Dehn function and length of proofs

Jan Krajíček*

Mathematical Institute†
Academy of Sciences, Prague

Abstract

We link Dehn function to length-of-proofs function in propositional proof complexity.

Introduction

Let $G = \langle A, R \rangle$ be a finite presentation of a group, with finite sets of generators A and relators R, words from $(A \cup A^{-1})^*$, with $A^{-1} = \{a^{-1} \mid a \in A\}$. Let w be a word in the free group $F(A)$ generated by A that equals to 1 in G, that is, it belongs to the word problem set $W(G) := \{w \in F(A) \mid w = 1 \text{ in } G\}$. One may prove that $w \in W(G)$ by consecutively inserting or deleting from w conjugates of relators from R until reaching 1, i.e. the empty word. This is clearly reminiscent to formal logical systems where one derives formulas in consecutive steps applying one of finitely many rules. We shall consider as an example of a logical system a variant of usual text-book calculus for propositional logic called Extended Frege system EF. The proof system works with boolean circuits (instead of usual formulas - that is the meaning of the qualification "extended") and is based on finitely many axiom schemes and inference rules, cf. [10].

Counting the number of steps in a derivation of $w = 1$ corresponds to the number of steps in proofs. Moreover, knowing a bound to the number of steps in the derivation one can bound also the size of intermediate words

*Partially supported by grant # A 101 99 01 of the Academy of Sciences of the Czech Republic and by project LN00A056 of The Ministry of Education of the Czech Republic.

†Also member of the Institute for Theoretical Computer Science of the Charles University. A part of this work was done while visiting the Mathematical Institute, Oxford.

(perhaps in another derivation with at most linearly more number of steps) and hence also the size, i.e. the total number of symbols, in some derivation of $w = 1$. This holds analogously (and with analogous bounds) for propositional calculus EF. In fact, this property is why one allows the calculus to operate with circuits rather than with formulas. (The minimal number of steps in proofs in EF and in usual calculus operating with formulas, a Frege system F, are proportional to each other so one could take F and the number of steps instead of EF and the size.)

Further, based on (the proof of) the undecidability of the word problem one can have finitely presented groups $G = \langle A, R \rangle$ and polynomial time set $X \subseteq A^*$ such that $W(G) \cap X$ is coNP-complete. For this group, no non-deterministic polynomial time algorithm can accept $X \cap W(G)$ and hence neither $W(G)$ (as X is polynomial time), unless the computational complexity conjecture that the class NP is not closed under complementation fails. coNP-complete sets are interesting because the set of propositional tautologies is coNP-complete. And the same remark applies to propositional logic; unless $NP = coNP$, no non-deterministic polynomial time algorithm accepts exactly the set of tautologies.

These similarities are clear but rather on surface. What sparked my interest in a possibility of a deeper relation of geometric group theory to propositional proof complexity is the following result by Birget et. al.[3]. Obviously, the simple non-deterministic procedure for verifying that $w = 1$ in G mentioned at the beginning is an example of a non-deterministic algorithm. There is a (possibly another) non-deterministic algorithm whose time is polynomially related to the number of steps taken by the simple algorithm (by the remark that one can also bound the size, cf. Section 3). What Birget et.al.[3] proved is that this simple algorithm is, in a sense, optimal. Namely, if $W(G)$ can be accepted by a non-deterministic algorithm running in time $T(n) \geq n$, then G can be embedded into some other finitely presented group H such that the simple algorithm on H works in the number of steps that is polynomial in $T(n)$.

Interestingly, this is also analogous to propositional logic. Namely, any non-deterministic algorithm for the set of tautologies can be simulated by EF that can use possibly an extra polynomial time set of axioms. Simulated means that there is a polynomial time algorithm, transforming computations of the original algorithm into proofs in the augmented EF, and hence the number of proof steps is polynomial in the time of the original algorithm.

This leads to the following informal correspondence that we shall make more precise later on in the paper. A language corresponds to (a simple subset of) a word problem in a finitely generated group G, a proof system for the language (see Section 1) corresponds to a finitely presented group H into which G embeds. The bigger set of relators of H corresponds to extra axioms added to EF and the fact that G embeds in H rather than only homomorphically maps into it corresponds to the conservativity of the extension. Proof complexity (of these extensions of EF) corresponds to complexity of Dehn functions. Our motivation for studying this connections is that Dehn function and complexity of word problem in general are linked to various geometric constructions and situations (we mention few in Section 3) which is something we lack for propositional proof system (with a small exception of cutting plane systems and its generalizations to discretely ordered modules which are, however, much weaker than EF). Propositional proof complexity relies mostly on finite combinatorics (and connections to bounded arithmetic and computational complexity), and it would be - in my view - useful to enlarge the set of methods we have at our disposal. The few cases when this was achieved (e.g. using algebraic or model theoretic methods) gave not only lower bounds but also results of a structural type. Such results are known for Dehn function too, see Section 3.

The paper is organized as follows. In Section 1 we recall some basic definitions and facts from proof complexity. Sections 2 and 3 similarly recall basics of finitely presented groups, the word problem and Dehn function. These sections contain also material that is well-known to researchers in the respective fields; we spell it out explicitly in order to make the whole presentation readily accessible to both logicians/complexity theorists and group theorists. In Section 4 we show a way how to embed propositional logic into word problem, using Birget's reduction [2], and we propose the notion of finitely presented groups embedding a fixed finitely generated group G_{TAUT} as a natural analogy to propositional proof systems. In Section 5 we describe a precise connection between propositional proof systems and these group-based proof systems. We conclude with a problem about complexity of proving primality.

More background information can be found in [6, 19] and in [15].

1 Proof systems

Finite objects of a given type (numbers, formulas, graphs, polynomials, etc.) are considered as encoded by words over a fixed finite alphabet A with at least two letters. Let A^* be the set of words over A. The following definition generalizes a notion that was defined for propositional logic (i.e., L is the set of propositional tautologies) by [10].

Definition 1.1 *A proof system for a language $L \subseteq A^*$ is a binary relation $P \subseteq A^* \times A^*$ (a word x such that $P(x, y)$ is called a P-proof of $y \in L$) satisfying the following three conditions:*

1. *Soundness: If $y \in A^*$ has a P-proof then $y \in L$.*

2. *Completeness: Any true $y \in L$ has a P-proof.*

3. *Effectivity: There is a polynomial-time algorithm deciding on inputs x, y whether the relation $P(x, y)$ holds.*

Various variants of this notion, with probabilistic notions in particular, are also very important in computational complexity theory.

For a language L whose membership is itself decidable effectively (e.g., to be an even number), or at least defined by an existential condition (e.g., to be a composite number), it is trivial to construct a proof system (a proof of $y \in L$ just incorporates witnesses to the existential quantifiers). Non-trivial proof systems occur when L is defined by a universal condition (e.g., to be a prime number) or by a combination of existential and universal conditions. Let us consider several examples of such proof systems.

Primality proofs. A proof system for the set of primes is constructed as follows. A proof of the primality of p will consist of a primitive root g modulo p, together with an extra information. This extra information will be a proof of the statement that the order of g modulo p is $p-1$; this statement is universal (no exponent smaller than $p-1$ is the order). Instead of checking for all $x < p-1$ that $g^x \not\equiv 1 \pmod{p}$ we check this for all x of the form $\frac{p-1}{q_i}$, for all prime divisors q_i of $p-1$. To certify that q_i are all prime divisors of $p-1$ we include into the extra information a decomposition of $p-1$ into powers of q_i, $p-1 = \prod_i q_i^{e_i}$, and proofs of the primality of all q_i. This process is iterated (obviously only finitely many times as all $q_i < p$), obtaining the full proof of the primality of p. The size of the full proof is only polynomially longer than the number of digits in p, see [20].

Propositional Calculus. L is the set of propositional tautologies TAUT in the DeMorgan propositional language. Any usual propositional calculus forms a proof system for L. The effectivity condition follows from the schematic way of defining the axioms and the inference rules of the calculi.

Non-3-colorability. L is the set of undirected graphs that are not 3-colorable. Hajós [13] devised three simple rules of forming a new graph from one or two graphs given, and he proved that a graph is not 3-colorable iff it can be obtained from the complete graph on 4 vertices by a repeated use of the rules.

Nullstellensatz. L is the set of finite families of polynomial equations $f_i = 0$ over a finite (to assure a simple encoding in A^*) field that have no solution in the algebraic closure of the field. A proof system for L is given by Hilbert's Nullstellensatz: a proof in the Nullstellensatz proof system NS is a list of polynomials g_i over the same field such that the identity

$$\sum_i g_i \cdot f_i = 1$$

holds in the ring of polynomials over the field. To be able to check the identity efficiently represent the polynomials as sums of monomials (the so called dense notation) rather than by general terms in the language of rings.

A crucial question to be asked about a proof system is how long are its proofs, i.e., what is the minimum length of a proof of an instance $y \in L$. For example, it is a major open problem in complexity theory and mathematical logic whether the set of propositional tautologies admits a proof system in which every tautology has a proof that is at most polynomially longer (it is also not ruled out, in fact, that the standard text-book propositional calculus has this property). This is an equivalent formulation of the $NP =_? coNP$ problem that asks whether a complement of a language with polynomially long proofs also admits such a proof system. The famous $P =_? NP$ problem of Cook [8] asks whether every language admitting a proof system with polynomially long proofs is, in fact, itself decidable in polynomial time (informally: Is it easy to find a proof if we know it exists?). The link between proof systems and machines stems from thinking of proofs as of computations: any particular accepting computation of a deterministic or a non-deterministic Turing machine on an input y is a proof that the machine accepts y. Every proof system can be reduced to this form.

In the four examples described earlier only the proofs of primality are known to be polynomially long (but it is unknown whether the primality of a number can be decided in polynomial time). All the other problems are examples of coNP-complete problems: they admit polynomially long proofs in some proof system iff $NP = coNP$, and they are decidable in polynomial time iff $P = NP$, cf.[11]. In particular, as $NP \neq coNP$ is the conjectured statement, it is thus expected that no proof system for any of those problems admit polynomially bounded proof system. For (fragments) of the particular proof systems considered earlier super-polynomial lower bounds are known (the reader can find some information in surveys [21, 23]).

The research into complexity of propositional logic is developed most among all types of proof systems. This is due to links to bounded arithmetic theories: lower bounds for proof systems yield independence results for corresponding theories of bounded arithmetic, and algorithms definable in the theories yield constructions of short proofs, cf. [15].

The most important propositional proof system is a Frege system F. A Frege system is any calculus based on finitely many axiom schemes and inference rules (that are sound and implicationally complete) in some complete propositional language. An example of such a system is the usual text-book calculus based on few axioms schemes and modus ponens. The notion of a Frege system is quite robust: their strength does not depend on the language, on the particular axiomatisation or the particular list of rules, or on formats of proofs (tree-like or sequence-like), cf. [10, 15]. This means, that proofs in any such system can be transformed into proofs in any other Frege system with at most polynomial increase of the length (and, in particular, super-polynomial lower bounds for one of them imply such bounds for all of them).

No non-trivial lower bounds are known for F so far. However, for some rather strong subsystems of F lower bounds are established (cf. [21, 23]).

Extended Frege system, discussed in the introduction is obtained from F by allowing to introduce and use in proofs abbreviations for formulas. Formally, in a proof one can introduce axiom of the form $r \equiv \phi$, where r is new atom not occurring in ϕ or the proven formula, or earlier in the proof, and then to use it in later parts of the proof. EF can be equivalently formulated as a Frege system operating with circuits.

Definition 1.2 *Let $P(x,y)$ and $R(x,y)$ be two proof systems for L. Then P polynomially simulates R iff there exists a polynomial time function $f(x,y)$ such that for all $x, y \in A^*$:*

$$R(x,y) \to P(f(x,y),y)$$

That is, there is a polynomial time algorithm transforming R-proofs into P-proofs.

Theorem 1.3 (Cook [9]) *Let P be any proof system for the set $TAUT$ of propositional tautologies (in de Morgan language). Then there is a polynomial time set Con_P of tautologies such that EF augmented by Con_P as extra axioms polynomially simulates P.*

In fact, more holds. First, Con_P is not only polynomial time, but also log-space and sparse. Second, if P is sufficiently strong (e.g. it simulates EF) then P proves all tautologies in Con_P in polynomial size and hence P is equivalent (up to polynomial simulation) to $EF + Con_P$.

Definition 1.4 *Let $X \subseteq TAUT$, and denote $EF + X$ by P. The length-of-proof function for P is function $k_P : \mathbf{N} \rightarrow \mathbf{N}$ defined by*

$$k_P(n) := \min_{\tau \in TAUT, |\tau| \leq n} ||\tau||_P$$

where $||\tau||_P$ is the minimal number of steps in a P-proof of τ.

The following theorem follows from Theorem 1.3 and the fact that the minimal size of a EF-proof of a formula of size n is bounded by $(O(k_{EF}(n))$, cf. [10] or [15].

Theorem 1.5 *The statement that $NP \neq coNP$ is equivalent to the statement that for no polynomial time subset $X \subseteq TAUT$ is $k_{EF+X}(n)$ bounded by a polynomial.*

2 Finitely presented groups

Given finite presentation $\langle A, R\rangle$ of a group G, the word problem of G is the problem to decide membership in the set $W(G) := \{w \in F(A) \mid w = 1 \text{ in } G\}$. Strictly speaking, $W(G)$ depends on the particular presentation and not just on G. However, given two presentations of the same group, the two word problems are reducible to each other by a polynomial time algorithm. Note that G is $F(A)/\langle\langle R\rangle\rangle$, $\langle\langle R\rangle\rangle$ the normal subgroup of $F(A)$ generated by R, and $W(G) = \langle\langle R\rangle\rangle$.

Many interesting groups, notably various groups occurring in geometry or topology (like some fundamental groups or braid groups), are finitely presented. Motivated by geometrical problems about low dimensional manifolds M.Dehn formulated in 1912 the word problem for groups. P.S.Novikov [18] and W.Boone [4, 5] constructed a finitely presented group whose word problem is undecidable.

The most fundamental result in the area is Higman's embedding theorem, [14]. A finitely generated group is recursively presented iff it can be presented as $\langle A, R\rangle$ with A finite and R recursively enumerable set.

Theorem 2.1 (Higman [14]) *Finitely generated group is recursively presented iff it can be embedded in a finitely presented group.*

The theorem does imply also the Novikov-Boone undecidability result. Take G generated by a, b, c, d with relators $a^{-i}ba^i = c^{-i}dc^i$, for i from some recursively enumerable but non-recursive set S. Then $a^{-i}ba^i = c^{-i}dc^i$ holds in G iff $i \in S$, cf. [17, Thm.7.2]. Let H be a finitely presented group embedding G, guaranteed to exists by Higman's theorem. Then $W(H)$ is undecidable as S reduces, polynomially even, to $W(H)$.

Note that the theorem is also analogous to a theorem in logic: A theory is recursively enumerable iff it has a finitely axiomatisable conservative extension. Theorem of Birget et.al. [3] mentioned in the introduction is, in a sense, a feasible version of Higman's embedding theorem, with the condition that R is recursively enumerable replaced by the condition that it is in the class NP.

3 Dehn function

Let $G = \langle A, R\rangle$. We make the following technical assumptions on R:

1. R does not contain any word aa^{-1} or $a^{-1}a$, for $a \in A$.
2. R is closed under cyclic permutations of words.
3. R is closed under inverses.

Clearly, any finite R can be replaced by a finite R' obeying the conditions and defining the same group. A reduction step allows to replace a word of the form w_1w_2 by $w_1uru^{-1}w_2$ and vice versa, where $r \in R$ and u is arbitrary word.

Lemma 3.1 *Let $w \in F(A)$. Then w can be reduced in N steps to* 1 *iff w can be written in $F(A)$ in the form*

$$w = \Pi_{i=1}^{N} u_i r u_i^{-1}$$

with $r_i \in R$ and u_i arbitrary.

Proof :

The if-part is trivial. For the only if-part let $x = w_1uru^{-1}w_2$ and $y = w_1w_2$. Then $x = yw_2^{-1}ur(w_2^{-1}u)^{-1}$ and $y = xw_2^{-1}ur^{-1}(w_2^{-1}u)^{-1}$. So every reduction step can be performed by multiplying on the right by a word of the form uru^{-1}. Simple induction concludes the proof.

q.e.d.

Denote by $a(w)$ the minimal N such that w can b expressed as in Lemma 3.1.

Definition 3.2 *The Dehn function $D(n)$ of the presentation $\langle A, R\rangle$ for group G is*

$$\max_w a(w)$$

where w ranges over words from $F(A)$ of length n that are equal to 1 in G.

Assume $\langle A, R\rangle$ and $\langle A', R'\rangle$ are two presentation of the same group G, with Dehn function $D(n)$ and $D'(n)$. Then D and D' are equivalent in the equivalence relation $f \simeq g$ given by:

- There is constant $c \geq 1$ such that for all n:

$$f(n) \leq c \cdot g(c \cdot n) + c \cdot n + c \text{ and } \; g(n) \leq c \cdot f(c \cdot n) + c \cdot n + c$$

see [6, 19]. Hence we speak of the Dehn function of G, meaning the $\simeq$-class of the Dehn functions of finite presentations of G.

The Dehn function has a beautiful geometric interpretation. A *van Kampen* diagram Δ is a planar, connected, simply-connected, labelled 2-dimensional complex. Its edges are directed and labelled by generators from A. The labelling of edges determines labelling of faces. It is required that the word of labels of its border edges when read from any point is a relator from R.

Diagram Δ has a fixed base point on its boundary. If reading the word on the boundary cycle starting from the base point and going counter-clockwise gives word w, we say that Δ is van Kampen diagram for w. The following is fundamental van Kampen lemma, called also the fundamental lemma of combinatorial group theory (cf. [22, Chpt.12]). The area of a diagram is the number of its faces.

Lemma 3.3 (van Kampen) *Let $w \in W(G)$ be arbitrary. Then $a(w)$ is the minimal area of a van Kampen diagram for w.*

This is not difficult to prove by induction similar to the proof of Lemma 3.1. The lemma explains why Dehn function is also called an isoperimetric function.

The geometric context is deep, derived from Plateu's problem. An isoperimetric function of a (2-dimensional, genus 0, smooth and closed) Riemannian manifold bounds the area of a disc in terms of the length of its boundary (a null-homotopic curve). It is a striking theorem that the Dehn function of the fundamental group of a manifold is $\simeq$-equivalent to the smallest isoperimetric function for the manifold, cf. [6].

Another measure associated with a van Kampen diagram Δ is its diameter, more precisely the largest distance (in the combinatorial sense of graph theory) from a base point to a vertex. Let $\Phi_G(n)$ be the minimal number bounding the diameter of some van Kampen diagrams for all words from $W(G)$ of length at most n. Then, in fact, $\Phi_G(n)$ is optimal upper bound on the length of conjugating elements u_i in the expression in Lemma 3.1. In general, $D_G(n) \leq \exp(\exp(\Phi_G(n)))$, although a single exponential bound is conjectured, cf. [7].

We shall need an upper bound in the opposite direction which is much simpler. In general, the number of steps taken by the simple algorithm that consecutively reduces w to 1 does not need to bound also the time of this algorithm. This is because the intermediate words may be long and the algorithm will consume time in just writing them down to a memory. However, this can be remedied, as follows, for example, from the following lemma.

Lemma 3.4 ([19, L.1]) *Let Δ be a van Kampen diagram for w with e edges. Then w is equal in the free group $F(A)$ to a word of the form*

$$u_1 r_1 u_2 r_2 \dots r_k u_{k+1}$$

where all $r_i \in R$, $u_1 u_2 \dots u_{k+1} = 1$ in $F(A)$, and $\sum_{i=1}^{k+1} |u_i| \leq 4e$.

As the number of edges in a van Kampen diagram is bounded by a constant (the maximal length of $r \in R$) times its area, we have that $\sum_{i=1}^{k+1} |u_i| \leq O(D_G(n))$, where n is the length of w. Hence a non-deterministic algorithm that first guesses the word $u_1 u_2 \dots u_{k+1}$ and verifies that it equals 1 in $F(A)$, and then inserts relators r_i and verifies that $u_1 r_1 u_2 r_2 \dots r_k u_{k+1}$ equals to w

in $F(A)$ runs in time $O(D_G^2(n))$. The word problem a group with polynomial Dehn function is therefore in the class NP. Next theorem gives a form of converse.

An embedding of a finitely generated group G in another finitely generated group H has bounded distortion iff the word length of any element of G in G is within a constant factor of the word length of the element in H.

Theorem 3.5 (Birget et.al. [3]) *Let G be a finitely generated group with $W(G)$ accepted by a non-deterministic Turing machine in time $T(n) \geq n$ such that $T^4(n)$ is superadditive (i.e. $T^4(m+n) \geq T^4(m) + T^4(n)$). Then G can be embedded with a bounded distortion into a finitely presented group H whose Dehn function is $\simeq$-equivalent to $n^2 T^4(n^2)$.*

In particular, $W(G)$ is in NP iff G can be embedded with bounded distortion into H with polynomially bounded Dehn function.

(For earlier work leading up to this result see [3]; In particular, Birget [1] proved analogous result for semigroups.)

One has to allow the embedding of G in a bigger group as there are finitely presented groups with word problem decidable in polynomial time but huge Dehn function. An example is Baumslag-Solitar group $G_{2,1} := \langle a, t |\ tat^{-1} = a^2 \rangle$. It embeds into $GL(2, \mathbf{Q})$ and so the word problem is solvable in quadratic time, but its Dehn function is exponential. In fact, Sapir constructed finitely presented groups with word problem in NP but whose Dehn functions are not even primitive recursive, cf.[19].

The Dehn function thus measures, up to embeddability, the complexity of the word problem (it has been known before that the word problem is decidable iff the Dehn function is recursive). This is similar to the length-of-proofs function. But Dehn function has further geometric significance besides Van Kampen diagrams, and it would be quite interesting if some similar context existed for $k_P(n)$. Let me give a couple more examples of properties associated to Dehn function, more can be found in [6, 19].

Gromov [12] proved that if $D_G(n)$ is $o(n^2)$ then it is, in fact, $O(n)$ and this happens iff the group is word-hyperbolic. The hyperbolic groups admit particularly simple algorithm for word problem, so called Dehn presentation. Namely, there is a list of words $u_1, v_1, \ldots, u_k, v_k$ such that $u_i = v_i$ in G and $|u_i| > |v_i|$, and whenever a non-trivial word w equals to 1 in G then one of u_i's occurs as a subword in w. So the algorithm replaces u_i by v_i and recurs. G can be presented with relators $u_i v_i^{-1}$.

Gromov [12] also found a way to abstract geometric patterns in the Caley graph of G that recur on larger and larger scale and he proved that some topological properties of these give information about the complexity of the word problem. Specifically, the asymptotic cone $Cone_\omega(G)$ is a large complete metric space (constructed by a form of ultraproduct from the Caley graph). He proved that if the asymptotic cone is simply connected then the Dehn function $D_G(n)$ is bounded by a polynomial. This is not a characterization, although a converse holds in a special case, cf. [6].

Some methods (combinatorial, geometric, algebraic) for proving lower bounds for Dehn functions are discussed in [6, Chpt.7].

4 Embedding propositional logic into the word problem

There are many "natural" finitely presented groups with the word problem in NP. For example, finitely presented matrix groups over a field, or automatic or hyperbolic groups, cf. [19]. The constructions behind Novikov-Boone's undecidability result or behind Higman's theorem give finitely presented groups of much bigger complexity but not, apparently, or arbitrary complexity with respect to polynomial time reducibility. A key step in the construction is, in one form or another, Boone's lemma (cf. [22, Chpt.12]) that does the following. For a non-deterministic Turing machine T accepting $X \subseteq \Omega^*$, Ω a finite alphabet, produces a finitely presented group $\mathcal{B}(T) = \langle A, R \rangle$, with $A \supseteq \Omega$, and embedding $h : \Omega^* \to A^*$ such that $h(w) = 1$ in $\mathcal{B}(T)$ iff $w \in X$, for $w \in \Omega^*$. So the fine complexity of $W(\mathcal{B}(T))$ may differ from that of X or $h''X$.

An embedding of propositional logic into the word problem would be easy and elegant if we had a finitely generated group whose word problem is coNP-complete. Unfortunately, no such group is known at present. However, we can do without such a group. The point is that in propositional proof complexity we are, a priori, interested in lengths of proofs, that is in the non-deterministic time complexity of TAUT. Thus we only need to embed TAUT into the word problem of a finitely generated group G such that the non-deterministic time complexity of $W(G)$ is the same as that of TAUT. This can be done; the reduction in [3] has likely such a property but, in fact, another suitable reduction has been defined by Birget [2].

Definition 4.1 (Birget[2]) *Let $L \subseteq \Omega^+$ be a language without the empty word. Let $\Omega_1, \ldots, \Omega_{2m}$ be $2m$ disjoint copies of Ω, each with some fixed bijection to Ω, for $m \geq 16$ a fixed integer. For $w \in \Omega^+$ let $(w)_i$ be the translation* *of w into Ω_i obtained by replacing letters from Ω by the corresponding letters from Ω_i.*

Define map $h : \Omega^+ \to \bigcup_{i \leq 2m} \Omega_i$ by:

$$h(w) := (w)_1(w)_2 \dots (w)_m((w)_{m+1}(w)_{m+2} \dots (w)_{2m})^{-1}$$

and group G_L by:

$$G_L := \langle \bigcup_{i \leq 2m} \Omega_i \ ; \ h''L \rangle$$

The mapping h is injective and linear time computable. The next theorem says that it is, in fact, the wanted reduction.

Theorem 4.2 (Birget[2, Cor.3.7]) *The mapping h is a linear time reduction of L to $W(G_L)$. Moreover, if L is accepted by a non-deterministic Turing machine in time $O(T(n))$, so is $W(G_L)$.*

This is proved by a small cancellation theory argument and the algorithm for $W(G_L)$ obtained is just the Dehn presentation (see Section 3).

Definition 4.3 *A group-based proof system for $L \subseteq \Omega^+$ is a finitely presented group $H = \langle A, R \rangle$ such that $\bigcup_{i \leq 2m} \Omega_i \subseteq A$ and the inclusion is an embedding of G_L into H.*

An H-proof of $w \in L$ is a sequence $\langle u_1, r_1, \dots, u_k, r_k, u_{k+1} \rangle$ of words from A^* such that

1. *$u_1 \dots u_{k+1} = 1$ in $F(A)$.*
2. *$u_1 r_1 \dots r_k u_{k+1} = h(w)$ in $F(A)$.*
3. *All $r_i \in R$.*

Lemma 4.4 *Let H be a group-based proof system for $L \subseteq \Omega$. The H is a proof system for L in the sense of Definition 1.1 too.*

Proof :

Theorem 4.2 implies the soundness and completeness conditions of Definition 1.1. The effectivity condition follows from the fact that $W(F(A))$ is polynomial time decidable.

q.e.d.

5 A simulation between propositional proof systems and group based proof systems

Let me first recall a translation of coNP-properties into tautologies (see e.g. [15]). Let $A(x)$ be a coNP-property of a set of words. Assume $A(x)$ has the form $\forall y; |y| \leq |x|^k \to B(x,y)$ with $B(x,y)$ a polynomial time predicate. Fixing length n to bound the length of words x we define a propositional formula $||A(x)||^n$ as in the proof of the NP-completeness of satisfiability (see e.g. [11]). The formula has n atoms $p_1, \ldots, p_n$ for bits of an x, $m = n^k$ atoms $q_1, \ldots, q_m$ for bits of a potential y, and atoms $r_1, \ldots, r_s$ for $s = n^{O(1)}$ bits of values on nodes of some fixed circuit C_n computing from $\overline{p}$, $\overline{q}$ the truth value of predicate $B(x,y)$. Formula $||A(x)||^n$ says, in a DNF form, that if $\overline{r}$ are computed by circuit C_n from inputs $\overline{p}$, $\overline{q}$ then the output of the computation is 1. Given a b of length at most n with bits $b(1), \ldots, b(n)$ let $||A(x)||^n(b)$ be the propositional formula with $b(i)$ substituted for p_i (the remaining atoms $\overline{q}$ and $\overline{r}$ remain unsubstituted). The point is that b satisfies $A(x)$ iff $||A(x)||^n(b)$ is a tautology.

We shall consider propositional proof systems now, i.e. proof systems for TAUT.

Theorem 5.1 *1. For a propositional proof system P there is a group-based propositional proof system H that polynomially simulates P. Moreover, if P has the form $EF + X$, for $X \subseteq TAUT$ a polynomial time set, then $k_P(n)$ and $D_H(n)$ are polynomially related.*

2. For any group-based propositional proof system H there is a polynomial time subset $X \subseteq TAUT$ such that $P := EF+X$ polynomially simulates H. Moreover, $k_P(n) = D_H(n)^{O(1)}$.

Proof :

Let P be a propositional proof system. Denote by $s_P(n)$ the minimal ℓ such that every tautology of size $\leq n$ has a P-proof of size $\leq \ell$. We can view P as a non-deterministic algorithm accepting $TAUT$ in time $s_P(n)^{O(1)}$ (by the effectivity condition of Definition 1.1). Hence $W(G_{TAUT})$ is accepted by a non-deterministic Turing machine T in time $s_P(n)^{O(1)}$ too, by Theorem 4.2. Moreover, a P-proof of tautology τ is, essentially, also a T-computation accepting $h(\tau)$: T checks that the word has the form of some $h(w)$ and that $w \in TAUT$, the latter being witnessed by any P-proof of size $\leq s_P(|w|)^{O(1)}$.

Take finitely presented group H in which G_{TAUT} embeds, constructed in [3] - Theorem 3.5, starting with algorithm T. This construction is quite effective: T is first simulated by a rewriting system S (an $\mathcal{S}$-machine in the terminology of [3]). System S is then simulated by a Boone's lemma type construction. The construction transfers first an accepting T-computation into an accepting S-computation, and the S-computation then transforms into a van Kampen diagram Δ for $h(\tau)$ in H. The area of Δ is proportional to the length of the original T-computation. Finally, Δ yields, via Lemma 3.4, an H-proof of $h(\tau) = 1$, i.e. of $\tau \in TAUT$.

The length of the H-proof is proportional to the area of Δ, i.e. to $s_P(n)^{O(1)}$. On the other hand, the bound is also (by Lemma 3.4 again) $O(D_H(|h(\tau)|) = O(D_H(O(n))$. As $k_P(n)$ is proportional to $s_P(n)$, for P of the form $EF + X$, this proves the first part of the theorem.

Now let H be a group-based propositional proof system. By Lemma 4.4 there is a polynomial time predicate $Prf_H(u,v) \subseteq (\bigcup_{i\leq 2m}\Omega_i)^* \times (\bigcup_{i\leq 2m}\Omega_i)^*$ expressing that u is an H-proof of $v \in TAUT$. Let $Taut(x)$ be the coNP-predicate $x \in TAUT$, and define two coNP predicates A and B by:

$$B(x,w) := [\exists u(|u| \leq |w|); Prf_H(u, h(x))] \to Taut(x)$$

where h is the linear time reduction from Theorem 4.2, and

$$A(x) := B(x, 1^{\ell(|h(x)|)})$$

where $\ell(n)$ bounds the size of H-proofs of $v \in W(H)$, for $|v| \leq n$. Note that $\ell(n) = O(D_H(n))$.

The set $X \subseteq TAUT$ is the set of all instances (substituting for bits of x) of $||A(x)||^n$, $n \geq 1$. Intuitively, $||A(x)||^n$ is a tautology expressing that whenever H proves that $x \in TAUT$ for $|x| \leq n$ then it is indeed so.

Now let u be an H-proof of $\tau \in TAUT$, $|\tau| = n$. The properties $|u| \leq |1^{\ell(|h(\tau)|)}|$ and $Prf_H(u,\tau)$ are true and their propositional translations in the antecedent of $||A(x)||^n(\tau)$ are just true boolean sentences and hence easy to prove in EF by evaluating the formulas to 1. Then, by the axiom $||A(x)||^n(\tau)$, the formula $||Taut(x)||^n(\tau)$ follows. But EF can also polynomially prove all formulas of the form $||Taut(x)||^n(\tau) \to \tau$ (see [15, Chpt.13]) and hence, by modus ponens, the formula τ itself.

The whole proof is polynomial in the length of u, i.e. $k_P(n) = D_H(n)^{O(1)}$. This concludes the proof of the theorem.

q.e.d.

For proof systems P, H polynomially simulating EF the two constructions are inverse to each other, up to polynomial simulation. Also, taking in the second part a proof system P that proves the soundness of the embedding h of $TAUT$ into $W(G_{TAUT})$ (a coNP-statement and hence such P exists of the form $EF + X$) the soundness of H can be expressed as simply the fact that the inclusion $G_{TAUT} \subseteq H$ is an embedding.

Let me conclude with a problem: Construct direct/combinatorial group-based proof system for the set of primes (encoded in binary, i.e. a subset of $\{0,1\}^+$) that has polynomial Dehn function. Such group-based proof system exists as the set of primes is in class NP, cf.[20]. The qualification direct/combinatorial means that it should not just use [3] to encode one of the known NP algorithms.

In fact, it would be interesting to have such a group-based proof system also for primes encoded in unary (i.e. the set of 1^p, p a prime). In this case we could use instead of G_L a simpler group with presentation

$$\langle \{a, b, c, d\} \ ; \ \{a^{-p}ba^pc^pdc^{-p} \mid p \text{ a prime } \}\rangle$$

(see the remark after Theorem 2.1). Polynomial upper bounds for proving primality are not known for any proof system simulated by EF. Lower bounds are known only for unary encoding and the Nullstellensatz based proof system (cf [16]).

Acknowledgement: I got interested in this topic after hearing two lectures by M. Bridson [6] in Oxford in Fall 1997. I am indebted to J.-C. Birget (Rutgers-Camden) and M. V. Sapir (Vanderbilt U.) for answering my queries, and to J.-C. Birget in particular, for comments on the draft of this paper.

References

[1] J.-C. Birget (1998) Time-complexity of the word problem for semigroups and the Higman embedding theorem, *Int. J. of Algebra and Computing*, **8**, pp.235-294.

[2] J.-C.Birget (2000) Reductions and functors from problems to word problems, *Theoretical Computer Science*, Vol.**237**, pp.81-104.

[3] J.-C.Birget, A.Yu.Ol'shanskii, E.Rips, and M.V.Sapir (to app.) Isoperimetric functions of groups and computational complexity of the word problem, *Annals of Mathematics*, to app.

[4] W.Boone (1958) The word problem, *Proc. Nat. Acad. Sci. USA*, **44**, pp.265-269.

[5] W.Boone (1959) The word problem, *Annals of Mathematics*, **70**, pp.207-265.

[6] M.Bridson (to app.) The geometry of the word problem, in: *Invitations to Geometry and Topology*, Oxford University Press, to app. in 2002.

[7] D. E. Cohen (1992) Isodiametric and isoperimetric inequalities for group presentations, *Int. J. of Algebra and Computing*, **1**, pp.315-320.

[8] Cook, S A. (1971) The complexity of theorem proving procedures, in: *Proc. 3^{rd} Annual ACM Symp. on Theory of Computing*, pp. 151-158. ACM Press.

[9] S.A.Cook (1975) Feasibly constructive proofs and the propositional calculus, in: *Proc. 7^{th} Annual ACM Symp. on Theory of Computing*, pp. 83-97. ACM Press.

[10] Cook, S. A., and Reckhow, A. R. (1979) The relative efficiency of propositional proof systems, *J. Symbolic Logic*, **44(1)**, pp.36-50.

[11] Garey, M.R., and Johnson, D. S. (1979) *Computers and intractability.* New York, W.H.Freeman and Co..

[12] M.Gromov (1993) Asymptotic invariants of infinite groups, in: *Geometric group theory*, Vol.**2**, eds. G.A.Niblo and M.A.Roller, London Math. Soc. Lect. Notes Ser. **182**.

[13] Hajós, G. (1961) Über eine Konstruktion nicht n-farbbarer Graphen, *Wiss. Zeitschr. Martin Luther Univ. Halle-Wittenberg*, A 10.

[14] G.Higman (1961) Subgroups of finitely presented groups, *Proc. Royal Soc. London*, Ser.A, **262**, pp.455-475.

[15] J. Krajíček (1995) *Bounded arithmetic, propositional logic, and complexity theory*, Encyclopedia of Mathematics and Its Applications, Vol. **60**, Cambridge University Press.

[16] J. Krajíček (2001) On the degree of ideal membership proofs from uniform families of polynomials over a finite field, *Illinois J. of Mathematics*, **45(1)**, pp.41-73.

[17] R.Lyndon, and P.Schupp (1977) *Combinatorial group theory*, Springer-Verlag.

[18] P.S.Novikov (1955) On the algorithmic unsolvability of the word problem in group theory, *Trudy Mat. Inst. Steklova*, **44**, p.143.

[19] A.Yu.Ol'shanskii, and M.V.Sapir (2001) Length and area functions on groups and quasi-isometric Higman embeddings, *Int. J. of Algebra and Computing*, Vol.**11(2)**, pp.137-170.

[20] Pratt, V.R. (1975) Every prime has a succinct certificate, *SIAM J. Computing*, **4**, pp.214-220.

[21] Pudlák, P. (1998) The lengths of proofs, in: *Handbook of Proof Theory*, S.R. Buss ed., Elsevier, pp.547-637.

[22] J.Rotman (1984) *An introduction to the theory of groups*, Allyn and Bacon.

[23] Urquhart, A. (1995) The complexity of propositional proofs, *Bulletin of Symbolic Logic*, **1(4)**, pp.425-467.

Mailing address:
Mathematical Institute
Academy of Sciences
Žitná 25
Prague 1, CZ - 115 67
The Czech Republic
`krajicek@math.cas.cz`