



Article Combinatorial Estimations on Burnside Type Problems

Anton Beletskiy^{1,*} and Ilya Ivanov-Pogodaev²

- ¹ Faculty of Mathematics, HSE University, 101000 Moscow, Russia
- ² School of Applied Mathematics and Computer Science, MIPT University, 141701 Moscow, Russia; ivanov-pogodaev@mail.ru
- * Correspondence: aabeletskiy@edu.hse.ru

Abstract: The Burnside problem, formulated by W. Burnside in 1902, is one of the most well-known and important open questions in the field of Group Theory. Despite significant progress made in the past century towards solving this problem, its complete solution remains unknown. In this paper, we investigate one of the approaches to solving the Burnside problem based on the application of an iterative theory of small cancellations and canonical forms developed by E. Rips in recent years. We present a novel self-contained exposition of this theory and utilize it to obtain new estimates on the infiniteness of initial approximations of Burnside groups where only a finite number of periodic relations is used for relatively small odd exponents (n > 120).

Keywords: Burnside; small cancellations; group theory

MSC: 20F05; 20F06; 20F65

1. Introduction

1.1. A Brief History of the Burnside Problem

The Burnside problem is a question in Group Theory that was originally formulated by W. Burnside in 1902 [1] and has since become one of the most well-known open problems in the field (quoting [2], "the comparison suggests itself of the influence of the Burnside problem on combinatorial group theory with that of Fermat's last theorem on the development of algebraic number theory"). In the last century, several modifications as well as alternative formulations of the problem have been suggested and extensively studied (for a survey of some of them, see, e.g., [3,4]). Despite the remarkably concise and natural formulation of the original problem, its complete solution remains unknown, although significant progress in this area has been achieved.

Let us state here the original problem posed by W. Burnside. Consider the free group generated by *m* symbols, denoted by a_1, \ldots, a_m . Now, for each word *x* in this alphabet (that is, for each element of the group), we impose the relation $x^n = e$, where *e* is the identity element. The quotient group obtained after imposing all such relations is called the Burnside group of rank *m* and exponent *n*, denoted by B(m, n).

The Burnside problem is the question of the finiteness of B(m, n) for a given m and n. Alternatively, we can ask about the finiteness of any group G generated by m elements, in which every $g \in G$ satisfies the relation $g^n = e$. It is clear that such a formulation of the problem is equivalent to the original one, as every such group is a factor group of B(m, n).

In 1902, Burnside himself proved (see [1], where the problem was originally formulated) that Burnside groups of exponent 3, that is, B(m, 3), are finite for $m \ge 2$. He also made some progress regarding the finiteness of *linear* (matrix) groups with similar relations [5], a problem which was later fully solved by I. Schur [6]. In 1940, Sanov (see [7]) proved the finiteness of B(m, 4). Then, after another 18 years, in 1958, the finiteness of B(m, 6) was established (by M. Hall [8]). This essentially exhaustively covers positive results in this



Citation: Beletskiy, A.; Ivanov-Pogodaev, I. Combinatorial Estimations on Burnside Type Problems. *Mathematics* 2024, 12, 665. https://doi.org/10.3390/ math12050665

Academic Editor: Bo Zhou

Received: 25 December 2023 Revised: 12 February 2024 Accepted: 20 February 2024 Published: 24 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). area. Thus, the question of the finiteness of B(m, n) for other small values of the exponent, including 5, remains open.

On the other hand, in 1965 E. S. Golod proved that the analog of the Burnside group of an *unbounded exponent* can indeed be infinite [9]. Other ingenious examples of constructing infinite finitely generated groups with similar properties (i.e., each element is periodic, but the exponent of the group is unbounded) can be found in [10–12].

Nevertheless, the most significant progress in resolving the Burnside problem was achieved by P. S. Novikov and S. I. Adian, who, in a series of joint works [13], proved that the group B(m, n) is infinite for all odd n > 4381 and $m \ge 2$. They accomplished this result by introducing a new generalization of small cancellation theory (see Section 2.1), which allowed them to consider the structure of relations in Burnside groups as a graded set, where within each rank, the relations satisfy the small cancellation condition. The direct proof of the result relied on a simultaneous inductive proof of more than a hundred related statements that established the necessary conditions for all ranks simultaneously.

It is worth noting that this work has since been recognized as "possibly the most difficult-to-read mathematical paper ever written" [14]. Furthermore, Adian later managed to lower the bound on the exponent from 4381 to 665, proving the infiniteness of B(m, n) for all odd n > 665 [15]. This result essentially provided technical refinements to the previous work, including more precise estimates of various parameters. Currently, these results represent the best-known estimates in this field, despite ongoing research. See, for example, the lecture by A. Atkarskaya [16], where she explores potential methods for further reducing the existing bounds, and also a recent preprint of a paper by A. Atkarskaya, E. Rips, and K. Tent [17].

In addition to P. S. Novikov and S. I. Adian, A. Y. Ol'shanski also worked on the Burnside problem—approaching it from a more geometric standpoint stemming from the analysis of van Kampen diagrams. His results are presented in the book [18]. It should be noted that despite the more intuitive nature of this approach, Ol'shanski's final estimates were significantly weaker than the already known estimates by Novikov and Adian (specifically, in [18], the infiniteness of B(m, n) is proven for odd $n > 10^{10}$).

Finally, much later (starting from 1982), E. Rips developed an alternative theory by working with iterative small cancellation theory (Refs. [19,20] and subsequent articles by the authors, such as [17,21]). This theory was based on his development of canonical forms for hyperbolic groups in the sense of Gromov [22]. In this work, we directly employ the apparatus developed by Rips and apply it to an analysis of Burnside groups as well as approximations of these groups obtained by factoring only by relations of a finite number of lower ranks (further elaborated on later).

A more detailed historical overview along with a deeper survey of the current state of this field can be found in the review by S. Adian [23], which also attempts to informally explain the essence of the inductive argument used by Adian and P. Novikov to prove the infiniteness of Burnside groups.

1.2. Goals and Objectives of the Present Paper

One of the planned goals of this project is the analysis of Burnside groups and other similar algebraic structures, including the use of computer methods. It can be observed that, especially in light of the recent advancements made by E. Rips [17], the analysis of these structures allows for a natural formulation in a "computer" or algorithmic language.

We present several intuitive justifications for our confidence in obtaining results in this area through computer analysis:

- Generalizations of small cancellation conditions (see Section 2.1) for Burnside groups allow for efficient algorithmic solutions to the word problem in these groups.
- Relations in Burnside groups enable the introduction of a graded structure, which allows for an iterative approximation of B(m, n) as a sequence of nested groups that introduce more and more relations. Each of these groups has only a finite number of relations, which can be effectively described for computer analysis.

• The apparatus of canonical forms developed by E. Rips provides a direct algorithm for constructing the canonical form of each word. Working with these canonical forms significantly simplifies the structure of the resulting Van Kampen diagrams. Although this theory does not initially involve manual (and rather laborious) computation of these forms, this approach is very natural and convenient for computer processing.

A preliminary step for working in this direction is the study, formalization, and rigorous presentation of the theory of canonical forms developed by Rips. The most substantial part of this work (Section 3) in terms of both volume and content is dedicated to this task. The size of this section may appear excessive, but it is necessary because the subsequent analysis will be based on the considerations described here, and therefore it is important to provide a sufficiently complete and rigorous exposition of the underlying theory.

Since the first nontrivial case that requires serious analysis is the restriction of relations to the first two ranks, the second task accomplished in this work is to refine Rips' estimates in the case when relations in the group are divided into ranks 1 and 2 (see Section 4). From the theory developed in Section 3.3, we obtain, for example, the result of the infinite-ness of the Burnside group when considering only the first two ranks of relations, even for relatively small exponents (odd n > 120). This result firstly serves as evidence of the potential of this theory for analyzing approximations of Burnside groups, and secondly, it significantly limits the number of cases requiring more detailed analysis.

Finally, in the last section (Section 5), we present some plans for further research.

2. Preliminaries

A part of this paper, which is dedicated to the exposition of previous advancements in this field upon which our work relies as well as other preliminary information, consists of two main sections that significantly differ in complexity and the nature of the presented material. In this section, we provide a brief (and quite informal) introduction to the so-called small cancellation theory. This area of geometric group theory was developed independently of the Burnside problem (and consequently has other applications; see, e.g., [24]), yet almost all progress achieved in the analysis of the structure of B(m, n) for large exponents has been based on considerations stemming from this theory. Therefore, we consider it necessary to present here a concise overview of some of its basic ideas and propositions.

The second part of what formally constitutes the preliminary information (Section 3, the iterative small cancellation theory by E. Rips) represents an exposition of the generalization as well as an alternative approach to the small cancellation theory developed by E. Rips. This exposition allows for a substantial simplification of the inductive analysis of Burnside groups. Although a significant part of this theory is presented in the publication [20] and later in [17,21], the latest advancements and concepts formulated by Rips have not been published yet.

Therefore, we deem it reasonable to include this text in a separate section since, despite the fact that these facts are already known (albeit to an extremely narrow circle of individuals), as far as the authors of this paper is aware, this text represents the first written and self-contained presentation of Rips' perspective formulated in this course. Hence, we consider the appearance of this exposition as one of the results of this work.

2.1. Small Cancellation Theory

Small cancellation theory is a branch of geometric group theory that studies groups satisfying the so-called small cancellation condition. Informally formulated as "distinct relations do not contain overly long common subwords", this condition allows for a series of statements about the hyperbolic structure of Van Kampen diagrams in such groups (see [25] for the earliest exposition of these ideas). Consequently, it enables the application of geometric methods to prove, for example, the solvability of the word problem in these groups. One of the best expositions of this theory, according to the authors of this paper, can be found in the book by A. Y. Ol'shanskii [18].

Small cancellation theory plays a key role in all known methods of obtaining estimates for the infiniteness B(m, n) for large exponents and is based on its generalization and iterative application.

In this section, we provide a brief introduction sufficient for understanding this work, omitting almost all proofs. Additionally, in several places where it does not affect the comprehension of the overall picture but facilitates the construction of an intuitive understanding of this field, we intentionally allow some imprecision in our definitions, guided by ease of understanding. Of course, rigorous definitions and proofs can be found in any textbook covering this branch of group theory, such as [18].

Let us consider a group *G* generated by a (finite) set of elements $A = \{a_1, \ldots, a_m\}$. If *G* is not free, there exist *relations* between the elements a_i : words in the alphabet $A \sqcup A^{-1}$ that equal *e* in *G* (often denoted simply as 1 in the group *G*). For example, \mathbb{Z}^2 can be regarded as a group generated by symbols a_1, a_2 with a single relation $a_1a_2a_1^{-1}a_2^{-1}$.

Naturally, it is not necessary to explicitly specify *all* relations in a group. Typically, it is sufficient to specify a small subset from which all the others necessarily follow. More precisely, we provide the following definition:

Definition 1. A relation w follows from relations v_1, \ldots, v_n if and only if there exist words u_1, \ldots, u_n such that

$$w = u_1^{-1} v_1^{\pm 1} u_1 \cdot u_2^{-1} v_2^{\pm 1} u_2 \cdot \ldots \cdot u_n^{-1} v_n^{\pm 1} u_n.$$

It is clear that this constructive definition is equivalent to w = 1 in any group *G* satisfying relations v_1, \ldots, v_n .

However, thus far we have not directly addressed the small cancellation condition. To introduce it, we formulate another equivalent way of looking at relations and the deducibility of one relation from others.

Recall that a *Van Kampen diagram* over a group *G* is a planar directed graph, where each edge is labeled with a generator a_i , and for each cell in the graph, the word obtained by reading its boundary is one of the generated relations. Here, reading the boundary means concatenating the generators written on the edges of the boundary, where whenever an edge is traversed in the opposite direction, the generator is included in the word with a negative exponent.

Lemma 1 (Van Kampen Lemma). A relation w follows from relations v_1, \ldots, v_n if and only if there exists a Van Kampen diagram over G where the boundaries of each cell yield cyclic shifts of the words v_i or their inverses and the outer boundary of the entire diagram yields the word w.

To provide an intuitive justification for this lemma, we think the following example (Figure 1) is best suited (adapted from [26]):



Figure 1. Simple Van Kampen diagram in group \mathbb{Z}^2 generated by *x* and *y* with the relation [x, y] = 0. Blue arrows stand for *x*, and green arrows stand for *y*.

On the left in the figure above, we can see a "geometric" derivation of the word $x^{-2}y^{-2}x^2y^2$ from the defining relation $[x, y] = x^{-1}y^{-1}xy$. On the right, we can see the decomposition of that Van Kampen diagram, which corresponds to the algebraic derivation:

$$x^{-2}y^{-2}x^{2}y^{2} = (x)^{-1}[x,y](x)(yx)^{-1}[x,y](yx)[x,y](y)^{-1}[x,y](y)$$

The formal proof of the Van Kampen lemma in its essence consists of the formalization of this "decomposition of a diagram" operation.

For a somewhat more interesting example, let us consider the following diagram (see Figure 2).



Figure 2. Van Kampen diagram arising in the proof of finiteness of *B*(2,3). Again, blue arrows stand for *x* and green ones for *y*, but this time, the relations are of the form $g^3 = e$.

Each cell in this diagram corresponds to the relation $w^3 = 1$, where w is equal to either x, y, or $x^{-1}y$. The outer boundary gives us the relation [[x, y], y] = 1. The deduction of this relation is a key step in proving the finiteness of the group B(n, 3) of Burnside exponent 3 (see [1]).

The theory of small cancellations connects the algebraic nature of relations in groups with the geometric language of Van Kampen diagrams. Here and further, we fix a set of relations \mathcal{R} . We assume that this set is closed under taking inverses and cyclic shifts.

We call a *piece* a word *w* that is part of two distinct relations in \mathcal{R} . Formally, this means that there exist relations $R_1, R_2 \in \mathcal{R}$ such that $R_1 = wr_1, R_2 = wr_2$. Finally, we say that $G = \langle A | \mathcal{R} \rangle$ satisfies the *small cancellation condition* $C'(\lambda)$ if for every piece *w* and every relation *R* containing it, we have $|w| < \lambda |R|$.

Intuitively, this condition asserts that any two generated relations in our group differ from each other sufficiently (they do not have long common subwords), and thus, each cell in the Van Kampen diagram has many (at least $\frac{1}{\lambda}$) neighbors.

Finally, we can state one of the main results in this area [27].

Theorem 1 (Greendlinger's Lemma). If \mathcal{R} satisfies the condition $C'(\frac{1}{6})$ and the (cyclically reduced) relation w = 1 follows from \mathcal{R} , then there exists a cyclic shift \bar{w} of w denoted by \bar{w} and a relation $r \in \mathcal{R}$ for which the common prefix u of the words \bar{w} and r has length $|u| > \frac{1}{2}|r|$.

Although this result may appear quite technical, its importance is hard to overestimate. Indeed, Theorem 1 enables us to construct a greedy algorithm that solves the word equality problem in *G*. To see how this becomes possible, note that it suffices for us to be able to check whether a given word *w* equals 1 in *G*. However, if w = 1, then by Greendlinger's lemma, we can find a relation *R* for which more than half of *R* occurs as a subword *u* in *w*. If R = us, then we can replace *u* with s^{-1} and reduce the resulting word. The length of the word *w* will then decrease, and the claim is then proved by an inductive argument.

This algorithm is known as the *Dehn algorithm* (for a thorough treatment of these ideas, see [28,29]), and its existence for groups with small cancellations is one of the reasons that makes their study interesting.

Let us finish this section off by noting that the constant $\frac{1}{6}$ in the Greendlinger's lemma above is exact—that is, for each constant $\lambda > \frac{1}{6}$, we can construct a group satisfying $C'(\lambda)$ and a Van Kampen diagram in that group that does not satisfy Greendlinger's lemma statement. To illustrate this point, consider the condition $C'(\frac{1}{5})$ and the following diagram (this example is adapted from [26]):

Indeed, every cell in Figure 3 has a perimeter of at least 10, and every piece has length of 2. Still, we can derive from these relations a word of length 2 (by applying Lemma 1 to the diagram in Figure 3), which would clearly contradict Lemma 1.



Figure 3. Van Kampen diagram in the group satisfying $C'(\frac{1}{5})$ with a perimeter of just 2, which contrasts with the statement of Lemma 1.

3. Iterative Small Cancellation Theory by E. Rips

In this section, we provide a brief self-contained exposition of the iterative small cancellation theory developed by E. Rips starting from [20] (1997) and further developed in subsequent works [17,21]. However, only a part of the constructions presented in the following subsections has been formally described by Rips in his publications. The remaining part of his results has not been published and was presented in a lecture course given by Rips at Tel Aviv University [30]. Since the notes from these lectures have only recently (with the direct involvement of A. Kanel-Belov and, to a lesser extent, the authors of this work) become publicly available, this work presents these materials in written form for the first time. It should be noted that the course given by Rips was quite informal and consisted mainly of intuitive explanations sufficient for subsequent rigorous development of the underlying theory. Therefore, understanding and rigorously proving his results, as well as their written exposition in this work, are, firstly, an important part of the progress in the field of the Burnside problem and generalizations of small cancellation theory in general and, secondly, a significant result of this work itself.

3.1. Diagrams for Semi-Canonical Words

So, we are working with a presentation of a group: $G = \langle A | \mathcal{R} \rangle$, where *A* is a (finite) set of generators and \mathcal{R} is a set of relations, which we will assume to be closed under taking inverses and cyclic shifts.

Our ultimate goal is to construct a canonical form for each word in *G* that would simplify the process of analyzing Van Kampen diagrams for the group. As a first step towards this goal, we introduce the concept of \mathcal{R} -diagrams.

Let us fix a constant $\tau \in (0, 1)$ (it is productive to think of τ as a sufficiently small number). Also, for each relation $R \in \mathcal{R}$, we define the relative length of each subword *S* in R as $L_R(S) = \frac{|S|}{|R|}$.

Definition 2. For a word W in the alphabet $A \cup A^{-1}$, we call an \mathcal{R} -diagram of W the collection of all subwords S (we distinguish the occurrence of the same subword at different positions in W) such that W = ASB and $S = R^{\alpha}R_1$, where $R \in \mathcal{R}$, $\alpha \ge 0$, $R = R_1R_2$, and the relative length $L_R(S) \ge \tau$. We call a diagram maximal if it contains only maximal subwords S in terms of inclusion.

This definition may seem somewhat convoluted, but in reality, it is quite natural. For a word W, we are interested in the potential possibilities for its cancellations, and each such possibility corresponds to a piece of relation from \mathcal{R} that occurs in our word. Moreover, even a repeated occurrence of a relation is considered as a potential cancellation opportunity. Finally, if we do not limit the length of the subwords of interest from below, we would include too many uninteresting words in the \mathcal{R} -diagram—for example, every symbol occurring in any relation. To avoid this, we only consider subwords of relations with a relative length of at least τ .

We assume that our group satisfies the conditions of small cancellation with a constant ε (this assumption will be implied everywhere until the consideration of graded systems of relations begins). In our terms, this means that if $R_1 = SP \in \mathcal{R}$, $R_2 = SQ \in \mathcal{R}$, then $L_{R_i}(S) < \varepsilon$, i = 1, 2. Moreover, we require the inequality $2\varepsilon < \tau$. This condition gives us an important restriction on the possible structure of maximal \mathcal{R} -diagrams:

Lemma 2. Under these conditions, in the maximal \mathcal{R} -diagram of a word W, no letter is included in 3 or more subwords S simultaneously.

Proof. Let S_1 , S_2 be subwords in the maximal diagram, let R_1 and R_2 be their respective relations, and let P be the area of their intersection: $S_1 = AP$, $S_2 = PB$. In this case, the relative length $L_{R_1}(P) < \varepsilon$, $L_{R_2}(P) < \varepsilon$. Let some third subword Q intersect P. From maximality, Q cannot be contained in S_2 . Thus, it must extend beyond the boundaries of S_2 either to the left (Figure 4, left) or to the right (Figure 4, right).



Figure 4. *S*₁ shown in blue, *S*₂ in green, and *Q* shown in red.

The first case is impossible, as the relative length of Q is at least $\tau > 2\varepsilon$, and either $Q \cap S_1$ or $Q \cap S_2$ would have a relative length greater than ε . The second case is impossible, as then Q would cover $S_2 \setminus P$, but $L_{R_2}(S_2 \setminus P) = \tau - L_{R_2}(P) > \varepsilon$. Thus, the intersection of three maximal subwords at one point is impossible. \Box

At this point, we have introduced the constants ε and τ , which are useful to think of as numbers close to zero. For further reasoning, we will also need the constants λ_0 and λ_1 , subject to the following conditions:

$$1 - \lambda_0 > \tau + 2\varepsilon$$

$$\lambda_0 - \lambda_1 > 2\varepsilon$$

$$\lambda_1 - \frac{1}{2} > 2\varepsilon$$

$$\tau > 2\varepsilon$$
(1)

The technical details that make these conditions necessary will become clear later. Thus, it is useful to think of λ_i as large numbers that are nevertheless distinct from 1 and $\frac{1}{2}$



Figure 5. The diagram of constants included in the following arguments.

Definition 3. A word W is called λ -semicanonical (denoted $W \in SC(\lambda)$) if for every subword S (and relation R whose subword is S) in the maximal \mathcal{R} -diagram of W, we have $L_R(S) < \lambda$.

Thus, we are interested in words where all occurrences of defining relations are sufficiently small. It is natural to study such words because if a word contains a large portion of a relation, it can be effectively canceled. The remainder of this subsection will be devoted to establishing the main properties of semicanonical words, constructing a group structure on them, and proving the equivalence between the obtained group and the original group *G*.

It is evident that the primary tool for manipulating words in *G* is finding a subword *S* such that there exists a relation of the form $R = SP \in \mathcal{R}$ and replacing *S* with the subword P^{-1} . We refer to each such replacement as an \mathcal{R} -turn. The schematic representation we will use to depict this transformation is shown in Figure 6. The black line corresponds to the original word, and the green line represents the resulting word.



Figure 6. Schematic representation of a turn on the word *S* (blue) using the relation R = SP. The resulting word is shown in green.

Furthermore, different λ -semicanonical words can actually be equivalent in group *G*. To address this issue, we introduce the following:

Definition 4. Two words W_1 and W_2 in $SC(\lambda_1)$ are called equivalent if there exists a sequence of \mathcal{R} -turns transforming the word W_1 into the word W_2 such that all intermediate words belong to $SC(\lambda_0)$.

Note that although this is indeed an equivalence relation (transitivity, reflexivity, and symmetry are evident from the construction), it is initially not obvious why such a definition (specifically, the requirement on λ_0 -semicanonicity of intermediate words) is natural and reasonable. Nevertheless, everything becomes much clearer after considering the following diagram in Figure 7:



Figure 7. Two *R*-turns; resulting word shown in red.

Here, the horizontal line denotes a sequence of characters in the word W_1 . The green and blue strips represent the elements S_1 and S_2 , respectively, of the maximal diagram. Suppose the first transformation we make is an \mathcal{R} -turn with respect to the first (green) subword. The second turn is with respect to the subword S_2 . The resulting word W_2 is indicated by the purple line. However, in this case, as can be seen from the diagram, after the first turn, the intersecting elements of the maximal diagram with S_1 can change because the intersection $S_1 \cap S_2$ is no longer present in the word. In this particular case, S_2 decreases, but it is clear that some turns (e.g., the inverse of the given one) can increase the length of neighboring subwords in the diagram. However, since at each stage two distinct words intersect by at most ε , the length of each subword *S* in the diagram can change by no more than 2ε (due to the turns immediately to the right and left of *S*). Thus, by requiring a sufficiently large difference between λ_1 and λ_0 (at least 2ε), we ensure that any sequence of turns on subwords that were initially part of the diagram does not create in the diagram subwords with a relative length greater than $\lambda_1 + 2\varepsilon$.

On the other hand, it is easy to see that each turn allowed by our definition is a turn using one of the subwords in the initial maximal diagram W_1 . Indeed, every turn replacing S with P^{-1} must satisfy $L_R(S) > \tau + 2\varepsilon$ (otherwise, the resulting word does not belong to $SC(\lambda_0)$). However, the subwords of S consisting of subwords from other relations have a relative length of at most 2ε (see the proof of Lemma 2). Thus, S has a part with a size of at least τ that does not come from other relations, which means that S was included in the initial maximal \mathcal{R} -diagram.

Finally, the first condition on the constants $(1 - \lambda_0 > \tau + 2\varepsilon)$ is necessary for each new subword to be included in the diagram again (otherwise, it may become too small and disappear from the diagram altogether, as we consider only segments of relations with a relative length of at least τ .

Combining Lemma 2 and Definition 4, we obtain the following lemma:

Lemma 3. In the conditions of Definition 4, the result of applying a sequence of turns to the subwords S_1, \ldots, S_k in the maximal \mathcal{R} -diagram W does not depend on the order of their application.

Proof. Note that the lemma statement is meaningful since the condition $1 - \lambda_0 > \tau + 2\varepsilon$ implies that all turns are indeed performed on some subwords in the initial diagram (possibly lengthened or shortened by $\leq 2\varepsilon$ due to previous turns). We proceed by induction on the number of turns. It is clear that it suffices to show that the new turn commutes with the last one made. If the subword used for the new turn is separated by at least ε symbols from the previous turn, commutativity is obvious. Otherwise, we have the situation depicted in diagram 7, and commutativity is evident from considering the diagram: applying turns 1 and 2 in any order leads to the word marked by the purple line.

Establishing that the result does not depend on the order of turns but only on the subwords on which we perform these turns, and based on Lemma 2, which postulates that all subwords in the maximal diagram are arranged sequentially and intersect by no more than ε , we can schematically depict each equivalence class of some λ_1 -canonical word in a more symmetric form. A typical example of such a depiction is shown in Figure 8. Here, each cell of the graph corresponds to a relation from \mathcal{R} and the corresponding subword in the maximal \mathcal{R} -diagram W (the reverse is also true: for each subword in the maximal \mathcal{R} -diagram, we obtain the corresponding segment in the one-layer map). It is clear that the length of each horizontal segment does not exceed λ_0 , and the length of each vertical bridge does not exceed ε .



Figure 8. One-layer map of an equivalence class of a λ_1 -canonical word.

Each word in the equivalence class *W* is represented as a path going from left to right in this diagram—choosing either the upper or lower segment for each segment. Each turn performed during the equivalent transformations of *W* is now represented as "switching" of the path from one alternative segment to another, and it is evident from this representation that the turns commute with respect to different subwords in the diagram. We call such a representation a *one-layer map* for the equivalence class of some λ_1 -canonical word W. These maps will be the basis for all subsequent reasoning.

The next step is to construct a group structure on the set of equivalence classes in $SC(\lambda_1)$ (i.e., on the set of one-layer maps). It is clear that we want to use the same map read in reverse as the inverse element in this group. However, if we define the multiplication of two maps as their immediate concatenation, the product of a map with its inverse will not be equal to the identity (the empty map). Therefore, a more elegant algorithmic definition of multiplication for two maps is introduced.

To multiply the map *C* by the map *D*, we start by comparing the maps C^{-1} and *D* symbol by symbol. We continue the comparison until we find the first place (either a symbol not participating in the \mathcal{R} -diagram or a cell in the map) where a difference occurs. Next, several possibilities arise:

The common part of C^{-1} and D is depicted in black, while the corresponding extensions are depicted in blue and green.

If the first variant is realized (the distinguishing point is a symbol not included in the \mathcal{R} -diagram), then the product of the maps is the concatenation of the blue and green parts. If the second variant is realized (the distinguishing point is a cell of the map, but the two possible extensions have disjointed subwords as bridges to this cell), then the product is obtained by concatenating the blue and green parts, with the insertion in the middle of one or both of the paths passing through the central cell depending on whether one or both of the lengths of these paths satisfy the condition $\leq \lambda_1$. Finally, in the third case, only the "short" segment of the central cell will be included in the product since its relative length does not exceed 2ε , which means that the length of the alternative path is too large.

It can be seen that with this multiplication algorithm, the map read in reverse indeed becomes the inverse element. It remains to verify that the introduced multiplication operation is associative, which is quite obvious (it is sufficient to note that the disappearance of the "merged" ends of the diagram does not affect the central cell, and therefore, the tails can be deleted in any order).

Thus, we have introduced a group structure on the set of one-layered maps. It is possible to prove the isomorphism of the group (denoted by G_1) obtained in this way with the original group *G*. One undeniable advantage of working with G_1 instead of *G* is that we have eliminated the freedom to choose different representations of the same element of the group *G*, as now different maps correspond to different group elements. The next (and crucial) step in building the theory is the introduction of the so-called *canonical form* of each element of G_1 : that is, the selection of a specific distinguished path for each map.

3.2. Canonical and Certified Forms

Before proceeding to the (quite counterintuitive) construction of canonical forms, let us try to formulate some motivations for this construction.

The idea of choosing a canonical representative for each one-layered map may seem somewhat useless in a sense—after all, the entire previous section was devoted to constructing equivalence classes from the original elements of the group *G*, from which we now want to return to considering specific elements. However, this idea becomes much more appealing if we recall the main idea of working with Burnside groups (which is our ultimate goal). This idea consists of grading the set of relations in such a way that within each rank, the relations satisfy the small cancellation conditions, and the relations in the previous rank are in some sense "small" compared to the subsequent ranks. In this case, the van Kampen diagrams, in general, begin to resemble the one shown in Figure 9:



Figure 9. Three-layer van Kampen diagram.

Cells of each specific rank are adjacent to each other not "ideally" but with layers of cells of lower rank in between. In such a situation, despite the fact that the higher rank satisfies the small cancellation condition, we cannot make any statements (in the spirit of Lemma 1) about the structure of the entire diagram as a whole because cells of higher rank do not abut each other closely.

However, let us assume that we have brought all the relations of rank 2 to canonical form with respect to rank 1 and repeated this operation inductively for all ranks. In this case, the boundaries of cells of the same rank, separated by a thin layer of cells of lower rank, will actually coincide since the word corresponding to the common boundary is written in canonical form on both sides. Hence, this word is the same on both sides (which means that we have chosen the same path on the one-layer map from both sides). Then, we would find that cells of higher ranks are actually directly glued to each other whenever they are close enough, allowing us to manipulate such structures much like ordinary small cancellation groups.

Now let us proceed directly to the formal construction. Let us consider a one-layer map (see diagram 8) of a λ_1 -canonical word. The idea is to choose one of the two horizontal segments—alternative paths—for each pair of segments and consider the path that goes along the selected segments as the canonical form of the diagram. We use the following criteria for the selection:

Definition 5. For two paths w_1 and w_2 in a one-layer map situated between bridges p_1 and p_2 (thus giving the relation $R = p_1 w_1 p_2^{-1} w_2^{-1}$), and given a set of generators x_1, \ldots, x_n , we consider path w_1 as preferable if:

- w_1 is shorter than w_2 in terms of the number of symbols.
- If w₁ and w₂ are equal in length, we compare the pairs w₁ and w₂ (respectively, w₂⁻¹ and w₁⁻¹) and compute the difference between their positions in the lexicographical ordering of all of the words of the same length in the alphabet x₁,..., x_n, x₁⁻¹,..., x_n⁻¹, denoted by d₁ (respectively, d₂). If |d₁| > |d₂|, we take the smaller of the words w₁ and w₂, and if |d₁| < |d₂|, we take the smaller of the words w₁ and w₂.
- If the words w_1 and w_2 are equal as character strings, we repeat the operation from the previous item for the words $p_1w_1, p_1^{-1}w_2$ and $p_2w_1^{-1}, p_2^{-1}w_2^{-1}$.

It is clear that this definition always allows us to obtain a single preferable path except in the case of $w_1 = w_2$, $p_1 = p_2$: that is, the case $R = (p_1w_1)^2$. Therefore, in the future, we assume that no relation is a complete square of a word.

Note 1. It is precisely here that the reason for drawing conclusions only about Burnside groups with odd exponents lies—otherwise, even in the first rank, we have a huge number of relations that are squares, which prevent us from unambiguously choosing canonical representatives.

Finally, let us note that we intentionally complicated the definition by including not only the words themselves but also their inverses in the comparison in order to ensure the equivariance of taking the canonical form with taking the inverse element in the group G_1 . Indeed, if we did not consider the lexicographic comparison of inverse elements, there could be a situation where when traversing a one-layer map in one direction, we prefer w_1 , and when traversing in the opposite direction, we prefer w_2^{-1} .

Definition 6. The canonical form of a one-layer map is defined as the result of concatenating all the preferable subwords (traversals of the one-layer map) connected by bridges. We denote the canonical form of a one-layer map C as can(C).

Note that for an arbitrary choice of representatives in each equivalence class, it is initially unclear how the canonical forms of the factors are related to the canonical form of the product. However, for our definition of the canonical form, the situation improves significantly. Specifically, we will prove the following lemma:

Lemma 4. $can(W_1)can(W_2)can(W_2^{-1}W_1^{-1}) = R_1^{\vee}R_2^{\vee}R_3^{\vee}R_4^{\vee}$, where R_i^{\vee} is the result of conjugating relation R_i by some word.

Thus, in a relation diagram composed of three canonical forms that multiply to give 1, there are at most 4 cells. Moreover, we will show that the diagram of such a "triangle" looks generally like the diagram in Figure 10.



Figure 10. General view of a triangle formed by three canonical words.

Proof. To prove this fact, let us first examine how the canonical form can change when one cell is modified in a one-layer map. It is clear that since we choose the preferred option independently in each cell, the canonical form can only change in the initial cell as well as in the case when new cells appear or disappear from our one-layer diagram. The only reason a cell can disappear is if one of its sides becomes too long (longer than λ_1) and therefore ceases to be included in the one-layer map (since we consider only sufficiently short subwords as options for transformations; see Definition 4). However, can the disappearance of cells affect cells in the one-layer map far from the one where the change occurred? Clearly, when moving through the diagram from the changed cell, if a certain cell survives, then all subsequent cells also remain unchanged, so the change can only spread along a sequence of cells, each of which disappears from the diagram. On the other hand, when a neighboring cell disappears, alternative paths are extended by no more than ε (because the vertical bridge is now included in one of the alternatives), so a chain of

disappearances occurs only when there is a chain of cells for which one of the sides has a relative length of at least $\lambda_1 - \varepsilon$. This situation is depicted in Figure 11.

Here, the blue path represents the canonical form in the initial map, and the cells in the chain on the right each have a length of at least $\lambda_1 - \varepsilon$ in the upper part. Then, when the red segment is added to the one-layer map, this chain of cells collapses, disappearing from the map. It is clear that since for each of these cells, the side that disappeared had a length of at least $\lambda_1 - \varepsilon > \frac{1}{2}$, the canonical form in this area does not change. Moreover, the only place where the canonical form can change is the first surviving cell. Moreover, the change can occur only if the relative lengths of both alternatives in this cell are in the interval $(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon)$, since otherwise, the disappearance of the neighboring cell would not have changed the preferred path.

Thus, despite the fact that a small change in the word can change the canonical form indefinitely far from the point of change, we can somehow "control" these changes: the canonical form changes only in one place, only in the presence of a chain of cells with near-maximal length of one of the paths, and only if at the end of this path there is a cell with relative side lengths close to $\frac{1}{2}$.



Figure 11. "Domino effect": change of the canonical form far from the word modification.

Taking this fact into account, the conclusion of the proof of Lemma 4 becomes obvious: indeed, according to the definition of the product in the group G_1 , we obtain a triangular construction from one-layer maps (see Figure 12) in which for each of the three sides, the change occurs only in the central (first differing) cell. Thus, the canonical forms of the initial words and their products can differ only in one place on each of the "rays" emanating from the central cell, which proves the statement. \Box



Figure 12. Possible arrangements of maps at the first point of difference.

However, our goal in introducing the canonical form is to achieve even more precise control over possible discrepancies. To achieve this, we will impose an additional condition on the maps under consideration.

First of all, let us note that Lemma 2 implies that all subwords in the maximal \mathcal{R} -diagram of a word are arranged sequentially (possibly with small overlaps) and can

therefore be ordered. Let us introduce two new constants, which we will denote by λ_2 and λ_3 , with conditions similar to those imposed on λ_1 :

$$\lambda_{1} - \lambda_{2} > 2\varepsilon$$

$$\lambda_{2} - \lambda_{3} > 2\varepsilon$$

$$\lambda_{3} - \frac{1}{2} > 2\varepsilon$$
(2)

Finally, let us fix a certain aperiodic sequence $l_1, l_2, l_3, ...$ of numbers 1 and 2. As a concrete example, the Morse–Thue sequence [31] can be used, which has the property that for no word *X* and symbol *a* does the sequence contain a subword of the form *aXaXa*, from which it follows that this sequence does not contain a cube of any word.

Definition 7. A reliable horizontal segment S of relation R in a one-layer map C is defined as follows: there exists a word W equivalent to the word represented by map C (i.e., a path in the map C) such that S is contained in the maximal \mathcal{R} -diagram of W, the relative length $L_R(S) \leq \lambda_3$, and for each n, the n-th subwords S_l (respectively, S_r) in the maximal \mathcal{R} -diagram encountered when moving left (respectively, right) in the diagram of W have a relative length $\leq \lambda_{l_n}$. In this case, W is called a witness for S.

This definition is not very intuitive, so let us examine it in more detail.

First, note that every preferred horizontal segment (i.e., one that belongs to can(*C*)) is reliable. Indeed, it is straightforward that can(*C*) serves as a witness for every segment in it (the length of each preferred segment does not exceed $\frac{1}{2}$, so all subwords in the \mathcal{R} -diagram have a length of at most $\frac{1}{2} + 2\varepsilon$).

Second, notice that the requirement for the existence of a witness is very weak: almost every horizontal segment with a length of λ_3 or less has the canonical form of the map as a witness. The only situation in which a segment is reliable but the canonical form is not a witness for it is when we have a sequence of horizontal segments of almost exactly the required length (i.e., of a length almost exactly equal to λ_{l_i}). Indeed, if the diagram contains a sequence of horizontal segments $S_0, S_1, S_2, ...$ for which the relative length $L_{R_0}(S_0) \in (\lambda_3 - \varepsilon, \lambda_3)$ and for all n > 0, $L_{R_n}(S_n) \in (\lambda_{l_n} - \varepsilon, \lambda_{l_n})$, then we have a witness for S_0 , but we cannot switch to an alternative, shorter path since the length of each horizontal segment satisfies very strict conditions that do not allow us to add a vertical bridge to take an alternative, globally shorter path. However, if such a situation does not occur, we can always "switch" to the canonical form:

Lemma 5. If there exists a cell R in map C for which the relative length of each horizontal segment does not exceed λ_3 , then for each horizontal segment to the left of this cell, the existence of a witness is equivalent to the existence of a part of the witness that reaches R. Similarly, for each segment to the right of R, the existence of a witness is equivalent to the existence of a witness that reaches R.

Proof. Since $\lambda_3 + 2\varepsilon \le \lambda_1$, λ_2 , we can add vertical segments both to the left and to the right, thereby allowing us to switch to the beginning of the shorter horizontal segment in the next cell. Then we take the corresponding part of can(*C*) as a continuation of the witness. \Box

Definition 8. A map in which every segment is reliable is called certified. The result of leaving only the reliable segments in map C is called the certification result (or certified form) of C and is denoted as cert(C).

Now, by introducing separate canonical and certified forms for maps, in the future, we will almost always be interested in their composition: the canonical certified form can(cert(C)).

Lemma 6. $\operatorname{can}(\operatorname{cert}(W_1)) \operatorname{can}(\operatorname{cert}(W_2)) \operatorname{can}(\operatorname{cert}(W_2^{-1}W_1^{-1})) = R_1^{\vee} R_2^{\vee} R_3^{\vee} R_4^{\vee}$, where R_i^{\vee} is the result of conjugating relation R_i with some word.

Proof. This lemma is a direct analogue of the previous statement, but it is not obvious itself since diagram certification is a global process, and therefore, the certified form of the product may not coincide with the certified forms of the factors. However, using Lemma 5, we understand that the differences in certified forms occur only in the segment from the triangle center (see Figure 10) to the point of difference in canonical forms. Indeed, at the point of difference, the length of each horizontal segment does not exceed $\frac{1}{2} + \varepsilon < \lambda_3$, so from Lemma 5, we obtain that all segments farther from the center are either included or not included in the certified forms of both factors and the product simultaneously (since by using a segment with length $< \lambda_3$, we can switch to either $\operatorname{can}(W_1)$ or $\operatorname{can}(W_2^{-1}W_2^{-1})$).

Moreover, we can observe that since all diagram segments are now reliable and thus do not exceed λ_3 in terms of relative length, the situation depicted in Figure 11 is impossible. Therefore, the *only* reason for differences in canonical forms can be differences in certified diagrams. However, again, using Lemma 5, this imposes length conditions on all segments between the triangle center and the point of difference. Recall that the sequence (l_n) was specifically chosen by us to be aperiodic, which gives us the following fact:

Note 2. If in the conditions of Lemma 6 we have $W_1 = A_1^{n_1}$, $W_2 = A_2^{n_2}$, then the distinguishing points in canonical forms are at most $2 \max(|A_1|, |A_2|)$ away from the center.

Proof. Indeed, we impose some aperiodic conditions and, moreover, conditions that do not contain subwords of the form aXaXa, on the lengths of segments of the map between the center of the triangle and the distinguishing point of canonical forms, while the maps of these words are periodic. As a result, the length of the segment on which both conditions hold cannot exceed two periods. \Box

Here we encounter for the first time the specificity of relations in Burnside groups (periodicity). Since this case is of particular interest to us, let us dwell on it a little more and also prove the following lemma:

Lemma 7. Let n > 4 and C be a one-layer map. Then for the map C^n , the certified (and therefore canonical) forms have a periodic structure, except possibly for two boundary periods of C on each side.

Proof. If in the certified form *C* there are no cells (i.e., pairs of alternative paths), then the statement is proven. If at least one cell is present, then the length of each side in it does not exceed λ_3 . In this case, according to Lemma 5, the witness only needs to reach the first such cell. And since this cell appears in every instance of the map *C*, the existence of a witness is equivalent for all internal periods (as it is sufficient to reach the nearest copy of this cell on the right and left). \Box

At this point, we move from constructing the necessary structures for analyzing Burnside groups (and groups with a graded structure of small cancellations) to more applied activities. In the next section, we will consider graded Van Kampen diagrams and apply the apparatus developed in the two previous sections to analyze them.

3.3. Ranking of Relations

Starting from this section, we will finally study the topology of Van Kampen diagrams arising in groups with a graded structure of small cancellations. Here and below, we will usually require $\varepsilon \leq \frac{1}{10}$: that is, each cell is surrounded by at least ten other cells of the same rank.

The foundation for all subsequent reasoning in this section will be the distribution of cells in the diagram according to their distance from a certain fixed central cell.

More formally, let us fix a cell R_0 ; then the distance from each cell R in the diagram to R_0 is defined as the minimum distance between the vertices corresponding to these cells in the dual graph. The left side of Figure 13 shows a typical (topological) appearance of the diagram after distributing the cells by distance.

By classical reasoning in small cancellation theory, it can be proved that if all cells in the diagram, except possibly the central cell, satisfy the small cancellation condition with a constant $\varepsilon < \frac{1}{6}$, then no cell can behave as shown in Figure 13 in the central or right diagram (otherwise, the cells in the closed region would have too few neighbors in the end).

Lemma 8. No cell from the ring at distance n > 1 can touch three or more cells at distance n - 1. Also, each cell at distance n has at least three neighbors at distance n + 1.

Proof. For the first statement: otherwise, one of these cells has one neighbor at distance n, at most two neighbors at distance n - 1, and at most two neighbors at distance n - 2 (we consider a minimal counterexample). For the second statement: otherwise, we have no more than two neighbors at the previous level, no more than two neighbors at the same level, and no more than two neighbors at the next level. Thus there would be a contradiction. \Box

From Lemma 8, it immediately follows that the number of cells at distance n from a given cell grows exponentially as a function of n.



Figure 13. Left: topology of the Van Kampen diagram. Right: examples of impossible topology.

However, we are interested in the topology of diagrams not only in groups with small cancellation conditions but also in groups where the small cancellation condition holds for each rank of relations separately. In this work, we focus on the case when there are two ranks of relations: $\mathcal{R} = \mathcal{R}_1 \sqcup \mathcal{R}_2$. However, generalization to an arbitrary number of ranks can be achieved by trivially repeating all the reasoning presented here. We will use the results of previous sections and bring all the words from \mathcal{R}_2 to canonical form with respect to the relations from \mathcal{R}_1 . Now our goal will be to formulate the necessary estimates for, informally speaking, the number of neighbors for each cell of rank 2 in the Van Kampen diagram over *G*.

Let us formalize the above. So for a Van Kampen diagram consisting of relations of two ranks, we introduce the concept of a *derived diagram* as follows. Let us introduce a certain ordering of rank 2 cells participating in the diagram and denote them as $R_1, R_2, ..., R_k$. Then for each cell of rank 1, distances $d_1, ..., d_n$ from it to each of the cells of rank 2 are defined. We call the *region* of the cell R_i the subdiagram that includes R_i itself as well as all cells for which $\forall j \neq i : d_i < d_j$ and all cells for which $\forall j < i : d_i < d_j$ and $\forall j > i : d_i \leq d_j$. Intuitively, we call the region of the cell R_i the area of the diagram for which R_i is the nearest cell of rank 2, using the introduced ordering to resolve cases of equal distances. The Van Kampen diagram, together with the ordering R_i , the numbers d_i , and the partition into regions, is called the *derived diagram*.

It is clear that since the relations from \mathcal{R}_2 are now written in canonical form, the cells of rank 2 cannot contain, informally speaking, too long parts of the relations from \mathcal{R}_1 . Therefore, the cells of rank 1 cannot be strongly adjacent to a cell of rank 2. More precisely, it will be sufficient for us that the relative length of the common boundary between a cell of rank 1 and rank 2 does not exceed $\lambda_1 < 1 - 4\varepsilon$, which means that each such cell of rank 1 has at least five neighbors of rank 1. As mentioned above, we also assume $\varepsilon < \frac{1}{9}$, which means that each cell of rank 1 that does not share a boundary with a cell of rank 2 has at least nine neighbors.

Taking into account the above and using reasoning that completely repeats the case of diagrams with relations of one rank, we conclude that, informally speaking, in the topology of each specific region there are no holes (see Figure 13). Formally, this means that the subsets of the region formed by the cells of distance $d \le n$ is simply connected for all n.

We are interested in analyzing the boundary between two regions. Before moving on to the analysis of the "big picture", we will prove a lemma that is quite trivial but plays a key role in the subsequent analysis. Note that due to the regularity of the cell layout structure at a given distance from the center of the region, we can fix a certain direction of traversal and have the ability to talk about the next and previous regions in the traversal order at a given distance from the center.

Now consider a cell R of rank 1 at distance n. Let us find a cell R_f at distance 1 obtained as follows: starting from R, at each step, we choose the first in the traversal order cell adjacent to the current cell and located at a distance one less. Similarly, we find a cell R_l obtained in a similar way but choose at each step the last in the traversal order cell among those adjacent to the current cell.

Then the following lemma holds.

Lemma 9. Cells R_f and R_l are adjacent.

Proof. Let us consider the process of iterative descent by distance to the center of the region. At the first step, we obtain two adjacent cells (a consequence of the first statement of Lemma 8). Furthermore, if at the next step we obtain non-adjacent cells, then in the region denoted by blue in Figure 14, a cell at the next level has no more than two neighbors, which again contradicts the second statement of Lemma 8. \Box



Figure 14. Impossibility of divergence of two extreme paths through adjacent cells.

Now let us look at the boundary between two regions. The general view is shown on the left side of Figure 15. Let us first consider the case when the cells of rank 2 for these regions are not close to each other: that is, the boundary between the regions is not a boundary of a rank 2 cell.

Without loss of generality, let us say that the cell of the region on the left has a smaller number. Then with a cell at distance k from the center to the left of the boundary, there can be cells at a distance k or k - 1 to the right of the boundary. The key fact that allows us to provide estimates of the number of neighbors for this region is the following lemma:



Figure 15. (Left): general view of the boundary between two regions. (**Right**): possible distances of cells on the other side of the given cell.

Lemma 10. *If the boundary between regions does not pass through the boundary of a rank 2 cell, then in the sequence of distances to the center during motion, there are no (non-strict) local maxima.*

Proof. Let us consider a local maximum in the left region (the second case is treated analogously). Suppose it is located at a distance of k from the center. It is clear that the neighboring cells in the same region have distances to the center of either k - 1 or k (since k + 1 is impossible due to local maximality). In this case, the given cell has at most four neighbors in its own region. Hence, this cell has at least five neighbors in the right region (see Figure 16).



Figure 16. Local maximum of distance on the boundary.

It is clear that each of these cells has a distance to the center of k or k - 1. However, in any sequence of 5 numbers k and k - 1, there exists a local maximum. Now let us repeat the reasoning for this local maximum, and we obtain that this cell must have at least five neighbors in the left region. However, by construction, it has exactly one. Thus, there is a contradiction. \Box

Thus, when moving along the boundary between regions, the distance to the center first decreases with each step and then increases with each step, with no more than two repeated distances in between (otherwise we would have a local maximum in a trivial way). This means that for the outermost cell on the boundary, one of the two sequences of cells



described in Lemma 9 goes along the boundary (see Figure 17) until the minimum distance on the boundary is reached. This allows us to prove the following important fact:

Figure 17. Paths obtained by choosing the rightmost (respectively, leftmost) cells on the previous layer, starting from the outermost cells on the boundary.

Lemma 11. *If the boundary between regions does not pass through the boundary of a rank 2 cell, then any shortest path from a cell on this boundary to a cell at a distance of 1 from the center of the region enters one of at most five consecutively located cells at a distance of 1 from the center.*

Proof. Let *T* and *B* denote the outermost cells on the boundary. It is evident that two cells, in which the leftmost and rightmost paths coming from *T* respectively enter (denoted by T_f and T_l in the terminology of Lemma 9) are the endpoints of any shortest path to the center starting from any cell on the boundary from *T* to the local minimum. The same holds for *B*. It has been previously shown that there is at most one local minimum, which means that there is at most one cell between the pair of cells T_f , T_l and the pair B_f , B_l , which proves the statement of the lemma.

Moreover, since each cell at a distance of 1 is an endpoint of a shortest path from a cell on the region boundary, Lemma 11 asserts that the number of boundaries that do not pass through rank 2 cells for a given region must be sufficiently large (since each boundary occupies no more than five cells at a distance of 1 from the center).

We still need to consider the case when the boundary between regions passes directly through the boundary of a rank 2 cell. In this case, we are dealing with the so-called adjacency region:

Definition 9. Let us consider two rank 2 cells R_1 and R_2 in the Van Kampen diagram. The portion of the diagram consisting of rank 1 cells for which the distance to each of R_1 and R_2 is one as well as the common segments of the boundaries of R_1 and R_2 is called the adjacency diagram of R_1 and R_2 .

It is precisely in the analysis of adjacent rank 2 cells that our previous efforts pay off. Since the rank 2 relations are now written in canonical form with respect to the rank 1 relations, the following statement is true:

Lemma 12. Let us consider the region of adjacency between cells R_1 and R_2 . Then except for at most two cells on each side of the adjacency region, this region consists entirely of the immediate border between R_1 and R_2 .

Proof. It is clear that the region of adjacency for R_1 and R_2 is precisely a one-layer map. Since each of the relations R_i is written in canonical form, the length of each of the alternative sections does not exceed $\frac{1}{2} + 2\varepsilon$, as we choose the shorter of the alternative sections in the canonical form. According to Lemma 5, in this case, the certified forms match everywhere, except perhaps two sections on each side of the map (the proof is similar to the proof of Lemma 7). Then the canonical forms coincide everywhere, except perhaps three extreme sections on each side (since the canonical form depends only on the geometry of the given cell; two extreme cells may differ in certified form, and one more cell is simultaneously present in both, but the side lengths in it may differ by ε due to differences in the previous ones).

Thus, we obtain the following result:

Theorem 2. If the relations within rank 1 satisfy the small cancellation condition with constant ε_1 , the relations within rank 2 satisfy the small cancellation condition with constant ε_2 , and the length of the common section of rank 1 and rank 2 cells with respect to a rank 2 cell does not exceed ε_3 , then for each Van Kampen diagram over $G = \langle A | \mathcal{R}_1 \sqcup \mathcal{R}_2 \rangle$, each region has at least $(\varepsilon_2 + 10\varepsilon_3)^{-1}$ neighbors.

Proof. From Lemmas 11 and 12, we deduce that if the border between regions does not have an adjacency region, then it corresponds to at most five rank 1 cells at a distance of 1 from the center of the region. If the border has an adjacency region, then it corresponds to at most two rank 1 cells on each side at a distance of 1 from the adjacency region plus at most three rank 1 cells on each side within the adjacency region plus the relative measure of the immediate common border between rank 2 cells, which gives the required estimate (Figure 18). \Box



Figure 18. The border between two rank 2 cells. The possible regions of difference in certified forms are shown in blue; the additional regions of difference in canonical forms are shown in red; the regions where the borders of rank 2 cells necessarily coincide are shown in green.

4. Immediate Results

As mentioned above, this work is an initial step in the project of studying Burnside groups for relatively small exponents based on the apparatus of canonical forms developed by Rips. The first important result in this direction presented in this work is a rigorous written exposition of the theory developed by Rips. Although these explanations are not, strictly speaking, previously unknown scientific achievements, working on this exposition required clarification and formalization of proofs for facts discovered by E. Rips as well as organization of a coherent written presentation of this theory starting from the immediate foundations and ending with an analysis of specific Burnside groups.

Now let us move on to the initial steps taken directly for the purposes of our work.

4.1. Graded Structure of Relations in B(m, n)

As mentioned earlier, relations in Burnside groups can be divided into ranks $\mathcal{R} = \mathcal{R}_1 \sqcup \mathcal{R}_2 \sqcup \mathcal{R}_3 \sqcup \ldots$ in such a way that:

- If $R_1 = sr_1 \in \mathcal{R}_k$, $R_2 = sr_2 \in \mathcal{R}_k$ are two distinct (i.e., not cyclic shifts or inverses of each other) relations of rank k, then $|s| < \varepsilon_k |R_1|$, $|s| < \varepsilon_k |R_2|$ for some $\varepsilon_k < 1$ (the small cancellation condition for each rank).
- If $R_1 = sr_1 \in \mathcal{R}_k$, $R_2 = sr_2 \in \mathcal{R}_l$, k < l are relations of different ranks, then $|s| < \varepsilon'_k |R_2|$ for some $\varepsilon'_k < 1$ (the smallness condition for previous ranks compared to the next one).

Initially, it is not obvious that such a partition is possible; this fact follows from the following lemma:

Lemma 13. Let $w_1 = p_1^n$ and $w_2 = p_2^n$ be relations in a group, where the words p_1, p_2 are not periodic and are not cyclic shifts or inverses of each other. Then the length of any common segment of these relations is less than $|p_1| + |p_2|$.

Proof. In the case of $|p_1| = |p_2|$, the statement is obvious: having a common segment with a length of at least $|p_1|$ implies that $w_1 = w_2$. Without loss of generality, let $|p_1| > |p_2|$. Let s be a common segment of w_1 and w_2 with $|s| \ge |p_1| + |p_2|$. Let $q_1(q_2)$ be the first $|p_1| (|p_2|)$ symbols of s. In this case, w_i becomes periodic with period q_i . Then, if $|p_1| = k|p_2|$, we have $q_1 = q_2^k$, which contradicts the non-periodicity of p_i . Thus, $|p_1| = k|p_2| + r$, $0 < r < |p_2|$. However, this means that the subword $s_{|p_1|+1}s_{|p_1|+2} \dots s_{|p_1|+|p_2|}$ is also equal to q_2 , which implies that w_2 contains two equal periods of q_2 shifted by $0 < r < |p_2|$. This means that the word w_2 is periodic with a period length of $(|p_2|, r)$, which again contradicts the non-periodicity of p_2 . \Box

Therefore, the powers of different non-periodic words intersect at most by the sum of their periods. This allows us to divide the relations into ranks. Let us introduce a sequence of numbers $l_0 = 0, l_1, l_2, l_3, \ldots$ and define

$$\mathcal{R}_{k} = \left\{ w^{n} \middle| l_{k-1} < |w| \le l_{k}, w - \text{non-periodic word} \right\}$$
(3)

Then, under certain conditions on l_k , which we will derive below, this grading will indeed satisfy the required conditions.

Note that E. Rips, like most authors of previous works on this topic, does not thoroughly investigate the relations that l_k must satisfy. This is because when analyzing the complete family of relations \mathcal{R} , the result does not depend on the specific partition into ranks. Therefore, different (correct) partitions are completely equivalent. However, in the context of our research plan on approximations of Burnside groups (which includes only a finite number of relations), a specific and meaningful choice of grading is important to us.

As mentioned above, in this work, we primarily focus on the analysis of the first nontrivial case that arises when considering the first two ranks: $\mathcal{R}_1 \sqcup \mathcal{R}_2$.

4.2. *Refinement of Estimates for the Case* $\mathcal{R} = \mathcal{R}_1 \sqcup \mathcal{R}_2$

Despite the fact that the theory constructed by E. Rips allows for precise estimates of the exponent at which the Burnside group becomes infinite, the exact calculation of these estimates is quite laborious due to boundary effects that arise each time when transitioning from the previous rank to the next (see Figure 19).



Figure 19. Boundary effects (i.e., differences in the canonical form along alternative paths) for one cell in the one-layer map of higher rank.

We observed an example of such boundary effects already for rank 2 in Theorem 2, where the constant in the small cancellation condition for the reduced diagram was a nontrivial combination of the small cancellation constants for the first and second ranks. Although the detailed analysis of the necessary exponent estimates is complex, when taking a sufficiently large exponent, the constants of small cancellations in the reduced diagrams of all ranks are clearly bounded from above, as the sizes of relations grow exponentially with the rank. Therefore, Rips operates with a sufficiently large odd exponent (e.g., an exponent greater than 2^{30}) in his reasoning to simplify proofs during the inductive transition through ranks. Naturally, such restrictions are insurmountable for any (manual or computer) direct analysis. Fortunately, for the case of a finite number of ranks, these restrictions can be significantly relaxed, as the specific form and size of boundary effects in this case can be manually controlled. In this section, we provide refined estimates for the case of the first two ranks. In the terminology of the previous section, let the *n*th powers of non-periodic words with lengths not exceeding *l* be assigned to the first rank, and let the *n*th powers of non-periodic words with lengths from l + 1 to L be assigned to the second rank. We prove the following theorem:

Theorem 3. Let $G = \langle A | \mathcal{R}_1 \sqcup \mathcal{R}_2 \rangle$ in the above terminology, where $|A| \ge 2$. Then, if

$$n > 120, \ n \text{ is odd} l < \frac{n - 24}{24}$$
(4)
$$L < \frac{n(n - 114)}{144}$$

the group G is infinite.

Proof. To prove the theorem, we consistently take into account all the necessary requirements and gradually obtain stronger constraints on *n*.

In constructing the theory in Section 3, we considered an arbitrary constant ε and assumed the fulfillment of the small cancellation condition in rank 1 for it. Note that not every value of ε will work. For the existence of constants τ , λ_0 , λ_1 , λ_2 , λ_3 , it is necessary and sufficient that

$$(1 - \lambda_0) + (\lambda_0 - \lambda_1) + (\lambda_1 - \lambda_2) + (\lambda_2 - \lambda_3) + (\lambda_3 - \frac{1}{2}) < \frac{1}{2}$$

from which, recalling that $\tau > 2\varepsilon$, we obtain

$$(4+2+2+2+2)\varepsilon < \frac{1}{2}$$

$$\forall a, b \leq l : a + b < \min(a, b) n\varepsilon$$

(the condition of small length of the common part for words with periods of lengths a and b). Considering the case a = l, b = l, we obtain

$$n > \varepsilon \frac{2l}{l}$$

which gives n > 48 (note that *n* is odd; see Remark 1). On the other hand, it is clear that the most stringent condition on *l* is obtained when considering the case a = 1, b = l:

$$l < 1 + n\varepsilon$$

Thus, although we can choose a value for *l* that satisfies the conditions for each odd n > 48, as *n* decreases, the number of words we can place in the first rank in such a way that the small cancellation condition $C'(\varepsilon)$ is satisfied rapidly decreases. In the limiting case, when n = 49, only *n*th powers of words of lengths 1, 2, and 3 can be in the first rank.

Now let us consider the constraints imposed on the second rank. Recall that our goal is to apply the analogue of Grindlinger's lemma (Theorem 1) to the derived Van Kampen diagram. For this, it is required that the relative measure of the boundary between two regions in an arbitrary map is less than $\frac{1}{6}$.

Using Lemma 13, we conclude that the length of the common part between a rank 1 cell and a rank 2 cell relative to the rank 2 cell does not exceed

$$\frac{(l+1)+l}{n(l+1)} < \frac{2}{n}$$

Now using the main result of Section 3, Theorem 2, we conclude that in this case, in order for the derivative map to satisfy the condition $C'(\frac{1}{6})$, it is necessary that:

$$\varepsilon_2 + 10 \cdot \frac{2}{n} < \frac{1}{6}$$

where ε_2 is the small cancellation constant within the second rank. From here, we obtain a condition on ε_2 :

$$z_2 < \frac{n-120}{6n}$$

ε

which clearly implies n > 120. Finally, we need to impose a condition on *L* necessary to satisfy the small cancellation condition with the constant ε_2 within rank 2. As in rank 1, this requires

$$\forall a, b \in [l+1, L]: a+b < \min(a, b)n\varepsilon_2$$

where the strictest estimate is achieved when a = l + 1, b = L:

$$\frac{l+l+1}{l+1} < n\varepsilon_2 < n \cdot \frac{n-120}{6n} = \frac{n-120}{6}$$

Now transferring terms and recalling that $l + 1 < n\varepsilon$, we obtain:

$$L < (l+1) \cdot \left(\frac{n-120}{6} - 1\right) < (l+1) \cdot \frac{n-114}{6} < n\varepsilon \cdot \frac{n-114}{6} < \frac{n(n-114)}{144}$$

Thus, the final constraints we obtained are as follows:

$$n > 120, n \text{ odd}$$

$$l < \frac{n - 24}{24}$$

$$L < \frac{n(n - 114)}{144}$$
(5)

Under these conditions, according to Theorem 2, the derivative Van Kampen diagram satisfies the condition $C'(\frac{1}{6})$, and thus, by Greendlinger's lemma (Lemma 1), some cell of some rank forms a boundary segment with a relative length exceeding $\frac{1}{2}$.

However, it is clear that due to the specifics of the relations in the Burnside groups, it is easy to construct an arbitrarily long word W that does not contain subwords of the form S^m , where $m > \frac{n}{2}$ —it is sufficient to take a sequence that does not have long periodic segments. An example of such a sequence is the Morse–Thue sequence (see, for example, [sequences]), which does not even have complete cubes as subwords. Such a sequence cannot be the boundary of a Van Kampen diagram for which it was a boundary, and therefore, it is not equal to the identity in the group *G*.

Thus, under these conditions, the group generated by m > 2 generators and relations $\mathcal{R}_1 \sqcup \mathcal{R}_2$ is infinite. \Box

This theorem shows that considering only the first two ranks of relations makes sense only for sufficiently small (odd) exponents (n < 120).

5. Planned Further Research

This work represents the first step in the investigation of approximations of groups B(m, n) based on the developments of E. Rips. The main result of this paper is the formalization and self-contained exposition of the theory of canonical forms as well as the direct and independent application of this theory to refine some estimates of the approximations of Burnside groups with a finite number of relations. Further research is planned in three directions:

- Generalizing the methods that allowed us to obtain estimates for the two-rank system of relations to an arbitrary finite number of ranks and studying the boundary effects that arise when the number of ranks increases in order to potentially lower the estimates;
- Using computer enumeration to study approximations of B(m, n) for small exponents (e.g., studying the finiteness of B(2, n) for odd n < 120);
- Extending the framework of the iterative small cancellation theory developed by E. Rips to other algebraic structures (rings and fields).

The first direction of research represents the most accessible and straightforward approach to the extension and generalization of this work. It is worth noting that despite the fact that for sufficiently large exponents the inductive transition along the ranks is performed almost automatically, the complexity of obtaining precise estimates and the manual analysis of the process rapidly increases with the growth in the number of ranks under consideration. Nevertheless, it seems possible to obtain estimates that combine a sufficiently high degree of accuracy and allow for an inductive proof to generalize to the case of an arbitrary, finite number of ranks.

At the same time, computer analysis (the second potential direction) may lead to improvement in our understanding of the structure of B(m, n) for small exponents. Finally, the third approach is arguably the most ambitious one, as obtaining an analogue of the small cancellation theory for other algebraic objects is a potentially important and largely open problem in modern mathematics.

Author Contributions: Conceptualization, A.B. and I.I.-P.; methodology, A.B.; formal analysis, A.B.; writing—original draft preparation, A.B.; writing—review and editing, I.I.-P.; visualization, A.B.; project administration, I.I.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Russian Science Foundation (grant No. 22-11-00177).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- 1. Burnside, W. On an unsettled question in the theory of discontinuous groups. *Quart. J.* **1902**, *33*, 230–238.
- 2. Chandler, B.; Magnus, W. *The History of Combinatorial Group Theory: A Case Study in the History of Ideas*, 1st ed.; Studies in the History of Mathematics and Physical Sciences; Springer: Berlin/Heidelberg, Germany, 1982.
- 3. Kostrikin, A.K. *Around Burnside*, 1st ed.; Ergebnisse der Mathematik und ihrer Grenzgebiete 20; Springer: Berlin/Heidelberg, Germany, 1990.
- 4. Kostrikin, A.I. The Burnside Problem. Transl. Ser. 2 Am. Math. Soc. 1964, 36, 63–99. [CrossRef]
- 5. Burnside, W. On criteria for the finiteness of the order of a group of linear substitutions. *Proc. Lond. Math. Soc.* **1905**, 2–3, 435–440. [CrossRef]
- 6. Schur, I.; Brauer, A.; Rohrbach, H. *Über Gruppen periodischer linearer Substitutionen*; Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin: Berlin, Germany, 1911; pp. 619–627.
- 7. Sanov, I.N. *Solution of Burnside's Problem for Exponent 4*; Uchenyye Zapiski Leningradskogo Gosudarstvennogo Universiteta: Leningrad, Russia, 1940; pp. 166–170
- 8. Hall, M.J. Solution of the Burnside problem for exponent six. Ill. J. Math. 1958, 2, 764–786. [CrossRef]
- 9. Golod, E.S. On nil-algebras and finitely approximable p-groups. Transl. Ser. 2 Am. Math. Soc. 1965, 48, 103–106. [CrossRef]
- 10. Aleshin, S.V. Finite automata and Burnside's problem for periodic groups. Math. Notes 1972, 11, 199–203. [CrossRef]
- 11. Grigorchuk, R.I. On Burnside's problem on periodic groups. Funct. Anal. Appl. 1980, 14, 41-43. [CrossRef]
- 12. Sushchanskij, V.I. Periodic *p*-groups of permutations and the unrestricted Burnside problem. Sov. Math. Dokl. **1979**, 20, 766–770.
- 13. Novikov, P.S.; Adyan, S.I. On infinite periodic groups I, II, III. Math. USSR Izv. 1969, 2, 209–236, 241–480, 665–685. [CrossRef]
- 14. Chandler, B.; Magnus, W. *Razvitie Kombinatornoj Teorii Grupp. Ocherk Istorii Razvitiya Idej*; transl. from the English ed.; Mir: Moskva, Russia, 1985.
- 15. Adian, S.I. The Burnside Problem and Identities in Groups; Springer: Berlin/Heidelberg, Germany, 2011.
- 16. Atkarskaya, A. Combinatorial Approach for Burnside Groups of Relatively Small Odd Exponents. 2021. Available online: https://www.youtube.com/watch?v=lJbdrPASjow/ (accessed on 1 December 2023).
- 17. Atkarskaya, A.; Rips, E.; Tent, K. The Burnside problem for odd exponents. arXiv 2023, arXiv:2303.15997.
- 18. Ol'shanskiI, A.Y. *Geometry of Defining Relations in Groups*, 1st ed.; Mathematics and Its Applications 70; Springer: Dordrecht, The Netherlands, 1991.
- 19. Rips, E. Generalized small cancellation theory and applications I. The word problem. Israel J. Math. 1982, 41, 1–146. [CrossRef]
- 20. Sela, E.R.Z. Canonical representatives and equations in hyperbolic groups. Invent. Math. 1995, 120, 489–512. [CrossRef]
- 21. Atkarskaya, A.; Belov-Kanel, A.; Eugene, P.; Rips, E. Structure of small cancellation rings. *J. Res. Math. Educ.* **2021**, *2*, 1–14. [CrossRef]
- 22. Gromov, M. Hyperbolic Groups. In *Essays in Group Theory*; Gersten, S.M., Ed.; Springer: New York, NY, USA, 1987; pp. 75–263. [CrossRef]
- 23. Adyan, S.I. The Burnside problem and related topics. Russ. Math. Surv. 2010, 65, 805–855. [CrossRef]
- 24. Roger C. Lyndon, P.E.S. Combinatorial Group Theory; Classics in mathematics; Springer: Berlin/Heidelberg, Germany, 2001.
- 25. Kampen, E.R.V. On Some Lemmas in the Theory of Groups. Am. J. Math. 1933, 55, 268–273. [CrossRef]
- 26. Klyachko, A.A. Combinatorial Group Theory and Geometry. Math. Prosv. S.3 2009, 13, 18-32.
- 27. Greendlinger, M. Dehn's algorithm for the word problem. Commun. Pure Appl. Math. 1960, 13, 67-83. [CrossRef]
- 28. Lyndon, R.C. On Dehn's algorithm. Math. Ann. 1966, 166, 208-228. [CrossRef]
- 29. Schupp, P.E. On Dehn's algorithm and the conjugacy problem. Math. Ann. 1968, 178, 119–130. [CrossRef]
- 30. Rips, E. Iterated Small Cancellation Theory for Groups and Rings. Available online: https://www.youtube.com/playlist?list= PLEguGWgkjHqPbv_yYULRfSZXls2dIX3NU (accessed on 1 December 2023).
- Allouche, J.P.; Shallit, J. The Ubiquitous Prouhet-Thue-Morse Sequence. In Sequences and Their Applications; Springer: London, UK, 1999; pp. 1–16.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.