B and some third vertex C'. If the second coordinates of the vertices C and C' are equal, then $\triangle ABC$ and $\triangle ABC'$ correspond to non-s-comparable words from the set Ω''. A contradiction. $a < j$ by assumption. Words that are colored in k_A (the first coordinate of vertex A) are monotonous. So the word $i_A i_B i_{C'}$, which is built from the second coordinates of the vertices A, B, C', is lexicographically less than $i_A i_B i_C$. That is why $i_{C'} < i_C$. Then the word $i_B i_{C'}$ is lexicographically less than the word $i_B i_C$. As words that are colored in k_B are monotonous, then $b > a$. A contradiction. $\square$

Let us finish the proof of Theorem 1.25. Consider a graph G_1, which is built from the graph G by replacing any multiple edge by the edge of least weight. By Lemma 6.1, there exist edges of all weights from 1 to t in the graph G_1.

Let us deal with the set E_{k_1,k_2} of edges between the vertices of the forms (k_1, i_1) and (k_2, i_2), where the numbers k_1 and k_2 are fixed and the numbers i_1 and i_2 are arbitrary. Consider two edges from the set

E_{k_1,k_2} of weight $j_1 < j_2$ with ends in some vertices (k_1, i_{1_1}), (k_2, i_{2_1}) and (k_1, i_{1_2}), (k_2, i_{2_2}), respectively. Then $i_{1_1} \leq i_{1_2}$ and $i_{2_1} \leq i_{2_2}$. Since we deal with the exemplars of different equivalence classes of the strong n-divisibility, one of these inequalities is strict. Since the second coordinates are less than $l+1$, the cardinality of the set E_{k_1,k_2} is less than $2l$.

As the first coordinate of the vertices is less than n, we see that the number of edges in the graph is no more than $(2l-1)(n-1)(n-2)$. Thus, we have obtained Theorem 1.25.

6.3. The Proof of Theorem 1.26. Suppose that a word W is not strongly n-divisible. As before, let an arbitrary positive integer m be bigger than $2n$. Let a set Ω'' be a set of nonoverlapping cyclic comparable subwords of the form z^m, where z is an acyclic $(n-1)$-letter word. Elements of the set Ω'' are called *exemplars*. We mean that elements of Ω'' are exemplars of different equivalence classes of strong n-divisibility. Let t be the cardinality of the set Ω''. Let us number all exemplars from left to right so the leftmost exemplar has number 1 and the rightmost one has number t. There are exactly $n-1$ different $(n-1)$-letter subwords in any exemplar.

Order these words as follows: $u \prec v$ if

(1) u is lexicographically less than v;
(2) the exemplar that comprises u is to the left from the one that comprises v.

The number of pairwise incomparable cyclic subwords of W is less than $n-1$ because the word W is not strongly n-divisible. By the Dilworth theorem, there exists a partition into $n-1$ chains of the considered $(n-1)$-letter words. Let us paint the words in the color of the chains these words belong to. Let us paint the first positions of words in the same colors as the respective words.

Consider an oriented graph G with vertices of the form (k,i), where $0 < k < n$ and $0 < i \leq l$. The first coordinate means a color and the second coordinate means a letter.

An edge of some weight j goes from (k_1, i_1) to (k_2, i_2) if

(1) there is a word cycle $i_1 i_2 \cdots i_{n-1}$ for some $i_3, i_4, \ldots, i_{n-1}$ in the jth exemplar;
(2) the colors of positions of letters i_1 and i_2 are k_1 and k_2, respectively.

So the graph G consists of oriented cycles of length $n-1$. The edges of any such cycle are of the same weight. Now we need an indicator that grows monotonically with the apparition of the new exemplars when we move from the beginning to the end of the word. Such an indicator is the number of pairs of connected vertices of graph G in the proof of Theorem 1.25. This indicator is the sum of the second coordinates of the nonisolated vertices of graph G in the current proof.

We need the following lemma.

Lemma 6.2 (basic)**.** *Let j be some positive integer. Suppose that $A_1, A_2, \ldots, A_{n-1}$ are vertices of the graph G and $A_1 \to A_2 \to \cdots \to A_{n-1} \to A_1$ is an oriented cycle of length $n-1$ with edges of weight j. Then there is no any other cycle of the same weight between vertices $A_1, A_2, \ldots, A_{n-1}$.*

Proof. Suppose the contrary. Consider the least positive integer j for which we can find another monochrome cycle between the vertices of the cycle of color j. Since j is the least, it can be considered that the color of this cycle is $k > j$. Let the cycle of color k have the form $A_{j_1}, A_{j_2}, \ldots, A_{j_{n-1}}$, where $\{j_p\}_{p=1}^{n-1} = \{1, 2, \ldots, n-1\}$. Let (k_j, i_j) be the coordinate of the vertex A_j. Consider the smallest number $q \in \mathbb{N}$ such that for some integer r the word $i_{j_r} i_{j_{r+1}} \cdots i_{j_{r+q-1}}$ is lexicographically bigger than the word $i_{j_r} i_{j_r+1} \cdots i_{j_r+q-1}$ (now and later we mean addition modulo $n-1$). Such q exists because the words $i_1 i_2 \cdots i_{n-1}$ and $i_{j_1} i_{j_2} \cdots i_{j_{n-1}}$ are strongly comparable. The sets $\{j_p\}_{p=1}^{n-1}$ and $\{1, 2, \ldots, n-1\}$ are equal so $q \geq 2$. Since q is the smallest, for any $s < q$ and for any r it is true that $i_{j_r} i_{j_{r+1}} \cdots i_{j_{r+s-1}} = i_{j_r} i_{j_r+1} \cdots i_{j_r+s-1}$. So for any $s < q$ and for any r the equality $i_{j_{r+s-1}} = i_{j_r+s-1}$ is true. The sequence of words of any color is monotonous, so for any r the word $i_{j_r} i_{j_{r+1}} \cdots i_{j_{r+q-1}}$ is no more than the word $i_{j_r} i_{j_r+1} \cdots i_{j_r+q-1}$. So the inequality $i_{j_{r+q-1}} \geq i_{j_r+q-1}$ is true for any r. By assumption, there exists r such that $i_{j_{r+q-1}} > i_{j_r+q-1}$. Since both sequences $\{j_{r+q-1}\}_{r=1}^{n-1}$ and $\{j_r + q - 1\}_{r=1}^{n-1}$ run through the set $\{1, 2, \ldots, n-1\}$ for one time, $\sum\limits_{r=1}^{n-1} j_{r+q-1} = \sum\limits_{r=1}^{n-1} (j_r + q - 1)$. But we got the equality $\sum\limits_{r=1}^{n-1} j_{r+q-1} > \sum\limits_{r=1}^{n-1} (j_r + q - 1)$. A contradiction. $\square$

Let us finish the proof of Theorem 1.26. Consider cycles of length $n-1$ of weight k and $j+1$ for some j. By the basic lemma 6.2, there exist numbers k and i such that the vertex (k, i) belongs to the cycle of weight $j+1$ but does not belong to the cycle of weight j. Let the cycle of weight j consist of vertices of the form $(k, i_{(j,k)})$, where $k = 1, 2, \ldots, n-1$. Let us introduce an indicator $\pi(j) = \sum\limits_{k=1}^{n-1} i_{(j,k)}$. Then by the main lemma 6.2 and by the monotony of sequences of the same colored words the inequality $\pi(j+1) \geq \pi(j)+1$ is true. Since we deal with noncyclic words, there exists an integer k such that $i_{(1,k)} > 1$. So $\pi(1) > n-1$. $\pi(j) \leq (l-1)(n-1)$ because for all j $i_{(j,k)} \leq l-1$. That is why $j \leq (l-2)(n-1)$. So $t \leq (l-2)(n-1)$. Thus, we obtain Theorem 1.26.

6.4. The Proof of Theorem 1.24. Let us present an example. We can assume that l is arbitrarily large. We assume that $l > 2^{n-1}$. We use constructions from the proof of Theorem 1.23. So the process of building an example is the same as building edges in a graph with l vertices. We divide this process into a few big steps. Let i be a positive integer from 1 to $l - 2^{n-1}$. Let the following pairs of vertices be connected in the following order during the ith big step:

$$\begin{gathered}(i,\, 2^{n-2}+i), \quad (i,\, 2^{n-2}+2^{n-3}+i), \quad (2^{n-2}+i,\, 2^{n-2}+2^{n-3}+i),\\ (i,\, 2^{n-2}+2^{n-3}+2^{n-4}+i), \quad (2^{n-2}+i,\, 2^{n-2}+2^{n-3}+2^{n-4}+i),\\ (2^{n-2}+2^{n-3}+i,\, 2^{n-2}+2^{n-3}+2^{n-4}+i), \quad \ldots,\\ (i,\, 2^{n-2}+\cdots+2+1+i), \quad \ldots, (2^{n-2}+\cdots+2+i,\, 2^{n-2}+\cdots+2+1+i).\end{gathered}$$

Herewith no edge is counted twice. That is why any vertex is connected only with vertices that differ from the selected vertex by a nonrecurring sum of powers of 2.

A vertex A is called a *vertex of the form* (k, i) if it is connected with k vertices of less value during the ith big step. For all i there exist vertices of the forms $(0, i), (1, i), \ldots, (n-2, i)$.

Consider words that begin from the letter corresponding to the vertex of the form (k, i). If such words end with the letter that connects with (k, i) during the ith big step, then we color them in i. We have correct coloring in $n-1$ color so the word is n-divisible.

We build $(n-2)(n-3)/2$ edges during the ith big step. So, $q = (l - 2^{n-1})(n-2)(n-3)/2$, where q is the number of edges in Γ. Thus, we obtained Theorem 1.24.

6.5. Estimating the Essential Height Using Theorem 1.24. We can obtain an exponential estimate of the essential height dealing with the case of words of degree 2. This estimate has polynomial growth when the degree of PI-identity is fixed and exponential growth when the number of generators is fixed. For this purpose we generalize some definitions that were introduced before. Hereinafter we use a bijection between word cycles and cycle words.

Construction 6.3. Consider an alphabet $\mathbb{A}$ from letters $\{a_1, a_2, \ldots, a_l\}$. Let us introduce a lexicographical order on them: $a_i > a_j$ if $i > j$. Consider any set of noncyclic pairwise strongly comparable word cycles of the same degree t. Let us enumerate the elements of this set by positive integers from 1. Order the words in word cycles as follows: $u \prec v$ if

(1) the word u is lexicographically less than the word v;
(2) the word cycle that contains the word u has smaller number than the word cycle that contains v.

Let us enumerate positions of letters in cycle words in numbers from 1 to t from the beginning to the end of some word from the respective word cycle.

Notation 6.4.

(1) Let $w(i,j)$ be a word degree t that begins from the jth position in the ith word cycle.
(2) Let a class $X(t,l)$ be the considered set of word cycles with order $\prec$.

Definition 6.5. We say that a class X is called *n-light* if it does not contain any antichain of length n. A class X is called *n-dark* if it contains an antichain of length n.

By the Dilworth theorem, the words in n-light classes can be colored in $n-1$ colors so that the sequence of monochrome colors is a chain. Then we need to estimate the number of elements in n-light classes X.

Definition 6.6. The maximal possible number of elements in n-light class $X(t,l)$ is denoted by $\beth(t,l,n)$.

Remark 6.7. Hereinafter the first argument of the function $\beth(\cdot,\cdot,\cdot)$ is less than the third one.

The following lemma lets us estimate $\beth(t,l,n)$ using the cases of small periods.

Lemma 6.8. $\beth(t,l^2,n) \geq \beth(2t,l,n)$.

Proof. Consider an n-light class $X(2t,l)$. Let us divide positions in all word cycles of $X(2t,l)$ into pairs of neighbors so that any position is exactly in one pair. Consider an alphabet $\mathbb{B} = \{b_{i,j}\}_{i,j=1}^{l}$, where $b_{i_1,j_1} > b_{i_2,j_2}$ if $i_1 \cdot l + j_1 > i_2 \cdot l + j_2$. The alphabet $\mathbb{B}$ consists of l^2 letters. Every pair from the partition consists of some letters a_i, a_j. Let us replace such a pair of letters a_i, a_j by the letter $b_{i,j}$. If we make this replacement for all pairs from the partition, then we get a new class $X(t,l^2)$. Suppose that there exists an antichain of length n from words $w(i_1,j_1), w(i_2,j_2), \ldots, w(i_n,j_n)$ in the class $X(t,l^2)$. Let us consider preimages of the words $w(i_1,j_1), w(i_2,j_2), \ldots, w(i_n,j_n)$ in the original class $X(2t,l)$. Let these preimages be $w(i_1,j_1'), w(i_2,j_2'), \ldots, w(i_n,j_n')$, respectively. Then the sequence $w(i_1,j_1'), w(i_2,j_2'), \ldots, w(i_n,j_n')$ is an antichain of length n in the class $X(2t,l)$. A contradiction. So the class $X(t,l^2)$ is n-light. Thus, we have proved the lemma. $\square$

Let us estimate $\beth(t,l,n)$ using the cases of the small periods.

Lemma 6.9. $\beth(t,l^2,n) \leq \beth(2t,l,2n-1)$.

Proof. Consider a $(2n-1)$-dark class $X(2t,l)$. We may assume that n words in the antichain begin from the odd positions of word cycles. Let these n words be $w(i_1,j_1), w(i_2,j_2), \ldots, w(i_n,j_n)$. Let us divide positions in the word cycles of $X(2t,l)$ into pairs of neighbors so that every position is exactly in one pair and the first position in every pair is odd. Then consider an alphabet $\mathbb{B} = \{b_{i,j}\}_{i,j=1}^{l}$. Order the letters of this alphabet as follows: $b_{i_1,j_1} > b_{i_2,j_2}$ if $i_1 \cdot l + j_1 > i_2 \cdot l + j_2$. $\mathbb{B}$ consists of l^2 letters. Every pair from the partition consists of some letters a_i, a_j. Let us replace such a pair of letters a_i, a_j by the letter $b_{i,j}$. If we make this replacement for all pairs from the partition, then we get a new class $X(t,l^2)$. Let the words $w(i_1,j_1), w(i_2,j_2), \ldots, w(i_n,j_n)$ map into words $w(i_1,j_1'), w(i_2,j_2'), \ldots, w(i_n,j_n')$. The sequence of these words is an antichain of length n in the class $X(t,l^2)$. So we get the n-dark class $X(t,l^2)$ with the same cardinality as the $(2n-1)$-dark class $X(2t,l)$. Thus, we have proved the lemma. $\square$

Now we need to connect $\beth(t,l,n)$ for any first argument and for the first argument that is equal to a power of 2.

Lemma 6.10. $\beth(t,l,n) \leq \beth(2^s, l+1, 2^s(n-1)+1)$, *where* $s = \ulcorner \log_2(t) \urcorner$.

Proof. Consider an n-light class $X(t,l)$. Let us add a new letter a_0 to the alphabet $\mathbb{A}$. Let a_0 be less than any other letter from $\mathbb{A}$. So we have the new alphabet $\mathbb{A}'$. Let us add the $(t+1)$th, $(t+2)$th, $\ldots$, 2^sth position to any word cycle from the class $X(t,l)$ and put the letter a_0 on these positions. So we get a new class $X(2^s,l+1)$. Suppose that it is not $(2^s(n-1)+1)$-light. Then for some j there exist a sequence of words $w(i_1,j), w(i_2,j), \ldots, w(i_n,j)$, which is an antichain in the class $X(2^s,l+1)$. So

(1) if $j > t$, then the sequence $w(i_1,1), w(i_2,1), \ldots, w(i_n,1)$ is an antichain in the class $X(t,l)$;

(2) if $j \leq t$, then the sequence $w(i_1, j), w(i_2, j), \ldots, w(i_n, j)$ is an antichain in the class $X(t, l)$.
So we obtained a contradiction with the assumption that the class $X(t, l)$ is n-light. Thus, the lemma is proved. □

Proposition 6.11. $\beth(t, l, n) \leq \beth(t, l, n+1)$.

By Lemma 6.10, we get that $\beth(t, l, n) \leq \beth(2^s, l+1, 2^s(n-1)+1)$, where $s = \ulcorner \log_2(t) \urcorner$. By Notation 6.7, we get that $t < n$. So $2^s < 2n$. Therefore,

$$\beth(2^s, l+1, 2^s(n-1)+1) \leq \beth(2^s, l+1, 2n^2).$$

By Lemma 6.8, the inequalities

$$\beth(2^s, l+1, 2n^2) \leq \beth(2^{s-1}, (l+1)^2, 2n^2) \leq \beth(2^{s-2}, (l+1)^{2^2}, 2n^2) \\ \leq \beth(2^{s-3}, (l+1)^{2^3}, 2n^2) \leq \cdots \leq \beth(2, (l+1)^{2^{s-1}}, 2n^2)$$

are true. By Theorem 1.23, we get that

$$\beth(2, (l+1)^{2^{s-1}}, 2n^2) < (l+1)^{2^{s-1}} \cdot 4n^4 < 4(l+1)^n n^4.$$

So we proved the following lemma.

Lemma 6.12. $\beth(t, l, n) < 4(l+1)^n n^4$.

We need to estimate the number of subwords with the same periods of non-n-divisible words to use Lemma 6.12 in the proof of Theorem 1.27.

Lemma 6.13. *If in any word W there exist $2n-1$ subwords such that a period repeats more than n times in every this subword, then their periods are pairwise non-s-comparable words and the word W is n-divisible.*

Proof. Suppose that there exist $2n-1$ subwords such that a period repeats more than n times in every such subword in some word W. Let x be one of the periods of these words. Then there exist nonoverlapping subwords $x^{p_1}v'_1, \ldots, x^{p_{2n-1}}v'_{2n-1}$ that are comparable with the word x, where $p_1, \ldots, p_{2n-1}$ are some positive integers that are bigger than n and the words $v'_1, \ldots, v'_{2n-1}$ are of length $|x|$. Then there exist n words from the set $\{v'_1, \ldots, v'_{2n-1}\}$ such that they are either n lexicographically greater than or n lexicographically less than x. We may assume that the words $v'_1, \ldots, v'_n$ are lexicographically greater than x. Then there exist subwords $v'_1, xv'_2, \ldots, x^{n-1}v'_n$, which are placed in lexicographically decreasing order. Thus, the lemma is obtained. □

From this lemma, we obtain Corollary 1.28.

Consider a non-n-divisible word W. Suppose that it contains such a subword that this subword contains some acyclic period x of length no less than n that repeats for more than $2n$ times. Then tails of x^2 that begin from the first, second,..., nth positions are pairwise comparable. So the word x^{2n} is n-divisible. We obtained a contradiction with non-n-divisibility of the word W. By Lemmas 6.13 and 6.12, we obtain that the essential height of the word W is less than

$$(2n-1)\sum_{t=1}^{n-1} \beth(t, l, n) < 8(l+1)^n n^5(n-1).$$

So $\Upsilon(n, l) < 8(l+1)^n n^5(n-1)$. Thus, we obtain Theorem 1.27.

REFERENCES

1. A. Ya. Belov, "On a Shirshov basis for relatively free algebras of complexity n," *Mat. Sb.*, **135**, No. 31, 373–384 (1988).
2. A. Ya. Belov, "Some estimations for nilpotency of nil-algebras over a field of an arbitrary characteristic and height theorem," *Commun. Algebra*, **20**, No. 10, 2919–2922 (1992).

3. A. Ya. Belov, "Rationality of Hilbert series for relatively free algebras," *Usp. Mat. Nauk*, **52**, No. 2, 153–154 (1997).
4. A. Ya. Belov, "Gelfand–Kirillov dimension of relatively free associative algebras," *Mat. Sb.*, **195**, No. 12, 3–26 (2004).
5. A. Ya. Belov, "Burnside-type problems, height and independence theorems," *Fundam. Prikl. Mat.*, **13**, No. 5, 19–79 (2007).
6. A. Ya. Belov, V. V. Borisenko, and V. N. Latyshev, "Monomial algebras," *J. Math. Sci.*, **87**, No. 3, 3463–3575 (1997).
7. A. Ya. Belov and M. I. Kharitonov, "Subexponential estimates in Shirshov theorem on height," *Mat. Sb.*, **203**, No. 4, 81–102 (2012), `arXiv:1101.4909`.
8. A. Ya. Belov and L. H. Rowen, *Computational Aspects of Polynomial Identities*, Research Notes Math., Vol. 9, Peters, Wellesley (2005).
9. J. Berstel and D. Perrin, "The origins of combinatorics on words," *Eur. J. Combin.*, **28**, 996–1022 (2007).
10. I. I. Bogdanov, "Nagata–Higman theorem for semirings," *Fundam. Prikl. Mat.*, **7**, No. 3, 651–658 (2001).
11. Gh. Chekanu, "Local finiteness of algebras," *Mat. Issled.*, **105**, 153–171, 198 (1988).
12. G. P. Chekanu, "Independence and quasiregularity in algebras," *Dokl. Akad. Nauk*, **337**, No. 3 (1994).
13. G. P. Chekanu and E. P. Kozhukhar, "Independence and nilpotency in algebras," *Izv. Akad. Nauk Respub. Moldova Mat.*, No. 2, 51–62, 92–93, 95 (1993).
14. Ye. S. Chibrikov, "On Shirshov height of a finitely generated associative algebra satisfying an identity of degree four," *Izv. Altaisk. Gos. Univ.*, No. 1 (19), 52–56 (2001).
15. Gh. Ciocanu, "Independence and quasiregularity in algebras. II," *Izv. Akad. Nauk Respub. Moldova Mat.*, No. 70, 70–77, 132, 134 (1997).
16. *Dnestr Copy-Book: A Collection of Operative Information* [in Russian], Inst. Mat. SO AN SSSR, Novosibirsk (1993).
17. V. Drensky, *Free Algebras and PI-Algebras: Graduate Course in Algebra*, Springer, Singapore (2000).
18. V. Drensky and E. Formanek, *Polynomial Identity Ring*, Adv. Courses Math., Birkhäuser, Basel (2004).
19. A. Kanel-Belov and L. H. Rowen, "Perspectives on Shirshov's height theorem," in: *Selected Papers of A. I. Shirshov*, Birkhäuser, Basel (2009), pp. 3–20.
20. A. R. Kemer, "Comments on Shirshov's height theorem," in: *Selected Papers of A. I. Shirshov*, Birkhäuser, Basel (2009), pp. 41–48.
21. M. Kharitonov, "Estimations of the particular periodicity in case of the small periods in Shirshov height theorem," `arXiv:1108.6295`.
22. M. I. Kharitonov. "Two-sided estimates for essential height in Shirshov height theorem," *Vestn. Mosk. Univ. Ser. 1 Mat. Mekh.*, No. 2, 24–28 (2012).
23. M. I. Kharitonov, "Estimates for structure of piecewise periodicity in Shirshov height theorem," *Vestn. Mosk. Univ. Ser. 1 Mat. Mekh.*, to appear.
24. A. A. Klein, "Indices of nilpotency in a PI-ring," *Arch. Math.*, **44**, No. 4, 323–329 (1985).
25. A. A. Klein, "Bounds for indices of nilpotency and nility," *Arch. Math.*, **74**, No. 1, 6–10 (2000).
26. A. G. Kolotov, "An upper estimate for height in finitely generated algebras with identities," *Sib. Mat. Zh.*, **23**, No. 1, 187–189 (1982).
27. Ye. N. Kuzmin, "On Nagata–Higman theorem," in: *Collection of Papers Dedicated to 60th Anniversary of Acad. Iliev*, Sofia (1975), pp. 101–107.
28. V. N. Latyshev, "On Regev's theorem on indentities in a tensor product of PI-algebras," *Usp. Mat. Nauk*, **27**, 213–214 (1972).
29. V. N. Latyshev, "Combinatorial generators of polylinear polynomial identities," *Fundam. Prikl. Mat.*, **12**, No. 2, 101–110 (2006).
30. A. A. Lopatin, *On the Nilpotency Degree of the Algebra with Identity* $x^n = 0$, `arXiv:1106.0950v1`.

31. M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge Univ. Press, Cambridge (2002).
32. S. P. Mishchenko, "A variant of the height theorem for Lie algebras," *Mat. Zametki*, **47**, No. 4, 83–89 (1990).
33. S. V. Pchelintsev, "The height theorem for alternative algebras," *Mat. Sb.*, **124**, No. 4, 557–567 (1984).
34. C. Procesi, *Rings with Polynomial Identities*, New York (1973).
35. Yu. P. Razmyslov, *Identities of Algebras and Their Representations* [in Russian], Nauka, Moscow (1989).
36. A. I. Shirshov, "On rings with identical relations," *Mat. Sb.*, **43**, No. 2, 277–283 (1957).
37. A. I. Shirshov, "On some nonassociative nilrings and algebraic algebras," *Mat. Sb.*, **41**, No. 3, 381–394 (1957).
38. V. A. Ufnarovsky, "An independence theorem and its consequences," *Mat. Sb.*, **128 (170)**, No. 1 (9), 124–132 (1985).
39. V. A. Ufnarovsky, "Combinatorial and asymptotical methods in algebra," *Itogi Nauki Tekh. Ser. Sovrem. Probl. Mat. Fund. Napr.*, **57**, 5–177 (1990).
40. V. A. Ufnarovskii and G. P. Chekanu, "Nilpotent matrices," *Mat. Issled.*, **85**, 130–141, 155 (1985).
41. E. Zelmanov, "On the nilpotency of nil algebras," in: L. L. Avramov and K. B. Tchakirian, eds., *Proc. 5th Nat. School in Algebra Held in Varna, Bulgaria, Sept. 24 – Oct. 4, 1986*, Lect. Notes Math., Vol. 1352, Springer, Berlin (1988), pp. 227–240.

Alexei Belov
Moscow Institute of Open Education, Moscow, Russia,
Jacobs University Bremen, Bremen, Germany
E-mail: `kanel@mccme.ru`

Mikhail Kharitonov
Moscow State University, Moscow, Russia
E-mail: `mikhailo.kharitonov@gmail.com`