Decomposition of Tautologies into Regular Formulas and Strong Completeness of Connection-Graph Resolution

W. BIBEL

Technical University Darmstadt, Darmstadt, Germany

AND

E. EDER

University Salzburg, Salzburg, Austria

Dedicated to J. A. Robinson

Abstract. This paper addresses and answers a fundamental question about resolution. Informally, what is gained with respect to the search for a proof by performing a single resolution step? It is first shown that any unsatisfiable formula may be decomposed into regular formulas provable in linear time (by resolution). A relevant resolution step strictly reduces at least one of the formulas in the decomposition while an irrelevant one does not contribute to the proof in any way. The relevance of this insight into the nature of resolution and of the unsatisfiability problem for the development of proof strategies and for complexity considerations are briefly discussed.

The decomposition also provides a technique for establishing completeness proofs for refinements of resolution. As a first application, connection-graph resolution is shown to be strongly complete. This settles a problem that remained open for two decades despite many proof attempts. The result is relevant for theorem proving because without strong completeness a connection graph resolution prover might run into an infinite loop even on the ground level.

Categories and Subject Descriptors: F.1.3 [**Computation by Abstract Devices**]: Complexity Classes—*machine independent complexity*; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*complexity of proof procedures*; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*mechanical theorem proving, proof theory*; I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving—deduction, resolution.

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Connection-graph resolution, decomposition of unsatisfiable formulas, regular graphs, semantic trees, strong completeness

---

This work was supported in part by the DFG under Bi 228/6-2 and by ESPRIT Basic Research Action 3125 (Medlar).

Authors' addresses: W. Bibel, Technical University Darmstadt, Darmstadt, Germany; E. Eder, University Salzburg, Salzburg, Austria.

1. *Introduction*

Resolution is a key deductive technique in automated deduction. It is used in numerous systems and studied extensively from various theoretical points of view. However, one of the fundamental questions about resolution has rarely been addressed. Namely, what is gained with respect to the search for a proof by executing a single resolution step?

In order to understand this question and its relevance, recall first that any resolution prover basically consists of a single loop, which in a successful proof is executed a number of times until the empty clause is produced. At each iteration of this loop, a pair of (parent) clauses is selected and their resolvent is added to the clause set. In other words, how can the progress that is achieved by performing a single iteration be described and measured?

For comparison, assume we want to compute the string representation of a given integer n in the alphabet $\{1\}$. Again a single loop will do with adding a 1 at each iteration to the string obtained so far. Here the progress achieved at any iteration is captured by the number of missing 1's, which each time decreases by 1. Note that this number is independent of the remaining computation; it only depends on the status achieved so far and on the desired result of the completed computation. Also note that this number is the key for proving the correctness and termination of the program basically consisting of this loop. So again, what is the analogue to this number in the case of resolution?

Of course, there are already answers to the question just stated in three variants, since any completeness proof of resolution such as those in Robinson [1965], Anderson and Bledsoe [1970], and Bibel [1983] must naturally provide one. The answers known so far in the literature are, however, not satisfactory, as will become clear shortly.

Connection-graph resolution [Kowalski 1975], or cg-resolution for short, is a refinement of resolution that records part of the history of previous resolution steps in the course of a proof attempt in order to prevent a particular resolution step to be performed more than once in the same or in similar ways. Since a multiple production of the same resolvent is obviously a redundant activity not useful for any resolution prover, cg-resolution might even be regarded as the proper form of resolution.

In case of a satisfactory answer to the above question, one could have expected that a completeness proof for resolution based on it could easily be carried over to cg-resolution which, as just pointed out, differs from resolution only in this single and trivial feature of avoiding redundant repetitions of resolution steps. Attempts to carry over known completeness proofs for resolution to cg-resolution failed, however. The answers, on which these proofs are based, are either too global (e.g., saturation of the set of possible resolvents) or too particular (e.g., those based on the number of excess literals). They seem not to reveal the true nature of a resolution step to a point where basic issues could be settled for both resolution and cg-resolution at the same time.

For instance, despite numerous proof attempts including Brown [1976], Bibel [1981b], Smolka [1982], and Eisinger [1989] and unpublished ones by C. A. Johnson, N. Murray, E. Rosenthal, and G. Wrightson, it remained an open problem for two decades whether cg-resolution is bounded on the ground level (i.e., a proof is found after finitely many steps), a property needed to establish the commutativity property (see Nilsson [1980] for this sort of Church–Rosser notion) for cg-resolution that guarantees a proof if one exists, no matter which sequence of steps is taken to find it.

The main result of this paper is an answer to this fundamental question about resolution proof steps. Since the essence of this answer is of a propositional nature, we restrict a great deal of the paper to the ground level. Our answer, for simplicity outlined in this introduction only for the ground level, says that any unsatisfiable formula may be decomposed into regular formulas each provable in linear time (by resolution) and that a relevant resolution step reduces at least one of the formulas in the decomposition (and does not affect the remaining ones). There may be resolution steps that do not contribute at all to finding a proof and are in this sense considered irrelevant.

Instead of regular formulas, one may as well think of all possible semantic trees for the given formula that differ in the sequence of the variables used in their generation; in this view one may rephrase the answer as saying that at least one of these trees is reduced by a relevant resolution step (and none is increased). In fact, it has been well-known in the resolution literature that any relevant resolution step properly reduces *some* semantic tree.[1] But to our best knowledge a fixed sequence of the variables is considered rather than all possible sequences as in our approach.

With this insight, the solution to the open problem about commutativity of cg-resolution comes along easily. That is, we prove that any sequence of cg-resolution steps applied to an unsatisfiable formula will eventually yield the empty clause, provided that the sequence satisfies a simple fairness criterion that gives any connection a finite chance to be selected as the kernel of a cg-resolution step. Were the proof search of cg-resolution not commutative in this sense, then backtracking to each choice point would be required even on the ground level. This would make cg-resolution incompetitive in comparison with resolution and thus worthless for practice. The result is therefore important also for practical purposes.

Since it appears only the discovery of the decomposition of formulas finally led to the solution of this longstanding open problem, it would not be a surprise if other problems of a practical or theoretical nature could as well be solved with it. In any case, it contributes a technique for establishing completeness of resolution refinements (in addition to those mentioned above). Roughly speaking, the decompositional feature may be described as follows:

On the ground level, we consider matrices, that is, sets of clauses, in the (propositional) variables $X_1, \dots, X_n$. Of special interest in our context are the complete matrices containing all possible Boolean combinations of these variables as clauses. These complete matrices can be proved fast, by resolution through eliminating the variables one after the other (similarly as in the Davis–Putnam procedure). Each step properly reduces the matrix if subsumed clauses are eliminated. Suppose M_1 is the result of carrying out this process to some extent. Now there are $n!$ different sequences of the variables so that we may obtain matrices $M_2, \dots, M_{n!}$ similarly as M_1, each with a different sequence of the variables (and an arbitrary extent of the corresponding process) in

[1] See, for example, Robinson [1968], Kowalski and Hayes [1969], Anderson and Bledsoe [1970], and Chang and Lee [1973].

mind. The union of any tuple of matrices obtained this way is obviously unsatisfiable. More interesting is that the converse in some way holds as well (see Theorem 3.1 in Section 3). More precisely, any minimal, unsatisfiable matrix may be decomposed in $n!$ matrices which are obtained from the complete matrix the way just described.

As we said, this decomposition is the key for our completeness proof of cg-resolution, or of resolution for that matter. It focuses on a minimal subset of the given unsatisfiable set of clauses. With the fairness criterion mentioned above, any connection in this subset will eventually be selected. The resolution step performed upon it at the same time properly reduces at least one of the matrices in the decomposition (without changing the nonreduced ones) and the result of this reduction is part of the decomposition of the resulting matrix. Since there are only finitely many decomposition matrices, the empty clause must be obtained in a finite number of steps independent of the sequence of steps taken.

In the next section, we first summarize well-known concepts related with cg-resolution in order to make the paper rather self-contained (Subsection 2.1) and then introduce the concepts needed especially for our proofs (2.2). We also clarify the exact relationship with the more familiar semantic trees with which the proof could have been developed as well (2.3). In Section 3, the proof of the decomposition theorem is given. Section 4 contains the proof of the strong completeness theorem for the ground level. Section 5 presents the straightforward generalization of the strong completeness result to the first-order level. In the concluding section, possible applications to theorem proving and to complexity theory are briefly discussed.

2. *Basic Concepts*

In this section, we introduce all the concepts needed for our main results on the ground level (while the details for the first-order level are deferred until Section 5). For completeness, standard terminology from the connection-graph resolution literature is briefly summarized in the first subsection. For more details, the reader is referred to the literature (e.g., Bibel [1993]).

2.1. Review of Standard Terminology.

All concepts in this paper are based on a denumerable alphabet of variables (or propositional variables), which are denoted by X, Y, Z. As usual, any such denotations can be used with indices or other decorations. In this case of variables, this means that, for instance, X_1, Y' denote variables as well.

Definition 2.1.1. For any variable X, X and its *complement* $\bar{X}$ are called *literals* that are denoted by K, L. If $K = Y$ for some Y, then $\bar{K} = \bar{Y}$; if $K = \bar{Y}$ for some Y, then $\bar{K} = Y$. $\{X, \bar{X}\}$ is called a *complementary* pair of literals.

Clauses are finite, possibly empty sets of literals, denoted by c, d, e. For a clause c, the set of variables occurring in c is denoted by $V(c)$. A clause is said to be *tautological* if it contains two complementary literals. A clause c *subsumes* a clause d if $c \subseteq d$.

A *matrix* is a finite family $(c_\xi)_{\xi \in \Gamma}$ of clauses with some index set Γ. Matrices are denoted by M, N. If $M = (c_\xi)_{\xi \in \Gamma}$ is a matrix then each $\gamma \in \Gamma$ is said to be an *occurrence* of the clause c_γ in the matrix M. If, moreover, $L \in c_\gamma$ then the pair (L, γ), also denoted by L^γ, is said to be an *occurrence* of the literal L in the matrix M.

If S is a finite set of clauses, then $(c)_{c \in S}$ is called the *matrix corresponding to S.*

A matrix $M = (c_\xi)_{\xi \in \Gamma}$ is *subsumption-minimal* if c_ξ does not subsume c_η for any $\xi, \eta \in \Gamma$ with $\xi \neq \eta$.

By defining matrices as families rather than sets of clauses, multiple occurrences of clauses can be taken into account (which seems necessary for handling connection graphs properly). In the case of a subsumption-minimal matrix, multiple occurrences of clauses are of course not possible, so the matrix is the set of its clauses.

Matrices as thus defined could more precisely be called *matrices in normal form*. In general, *nonnormal-form matrices*, the elements of their clauses, may be (general) matrices rather than literals. Since we are mostly concerned with matrices in normal form, we refer to them as matrices for short and refer to nonnormal-form matrices in the more general case.

Matrices represent propositional formulas. For affirmative proof methods, one uses a *positive* representation of formulas by matrices, and for refutational methods, a *negative* representation of formulas by matrices. The structure of a matrix is, however, invariant with respect to the sign of the representation. So it is left to the reader to think of a clause as of a disjunction (as in the negative representation) or of a conjunction of literals (as in the positive representation). For proof purposes, the difference is absolutely negligible.

Many concepts involving matrices are intuitively best understood if one thinks of matrices as of two-dimensional structures similar to the matrices in linear algebra. Used to the positive representation, we present clauses vertically as columns in the matrix. Hence, the matrix corresponding to the clause set $\{\{X, Y\}, \{\bar{X}, Y\}, \{X\}, \{\bar{X}, \bar{Y}\}\}$ two-dimensionally is displayed as

$$\begin{array}{cccc} X & \bar{X} & X & \bar{X} \\ Y & Y & & \bar{Y} \end{array}$$

or more elaborately as

$$\left[\begin{bmatrix} X \\ Y \end{bmatrix} \quad \begin{bmatrix} \bar{X} \\ Y \end{bmatrix} \quad [X] \quad \begin{bmatrix} \bar{X} \\ \bar{Y} \end{bmatrix} \right].$$

Definition 2.1.2. A *path p* through a matrix $(c_\xi)_{\xi \in \Gamma}$ is a family $(L_\xi)_{\xi \in \Gamma}$ of literals where $L_\xi \in c_\xi$ for each $\xi \in \Gamma$.

A *connection* in a matrix $M = (c_\xi)_{\xi \in \Gamma}$ is a two-element set $\{K^\gamma, \bar{K}^\delta\}$ of occurrences of two complementary literals, $K, \bar{K}$, in M where $\gamma \neq \delta$. Connections are denoted by k, ℓ.

A set of connections in a matrix is called a *mating.*

A connection $\{K^\gamma, \bar{K}^\delta\}$ in M is *tautological* if there is a connection $\{L^\gamma, \bar{L}^\delta\}$ in M with $K \neq L$.

A path $(L_\xi)_{\xi \in \Gamma}$ is *complementary* if there is a connection $\{L_\gamma^\gamma, \bar{L}_\delta^\delta\}$. A matrix M is called *complementary* if all its paths are complementary. Any mating rendering M complementary is called *spanning* for M.

If $M = (c_\xi)_{\xi \in \Gamma}$ is a matrix and $\Delta \subseteq \Gamma$, then the matrix $M = (c_\xi)_{\xi \in \Delta}$ is said to be a *submatrix* of M.

A complementary matrix M is called a *minimal matrix* if there is no proper submatrix in M which is complementary.

Obviously, any complementary matrix has a submatrix that is minimal. It is known that none of the clauses in a minimal matrix is tautological, which is why we shall concentrate on matrices without tautological clauses.

Complementarity of a matrix M is a necessary and sufficient condition for the formula represented by M to be a tautology (or, in the negative representation, to be unsatisfiable). Consequently, all proof procedures test for complementarity of matrices in one way or another. For instance, resolution can be seen this way [Bibel 1993; Bibel and Eder 1993].

Complementarity is characterized by a spanning mating. It has therefore been a natural idea to design a form of resolution that handles connections in an explicit way; it is called connection-graph resolution [Kowalski 1975] (cg-resolution) as follows:

Definition 2.1.3. A *connection graph* G is a pair (M, C) where M is a matrix and C is a mating.

If $M = (c_\xi)_{\xi \in \Gamma}$, then we say that the c_ξ with $\xi \in \Gamma$ are the *clauses* of G, or equivalently, that G *has* (or *contains*) the clauses c_ξ with $\xi \in \Gamma$. A connection graph is said to be *tautology-free* if it does not have a tautological clause.

G is called *full* if, for any connection k in M, we have $k \in C$; C is called the *full mating* for M in this case. For any matrix $M = (c_\xi)_{\xi \in \Gamma}$, the *complete mating* C_M is defined as

$$C_M = \{\{L^\gamma, \bar{L}^\delta\} \mid \gamma, \delta \in \Gamma, \gamma \neq \delta \text{ and there is no } \{K^\gamma, \bar{K}^\delta\} \text{ with } K \neq L\}.$$

Let $G = (M, C)$ and $H = (N, D)$ be two connection graphs. Then $G \leq H$ if M is a submatrix of N and $C \subseteq D$.

A *comp-graph* is a graph consisting of a complementary matrix and its complete mating. A *minimal comp-graph* is a comp-graph G such that for each comp-graph H with $H \leq G$ it holds $H = G$.

Let $G = (M, C)$ be a connection graph, $M = (c_\xi)_{\xi \in \Gamma}$, and $c_\xi \neq \emptyset$ for $\xi \in \Gamma$. Let $\ell = \{L^\gamma, \bar{L}^\delta\} \in C$. Let

(1) θ be some token with $\theta \notin \Gamma$;
(2) $\Theta := \Gamma \cup \{\theta\}$;
(3) $c_\theta := (c_\gamma \backslash \{L\}) \cup (c_\delta \backslash \{\bar{L}\})$;
(4) $M_\ell := (c_\xi)_{\xi \in \Theta}$;
(5) $C_\ell := (C \backslash \{\ell\}) \cup D$, where $D := \{\{K^\xi, \bar{K}^\theta\} \mid \xi \in \Gamma$, and $\{K^\xi, \bar{K}^\gamma\} \in C$ or $\{K^\xi, \bar{K}^\delta\} \in C\}$.
(6) $G_\ell := (M_\ell, C_\ell)$.

Then the pair (G, G_ℓ), written $G \vdash G_\ell$, is called a *cg-resolution step* and ℓ is called its *kernel.*

A connection graph G is called *refutable by cg-resolution* if $G \vdash^* G'$ holds for some connection graph G' which contains the empty clause.

A cg-resolution step is illustrated in Figure 1.

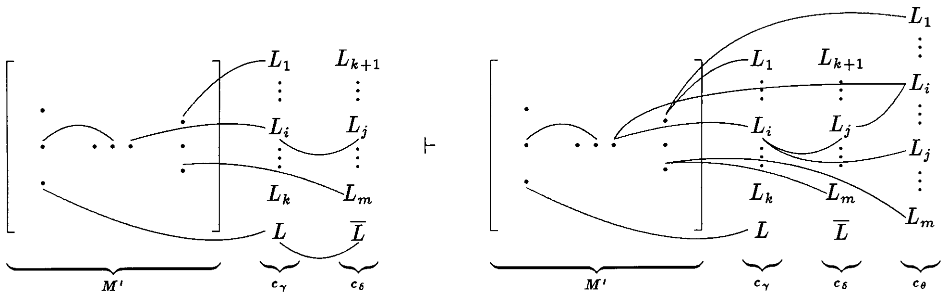

FIG. 1. A cg-resolution step.

cg-resolution was shown to be a sound and complete deductive rule in Bibel [1981a].[2] For practical purposes a stronger completeness property would be desirable that is introduced further below. Also for practical reasons, provers based on cg-resolution must incorporate the following two well-known reduction operations:

Definition 2.1.4. Let (M, C), $M = (c_\xi)_{\xi \in \Gamma}$, denote a connection graph with some clause c_γ and some literal $L \in c_\gamma$. If for no connection $k \in C$, we have $L^\gamma \in k$ then L^γ is called *pure*. In this case we say that $((c_\xi)_{\xi \in \Gamma \setminus \{\gamma\}}, C \setminus D)$ is obtained from (M, C) by *purity reduction* upon L^γ whereby $D = \{\{K^\gamma, \bar{K}^\xi\} \in C \mid \xi \in \Gamma\}$. The graph obtained from a graph G by applying purity reduction until it can be applied no more is called the *purity-reduced graph* $p(G)$.

For two clause occurrences $\gamma, \delta \in \Gamma$, we say γ *cg-subsumes* δ if

(1) $c_\gamma \subseteq c_\delta$ and
(2) if $\{L^\delta, \bar{L}^\xi\} \in C$ for any $L \in c_\gamma$ and $\xi \in \Gamma$ then $\{L^\gamma, \bar{L}^\xi\} \in C$.

In this case, we say that $((c_\xi)_{\xi \in \Gamma \setminus \{\delta\}}, C \setminus D)$ is obtained from (M, C) by *cg-subsumption* whereby $D = \{\{K^\delta, \bar{K}^\xi\} \in C \mid \xi \in \Gamma\}$. The graph obtained from a graph G by applying cg-subsumption until it can be applied no more is called the *subsumption-reduced graph* $s(G)$.

Purity reduction allows the deletion of clauses with unconnected literals as well as their connections. Subsumption allows the deletion of a subsumed clause (as well as its connections) that contains all the literals of the subsuming clause along with some of their connections and no further ones. It was shown in Bibel [1981a] that these two reduction operations retain the spanning property so that they can be applied without any harm with respect to provability at any point in the deductive process. For the purposes of this paper, we stipulate that they are applied after each cg-resolution step. In other words, if we write $(M, C) \vdash (M_\ell, C_\ell)$, we actually mean $(M, C) \vdash p(s((M_\ell, C_\ell)))$.

For a nonnegative integer n, the set of permutations of $\{1, \ldots, n\}$ is denoted by S_n.

[2] In Bibel [1981a], a more restricted form of cg-resolution was used; for our purposes, the simpler form defined here will do as well.

2.2. Additional Terminology. In the present subsection, we introduce the special terminology needed for the subsequent proofs of our main results. We begin by defining what we consider to be a fair derivation.

Definition 2.2.1. The set of natural numbers (i.e., nonnegative integers) is denoted by $\mathbb{N}$.

An *initial segment of* $\mathbb{N}$ is a (finite or infinite) set $I \subseteq \mathbb{N}$ such that $j < i$ implies $j \in I$ for all $i \in I$ and $j \in \mathbb{N}$.

A *derivation* is a sequence $(G_i)_{i \in I}$ of connection graphs where I is an initial segment of $\mathbb{N}$ and $G_{i-1} \vdash G_i$ for all $i \in I \backslash \{0\}$. We say $(G_i)_{i \in I}$ *starts with* G_0, and it *ends with* G_n if I is bounded and $n = \max(I)$.

A derivation $(G_i)_{i \in I}$ is a *refutation* if I is bounded and $G_{\max(I)}$ contains the empty clause. In this case, $(G_i)_{i \in I}$ is a *refutation of* G_0.

For two derivations $(G_i)_{i \in I}$ and $(H_i)_{i \in J}$, we say $(G_i)_{i \in I} \leq (H_i)_{i \in J}$ if $I \subseteq J$ and $G_i = H_i$ for all $i \in I$. A derivation is *maximal* if there is no strictly greater (with respect to $\leq$) derivation.

A derivation $((M_i, C_i))_{i \in I}$ is *fair* if it is finite or

$$\bigcap_{\substack{i \in I \\ i \geq n}} C_i = \emptyset$$

for any $n \in I$.

In words, the fairness condition means that the derivation has to be finite or each connection occurring in one connection graph of the derivation is resolved upon at some later stage of the derivation and therefore not present any more in the connection graphs thereafter.

Definition 2.2.2. cg-resolution is called *strongly complete* if, for any complementary matrix M and its full mating C, every fair maximal derivation starting with the connection graph (M, C) is a refutation.

Whether or not cg-resolution indeed enjoys this property of being strongly complete was an open problem for two decades, which is now settled positively in the present paper. Given this property, a cg-resolution prover may proceed in its search for a proof without any backtracking.

The solution to this problem will involve a very special sort of connection graph, which we are studying in some detail in the rest of the section. In the following, we assume that n is a fixed integer and that $X_1, \ldots, X_n$ are fixed pairwise distinct variables. Further, we assume that no other variables occur in the matrices we consider. Recall the notion C_M of the complete mating for a matrix M from Definition 2.1.3 used in the following definition:

Definition 2.2.3. The *complete matrix* $M^n(X_1, \ldots, X_n)$ in the variables $X_1, \ldots, X_n$, abbreviated M^n, is the matrix $M^n = (c)_{c \in N^n}$ corresponding to the clause set N^n, which is defined inductively as follows:

(1) $N^1 = \{\{X_1\}, \{\overline{X_1}\}\}$.
(2) If $N^{n-1} = \{c_1, \ldots, c_{2^{n-1}}\}$, then

$$N^n = \{c_1 \cup \{X_n\}, \ldots, c_{2^{n-1}} \cup \{X_n\}, c_1 \cup \{\overline{X_n}\}, \ldots, c_{2^{n-1}} \cup \{\overline{X_n}\}\}.$$

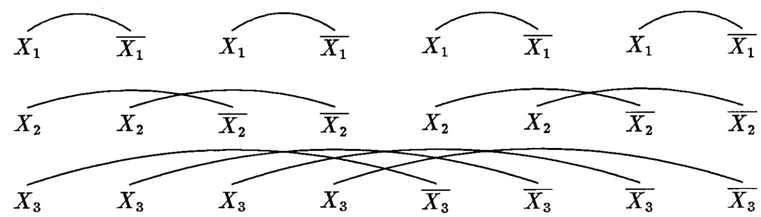

FIG. 2. The complete connection graph G^3 (or $G(X_1, X_2, X_3)$).

The *extension* of a clause c with $V(c) \subseteq \{X_1, \ldots, X_n\}$ in M^n is $\mathrm{Ext}(c) = \{d \in M^n \mid c \subseteq d\}$.

$G(X_1, \ldots, X_n) = (M^n, C_M{}^n)$, or abbreviated G^n, is called the *complete connection graph* (in the variables $X_1, \ldots, X_n$).

For illustration of these concepts, we present in Figure 2 the complete connection graph G^3, which has eight clauses, $c_1, \ldots, c_8$. The following lemma is easy to prove:

LEMMA 2.2.4. *For any clause c in a minimal complementary matrix M in the variables $X_1, \ldots, X_n$, there is a clause $e \in M^n$ such that $e \in Ext(c)$, but $e \notin Ext(d)$ for any clause $d \in M$ with $d \neq c$.*

Note that the complete mating C_{M^3} is not spanning for M^3 since there are a number of paths through M^3 that do not contain any connection from C_{M^3} (as the reader may easily check). The same applies to the general case C_{M^n}. The connections in C_{M^n} are still sufficient to prove M^n. For this reason, we introduce a weaker spanning concept such that C_{M^3} is weakly spanning for M^3 (or C_{M^n} for M^n for that matter).

Definition 2.2.5. Let M be a matrix. Then we denote by C_M^{taut} the set of tautological connections of M. A mating C in a matrix M is said to be *weakly spanning* for M if $C \cup C_M^{\mathrm{taut}}$ is spanning for M. A connection graph (M, C) is said to be *complementary* if C is spanning for M. It is said to be *weakly complementary* if C is weakly spanning for M.

As the reader may note, there are different orders in which the connections in M^n can be resolved. As we shall see, particularly simple matrices result if the order chosen corresponds to an ordering of variables, represented by a permutation in S_n.

Definition 2.2.6. Let c be a (nontautological) clause and let $\pi \in S_n$. Then the *π-front literal* of c is the literal $X_{\pi(i)}$ or $\overline{X_{\pi(i)}}$ of c with the smallest possible value of i.

Let $G = (M, C)$ be a tautology-free connection graph, $M = (c_\xi)_{\xi \in \Gamma}$, and $\pi \in S_n$. A *π-front connection* in G is a connection $\{K^\gamma, \bar{K}^\delta\}$ in C where K is the π-front literal of c_γ and $\bar{K}$ is the π-front literal of c_δ.

Let G be a connection graph, $\pi \in S_n$, ℓ be a π-front connection of G, $G \vdash G_\ell$ a cg-resolution step upon ℓ, and G_ℓ^{red} be the result of applying purity reduction to G_ℓ. Then we say that G_ℓ^{red} is obtained by *π-regular reduction* from G.

Let $\pi \in S_n$. Then *π-regular graphs* are connection graphs defined inductively as follows:

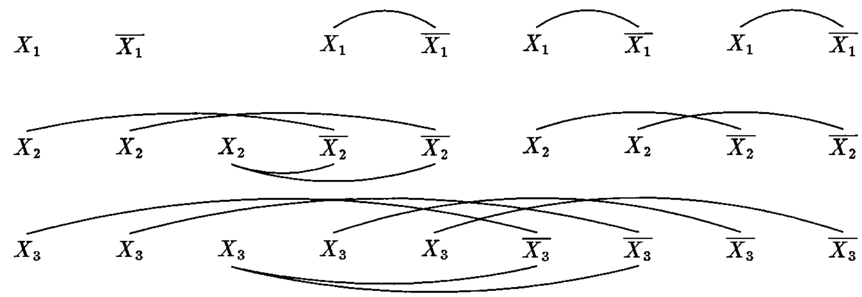

FIG. 3. Graph obtained from G^3 by cg-resolution.

(1) $G(X_1, \ldots, X_n)$, or G^n, is a π-regular graph.
(2) If G is a π-regular graph and G_ℓ^{red} is obtained from G by π-regular reduction, then G_ℓ^{red} is a π-regular graph.

A graph is called *regular* if it is π-regular for some $\pi \in S_n$.

Let $\pi \in S_n$. By a *π-clause*, we mean a nontautological clause c such that there is a nonnegative integer $j \leq n$ such that $V(c) = \{X_{\pi(j+1)}, \ldots, X_{\pi(n)}\}$.

For the graph G^3 presented above, let us consider the identity permutation π_1, $\pi_1(i) = i$ for $i = 1, 2, 3$. We introduce a systematic indexing for regular graphs by coding the permutation in the left lower index and the pattern of the π-front connections in the right lower index of the denotation for the graph. This way G^3 becomes ${}_{123}G^3_{1111}$ reflecting the permutation π_1 and the presence of all four π-front connections. Selecting as ℓ the leftmost π-front connection, that is, the connection between X_1 in the first clause and $\overline{X_1}$ in the second clause, cg-resolution upon ℓ yields the graph shown in Figure 3.

Since X_1 and $\overline{X_1}$ are pure in the leftmost two clauses, these are deleted in the subsequent purity operation, resulting in the π_1-regular graph ${}_{123}G^3_{0111}$ shown in Figure 4, which altogether is the result of π_1-regular reduction in this case. Note the change from 1 to 0 in the first position of the right lower index reflecting the removal of the first π-front connection. Since, in this case, only the first value of the permutation matters, we may denote the graph also by ${}_{1\,.\,.}G^3_{0111}$, whereby the dots are placeholders for arbitrary values.

At this point, again only the three connections in the first row are π_1-front connections, which can be resolved in order to obtain a π-regular graph. It is easy to see that, in this way, $G(X_1, \ldots, X_n)$ can be proved in $\sum_{i=0}^{n-1} 2^i = 2^n - 1$ steps (without counting the reductions separately), that is, in seven steps in the case of our example with $n = 3$, which has eight clauses; in other words, these proofs are linear in the number of clauses. These proofs are *regular resolution proofs* in the sense of this term as used in complexity theory [Tseitin 1968].

The following three lemmas are easy to prove:

LEMMA 2.2.7. *Every clause in a π-regular graph is a π-clause.*

LEMMA 2.2.8. *Every nontautological clause is a π-clause for some $\pi \in S_n$.*

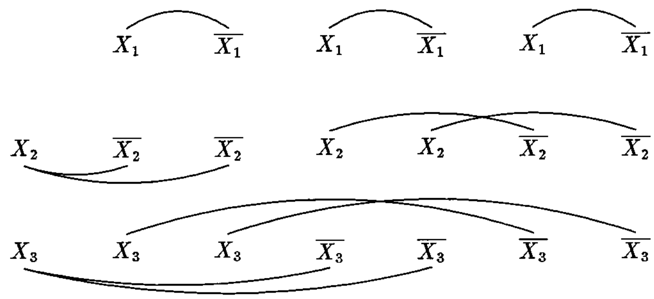

FIG. 4. The regular graph ${}_{1\,.\,.}G^3_{0111}$.

LEMMA 2.2.9. *Let $\pi \in S_n$ and let M be the matrix corresponding to a complementary subsumption-minimal set of π-clauses. Then (M, C_M) is a π-regular graph.*

The regular graphs are but a small fraction of the complementary graphs. However, any complementary graph is in some way composed of regular graphs. The composition operation basically consists of set union followed by subsumption and purity reduction.

Definition 2.2.10. Let $(G_i)_{i\in I}$, $G_i = ((c_{i\xi})_{\xi\in\Gamma_i}, C_i)$, be a finite family of connection graphs. Then the *sum* ${}_{i\in I}\, G_i$ of this family is the connection graph (M, C) defined as follows:

(1) $\Gamma := \{(i, \xi) \mid i \in I \text{ and } \xi \in \Gamma_i\}$;
(2) $M := (c_{i\xi})_{(i,\xi)\in\Gamma}$;
(3) $C := \{\ell \mid \ell = \{K^{(i,\xi)}, \bar{K}^{(j,\eta)}\}$ and either $i = j$ and $\{K^\xi, \bar{K}^\eta\} \in C_i$ or $i \neq j$ and ℓ nontautological$\}$.

The *union* $\bigcup_{i\in I} G_i$ of a finite family $(G_i)_{i\in I}$ of connection graphs is defined by

$$\bigcup_{i\in I} G_i := p\left(s\left(\sum_{i\in I} G_i\right)\right).$$

In the graphical representation of connection graphs, the sum ${}_{i\in I}\, G_i$ is obtained by writing the graphs G_i unconnected next to each other and adding all nontautological connections between the different components. From this, the union is obtained by applying the operations introduced in Definition 2.1.4. For practical purposes, the construction can be optimized by suppressing the addition of any connections between different components and adapting the subsumption operation accordingly.

Definition 2.2.11. A *regular family* is a family $(G_\pi)_{\pi\in S_n}$ such that G_π is a π-regular graph for each $\pi \in S_n$.

A (full) *decomposition* of a minimal comp-graph G is a regular family $(G_\pi)_{\pi\in S_n}$ such that $G = \bigcup_{\pi\in S_n} G_\pi$ holds.

Let $G = (M, C_M)$, $M = (c_\xi)_{\xi\in\Gamma}$, $G_\pi = (M_\pi, C_{M_\pi})$, $M_\pi = (c_\xi)_{\xi\in\Gamma_\pi}$ for any $\pi \in S_n$, and $G = \bigcup_{\pi\in S_n} G_\pi$. Then the decomposition $(G_\pi)_{\pi\in S_n}$ is said to *have property P* if the following two properties hold.

(1) For any $\gamma \in \Gamma$ and for any $\pi \in S_n$ such that c_γ is a π-clause, there exists a $\delta \in \Gamma_\pi$ such that $c_\gamma = c_\delta$.
(2) For any $\gamma \in \Gamma$, any $\{K^\gamma, \bar{K}^\delta\} \in C_M$ for some $\delta \in \Gamma$, and for any $\pi \in S_n$ such that $c_\gamma \cup (c_\delta\backslash\{\bar{K}\})$ is a π-clause with π-front literal K, there exist λ, $\rho \in \Gamma_\pi$ such that $c_\lambda = c_\gamma \cup (c_\delta\backslash\{\bar{K}\})$, $c_\rho = (c_\gamma\backslash\{K\}) \cup c_\delta$, and $\{K^\lambda, \bar{K}^\rho\} \in C_{M_\pi}$.

In other words, this means that G is the union of the regular graphs G_π, that is, that these form a decomposition in a certain sense of the word. Property P says first that any clause c_γ from the given matrix is contained in the component graph G_π, whenever π is such that c_γ is a π-clause. It says secondly that for any connected clauses c_γ and c_δ in G their union with the resolvent is contained in G_π, whenever π is such that this union is a π-clause and the connected literal is a π-front literal.

According to this definition, any full decomposition consists of $n!$, that is, exponentially many components. As we will see in the examples of the subsequent sections, G is often determined by a small subset of characterizing components, which in turn corresponds to some subset of S_n. For the remaining permutations not in this subset, the corresponding regular graphs are irrelevant since all their clauses are subsumed by those in the characterizing components. These observations give rise to further concepts worth being noted here, although they will not be used in the present paper in any substantial way.

Let $(G_\pi)_{\pi\in S_n}$ be a decomposition of a minimal comp-graph G. A component G_{π_0} in the decomposition is called *minimal* if $(G'_{\pi_0} \cup \bigcup_{\pi\in S_n, \pi\neq\pi_0} G_\pi) \neq G$ holds for any graph G'_{π_0} obtained from G_{π_0} by π_0-regular reduction. For a subset $S \subseteq S_n$, the regular family $(G_\pi)_{\pi\in S}$ is also called a *decomposition* (not necessarily full), provided $G = \bigcup_{\pi\in S} G_\pi$ holds. Any such decomposition is called *standard* if each of its components is minimal. It is called *minimal* if $\bigcup_{\pi\in S'} G_\pi \neq G$ holds for any proper subset $S' \subset S$.

As mentioned before, we will consider only full decompositions in this paper. In fact we may further assume that any such decomposition is standard without restriction of generality.

2.3. The Relationship with Semantic Trees. Regular graphs are closely related with semantic trees. This relationship will now be clarified.

One notices that, with a trivial inductive argument, the mating of any regular graph is complete. As we have seen in the case of G^3 (see Definition 2.2.5), complete matings are not necessarily spanning. On the other hand, they are sufficient for yielding resolution proofs. The reason is that resolution incorporates factorization. For instance, the resolvent $\{X_2, X_3\}$ of the resolution step upon the leftmost π-front connection, $\pi_1(i) = i$ for $i = 1, 2, 3$, in G_3 shown in Figure 2 is the result of factorizing $\{X_2^{c_1}, X_2^{c_2}\}$ and $\{X_3^{c_1}, X_3^{c_2}\}$, where the upper indices indicate the (occurrences of the) parent clauses of the operation.

If we perform all the factorizations of such a regular proof of G^n prior to the actual proof, then we get a graph F^n whose matrix is in nonnormal form. For instance, F^3 obtained this way from G^3 looks as shown in Figure 5.

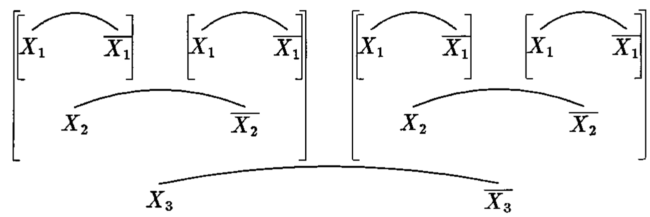

FIG. 5. The fully factorized form F^3 of the π_1-regular graph ${}_{123}G^3_{1111}$.

This observation leads to the following definition:

Definition 2.3.1. The *fully factorized form* of a π-regular graph G is inductively defined as follows.

(1) The graph

$$X_{\pi(1)} \qquad \overline{X}_{\pi(1)}$$

is in fully factorized form.

(2) If F_1 is the fully factorized form of the subgraph G_1 obtained from G by deleting all clauses with an occurrence of $X_{\pi(n)}$ and by deleting the occurrences of $X_{\pi(n)}$ from all remaining clauses and if F_2 is obtained similarly except that the roles of $X_{\pi(n)}$ and $\overline{X}_{\pi(n)}$ are exchanged, then

$$\begin{array}{cc} F_1 & F_2 \\ X_{\pi(n)} & \overline{X}_{\pi(n)} \end{array}$$

is the fully factorized form of G.

It is easy to see that the mating in such a fully factorized form of a π-regular graph is spanning while the mating in the graph itself is only weakly spanning, which illustrates the relationship of these two notions.

The close relationship of regular graphs and semantic trees is particularly obvious if the graph is in fully factorized form. This is illustrated by the semantic tree shown in Figure 6 for the matrix of G^3. Except for the root of the tree, each node corresponds to a literal occurrence in F^3 shown in Figure 5, and vice versa. If one of the figures is turned upside down, it becomes obvious that they represent just different ways of depicting the same construct. This observation, which was made for this particular graph, is true in general. Hence, one could have developed the entire paper on the basis of semantic trees. Readers more familiar with semantic trees might therefore prefer to think in terms of semantic trees whenever the paper refers to regular graphs.

3. *Decomposing Minimal Comp-Graphs into Regular Graphs*

In this section, we show that every minimal comp-graph can be decomposed into regular graphs (Theorem 3.1). We illustrate this decomposition of a complementary matrix by the following graphs in X_1, X_2, and X_3, starting with G^3 from Figure 2 in the previous section.

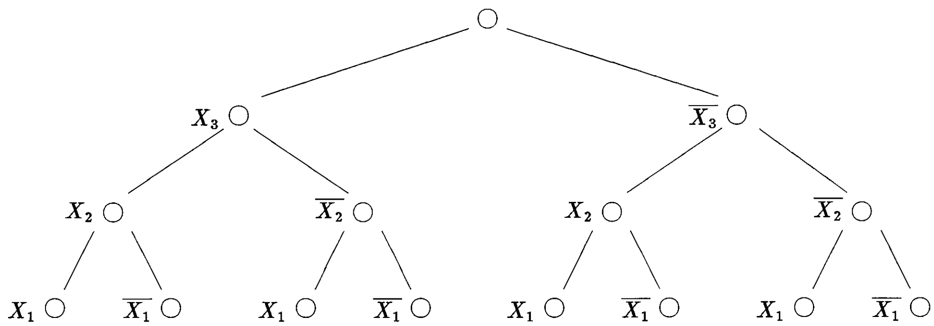

FIG. 6. The semantic tree corresponding to F^3 from Figure 5.

Consider the three permutations π_1, π_2, and π_3, defined by $\pi_1(i) = i$ for $i = 1, 2, 3$, $\pi_2(1) = 3$, $\pi_2(2) = 2$, $\pi_2(3) = 1$, $\pi_3(1) = 2$, $\pi_3(2) = 1$, $\pi_3(3) = 3$. As discussed earlier, one π_1-regular reduction of ${}_{1\,..}G^3_{1111}$ yields the graph ${}_{1\,..}G^3_{0111}$ already shown in Figure 4. Similarly, we obtain the graph ${}_{2\,..}G^3_{0111}$ shown in Figure 7 by π_2-regular reduction from ${}_{2\,..}G^3_{1111}$. Finally, the graph ${}_{3\,..}G^3_{0111}$ shown in Figure 8 is obtained by π_3-regular reduction from ${}_{3\,..}G^3_{1111}$.

The union of these three graphs ${}_{i\,..}G^3_{0111}$, $i = 1, 2, 3$ yields the graph G shown in Figure 9, which already looks pretty irregular. Similarly, G' shown in Figure 10 is $\bigcup_{i=1}^3 {}_{i\,..}G^3_{0110}$, whereby ${}_{i\,..}G^3_{0110}$ is obtained from ${}_{i\,..}G^3_{0111}$ by π_i-regular reduction upon the rightmost connection in row $\pi_i(1)$ of ${}_{i\,..}G^3_{0111}$.

As we see, the composition (or union) of regular graphs yields (fairly irregular) graphs. We now show that, conversely, every minimal comp-graph can be decomposed into such regular graphs.

THEOREM 3.1 (DECOMPOSITION). *Every minimal comp-graph has a decomposition satisfying property P.*

PROOF. The proof proceeds by constructing the decomposition of the given minimal comp-graph $G = (M, C_M)$ with $M = (c_\xi)_{\xi\in\Gamma}$ in an inductive way from G^n. The induction is on the difference w_Γ of the number of literals in the complete matrix $M^n(X_1, \ldots, X_n)$ and in M, that is, $w_\Gamma = n \times 2^n - {}_{\xi\in\Gamma} |c_\xi|$, whereby $X_1, \ldots, X_n$ are the variables occurring in M.

The base case in this inductive definition is $G = G^n$, or $w_\Gamma = 0$. In this case, we consider the regular family $(G(X_{\pi(1)}, \ldots, X_{\pi(n)}))_{\pi\in S_n}$. Both parts of property P trivially hold in this case.

Assume $G \neq G^n$ and hence $w_\Gamma > 0$. Then there is at least one clause c_γ, $\gamma \in \Gamma$, with $|c_\gamma| < n$. Let $\gamma_0, \gamma_1 \notin \Gamma$, $\Gamma'' = (\Gamma\backslash\{\gamma\}) \cup \{\gamma_0, \gamma_1\}$, $X_i \notin V(c_\gamma)$, $c_{\gamma_0} = c_\gamma \cup \{X_i\}$, $c_{\gamma_1} = c_\gamma \cup \{\overline{X_i}\}$, $M' = s((c_\xi)_{\xi\in\Gamma''}) = (c_\xi)_{\xi\in\Gamma'}$ for some $\Gamma' \subseteq \Gamma''$, and $G' = (M', C_{M'})$.

More informally, we perform sort of an extension step that is the inverse of a resolution step, or more specifically, the inverse of a π-regular reduction step. The structure of the remaining proof is illustrated in the left part of the diagram shown in Figure 11 (while the right part will be used for illustration of the proof of the strong completeness theorem in the next section). By the induction hypothesis, we may assume that there is a decomposition of the comp-graph G'. This decomposition will be transformed by reductions upon π-front connections

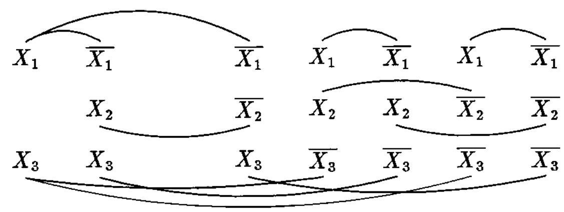

FIG. 7. The regular graph ${}_{2\,..}G^3_{0111}$.

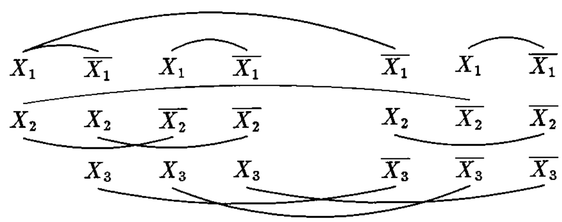

FIG. 8. The regular graph ${}_{3\,..}G^3_{0111}$.

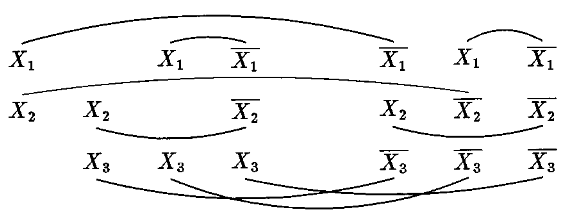

FIG. 9. The (irregular) graph G composed of ${}_{i\,..}G^3_{0111}$, $i = 1, 2, 3$.

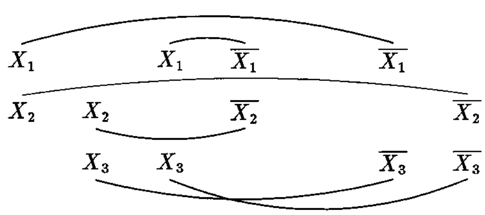

FIG. 10. The (irregular) graph G composed of ${}_{i\,..}G^3_{0110}$, $i = 1, 2, 3$.

that correspond to the connection used for the extension step. We finally establish that the resulting family is a decomposition of the given graph.

Resuming the formal proof we first note that G' obviously is a comp-graph by construction, given that G is one. We then claim that $w_{\Gamma'} < w_{\Gamma}$. The main question in this regard is what the effect of subsumption might be under the given circumstances. Since G is minimal, only c_{γ_0} or c_{γ_1} could be subsumed in $M'' = (c_\xi)_{\xi \in \Gamma''}$ by some other clause. For the same reason, both c_{γ_0} and c_{γ_1} cannot be subsumed in M'' by some other clauses. Namely, were there clauses $c_{\rho_0} \subset c_{\gamma_0}$ and $c_{\rho_1} \subset c_{\gamma_1}$ in M'', then the union of the extensions of c_{ρ_0} and c_{ρ_1} would contain the extension of c_γ, thus violating the property of minimal matrices stated in Lemma 2.2.4. Therefore, either one clause out of c_{γ_0}, c_{γ_1} is subsumed, resulting in $w_{\Gamma'} = w_\Gamma - 1$, or none, resulting in $w_{\Gamma'} = w_{\Gamma''} = w_\Gamma - |c_\gamma| - 2$.

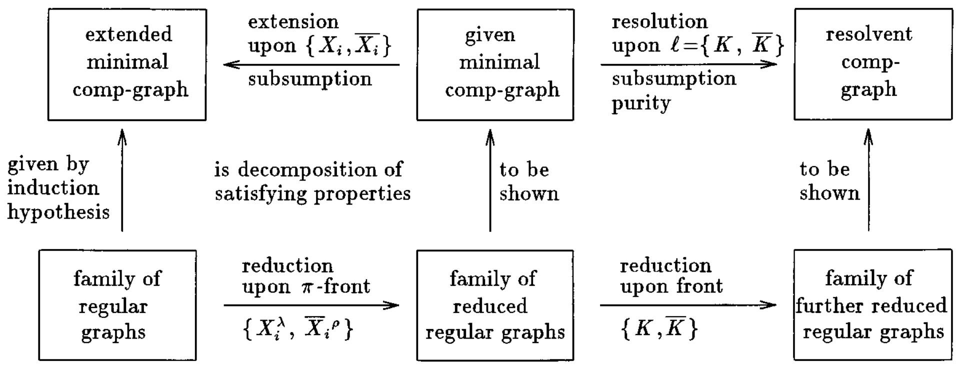

FIG. 11. The structure of the proofs of the two main theorems.

In any case, the claim is established. For simplicity, we may assume in the following, without restricting generality, that $\gamma_0 \in \Gamma'$.

We can therefore assume by the induction hypothesis that the theorem holds for G', which obviously is a minimal comp-graph. In other words, there is a decomposition $(G'_\pi)_{\pi \in S_n}$ of G' satisfying property P. For any $\pi \in S_n$ such that the matrix M'_π of G'_π contains a clause $c_\lambda \supseteq c_{\gamma_0}$ and X_i^λ is the π-front literal in c_λ, let G_π be the regular graph obtained from G'_π by π-regular reduction upon the π-front connection $\{X_i^\lambda, \bar{X}_i^\rho\}$ for some $\rho \in \Gamma_\pi$. For any other $\pi \in S_n$, let $G_\pi = G'_\pi$.

We claim that $(G_\pi)_{\pi \in S_n}$ is the desired decomposition of G satisfying property P and are done with the entire proof once this claim is established. We first show that $M = \tilde{M}$, where $\tilde{M}$ is the matrix of $\bigcup_{\pi \in S_n} (G_\pi)$. Because of the first part of property P, there are permutations π such that M'_π contains a clause $c_{\delta_0} = c_{\gamma_0}$ and $X_i^{\delta_0}$ is its π-front literal. Because of the regularity of the M'_π, $c_{\delta_1} = (c_{\delta_0} \setminus \{X_i^{\delta_0}\}) \cup \{\bar{X}_{i_1}^{\delta}\}$ is also a clause of M'_π. By definition, M_π thus contains c_γ. Else, M_π contains only clauses also contained in M'_π, a fact which is true also for any other π. With this observation, the equality is obvious, since also M differs from M' only in c_γ.

We also see immediately that the first part of property P carries over for the same reasons. Similarly, by those π-regular reductions just described only the π-front connections disappear from $\bigcup_{\pi \in S_n} (G'_\pi)$ in the transition to $\bigcup_{\pi \in S_n} (G_\pi)$ while all other connections remain the same, which again reflects exactly the transition from G' to G with respect to their connections. Thus, $G = \bigcup_{\pi \in S_n} (G_\pi)$. Since any connection in G is also contained in G' where the second part of property P holds by the induction hypothesis, it is clear that this part of property P is also true for G. Q.E.D.

For any minimal comp-graph G and any $\pi \in S_n$ the minimal component G_π of a decomposition may be algorithmically determined by starting with the complete π-regular graph and performing as many π-regular reductions as possible (which, in a way, simulates the proof just presented). By carrying this out for all possible permutations, the standard decomposition of G may be obtained this way. Since this will usually amount to computationally very expensive efforts, we are not suggesting such an approach for theorem proving purposes unless it is taken in some restricted way.

4. *Strong Completeness of Ground cg-Resolution*

As announced in the introduction, we apply the decomposition theorem 3.1 to prove strong completeness of cg-resolution in the present section for the ground level and in the subsequent section for the general level. The significance of this result, apart from solving a longstanding problem, lies in a clarification of the fundamental question about resolution discussed in the introduction. There we mentioned the representation of natural numbers as an analogue example and pointed out the role of the number of missing 1's as a measure for the progress in the computation. We are now in a position to define an analogue measure for resolution.

Definition 4.1. The *connectivity measure* $\mu(\mathcal{F})$ of a regular family $\mathcal{F} = (M_\pi, C_\pi)_{\pi \in S_n}$ is the sum $\quad_{\pi \in S_n} |C_\pi|$.

In other words, we count the number of connections in each component and take the sum of the results. Having this measure, we are now in a position to prove the strong completeness theorem. Its proof uses a result from Bibel [1981a], which is stated first, using the notation from Definition 2.1.3.

THEOREM 4.2. *A connection graph G is complementary iff G_ℓ is complementary.*

THEOREM 4.3. *cg-resolution* (*on the ground level*) *is strongly complete.*

PROOF. We are given an arbitrary complementary connection graph H with a full mating. The theorem then claims that any fair maximal derivation $\Delta = ((N_i, D_i))_{i \in I}$ (see Definition 2.2.1) starting with H is a (finite) refutation.

Since H is complementary, its matrix has a minimal complementary submatrix M. Since the mating D_0 of H is full, $G = (M, C_M) \leq (N_0, D_0) = H_0 = H$ holds (for the notation see Definition 2.1.3. If the kernel ℓ of the first step of the derivation Δ is not in C_M, then $G \leq (N_1, D_1) = H_1$ holds as well according to an obvious lemma, which is stated below for completeness. By induction on the number of initial steps with a kernel not in C_M and because of the fairness of the derivation, we may therefore assume that $G \leq H_j$ for some minimal $j \in I$ and that the kernel ℓ of the $j + 1$-st step of Δ is in fact a member of C_M. In accordance with Theorem 3.1, we may further assume that there is a decomposition $(G_\pi)_{\pi \in S_n}$ of G satisfying property P. Under these assumptions, the theorem is proved by induction on $\mu(G)$.

If H_j contains the empty clause (either in the subgraph G or in some other part of H_j), then the theorem trivially holds. This provides the base case of the induction. So we assume that H_j does not contain the empty clause and that $\mu(G) > 0$.

The structure of the remaining proof is now similar to the one for Theorem 3.1. It is illustrated by the right part of Figure 11. On the one hand (illustrated by the upper part of the figure), we consider the cg-resolution step upon ℓ transforming H_j into H_{j+1}. On the other hand, some of the G_π's are reduced by π-regular reduction thus reducing the measure. Thereby, the four properties, namely "being a decomposition of", "being a comp-graph" (not necessarily a minimal one), property P, and $G \leq \bigcup_{\pi \in S_n} (G_\pi) \leq (N_j, D_j)$, satisfied before this pair of transformations, are shown to be preserved for the respective graphs and also the measure is shown to be strictly reduced by the transformations so that the induction hypothesis can be applied to establish the theorem. The details of this outline follow in the remaining proof:

Let $M = (c_\xi)_{\xi\in\Gamma}$, $M_\ell = (c_\xi)_{\xi\in(\Gamma\cup\{\theta\})}$, and $\ell = \{K^\gamma, \bar{K}^\delta\}$ for $\gamma, \delta \in \Gamma$. That is, $N_{j+1} = (N_j)_\ell$ is obtained by substituting M by M_ℓ in N_j and $G_\ell = (M_\ell, (C_M)_\ell) \leq (H_j)_\ell = ((N_j)_\ell, (D_j)_\ell) = H_{j+1}$.

For any $\pi \in S_n$ such that the matrix M_π of G_π contains a clause $c_\rho \supseteq c_\gamma \cup (c_\delta \backslash \{\bar{K}\})$, and K^ρ is the π-front literal in c_ρ, let G'_π be the regular graph obtained from G_π by π-regular reduction upon the π-front connection $\{K^\rho, \bar{K}^\lambda\}$ for some $\lambda \in \Gamma_\pi$. For any other $\pi \in S_n$, let $G'_\pi = G_\pi$. Further, we let $G' = (M', C') = \bigcup_{\pi\in S_n} G'_\pi$.

We first show that $M' = M_\ell$. M is the matrix of $\bigcup_{\pi\in S_n} G_\pi$. M_ℓ differs from M by its added resolvent C_θ and by the possibly deleted one or two parent clauses (according to the convention noted at the end of Section 2.1). Exactly as in the proof of Theorem 3.1, we also see that M' also differs from M by one additional clause c resulting from π-regular reduction of pairs of clauses that subsume the parent clauses of the resolution step. Because of the second part of property P there are such π-front connections that lead to the resolvent c_θ, so that $c = c_\theta$. Since subsumption and purity reductions are included in the union of a decomposition in exactly the same way, also here one or two parent clauses are deleted exactly as for M_ℓ.

As with any decomposition, we have $C' = C_{M'}$. Since no tautological connection can be generated from resolving upon a connection from a complete mating, we have $(C_M)_\ell \subseteq C'$. There are cases where equality between these two sets does not hold as one might first expect, which explains why we prove $\leq$ rather than $=$. One such example is obtained by resolution upon the leftmost $\{X_1, \overline{X_1}\}$-connection in the matrix

$$\{\{X_1, X_2\}, \{\overline{X_1}, X_3\}, \{X_1, \overline{X_2}\}, \{\overline{X_1}, \overline{X_3}\}\},$$

which is the matrix of the minimal comp-graph ${}_{2\,.\,.}G^3_{1010} \cup {}_{3\,.\,.}G^3_{0101}$. Altogether, we have established $G_\ell \leq G'$ with $M_\ell = M'$. $G' \leq H_{j+1}$ holds by construction; thus, out of the four properties to be established, the inequality holds.

Also, we do have the desired decomposition of the graph G' which “majorizes” G_ℓ (in the sense of $G_\ell \leq G'$). Note that we allow G' to be nonminimal (as it happens in the example just mentioned). But G' is complete and, because of Theorem 4.2, also complementary (i.e., G' is a comp-graph, thus establishing also the second of the four properties). Since there is at least one π-regular reduction leading to G', we also have $\mu(G') < \mu(G)$.

Thus, in order to be able to apply the induction hypothesis, we only need to establish property P for G'. Since G' is obtained by π-regular reduction of the decomposition of G in exactly the same way as in the proof of Theorem 3.1, the arguments used there apply in exactly the same way here to establish this last point in the proof. QED.

As mentioned in the previous proof, the following obvious lemma has been used in it.

LEMMA 4.4. *If $G \leq H$ and ℓ is not a member of the mating of G, then $G \leq H_\ell$.*

Since the reduction operations of purity and subsumption are implicitly performed in the process of constructing the union of a regular family, our proof has established the following even stronger theorem.

THEOREM 4.5. *cg-resolution with purity reduction and subsumption is strongly complete.*

This means that, no matter whether and to what extent we apply purity reduction and subsumption, an attempted cg-resolution refutation, starting from a complementary matrix, will successfully terminate after a finite number of steps, provided that the connections are resolved away in a fair manner.

As the main result of this section, we have seen that any cg-resolution step (or resolution step for that matter) either does not contribute to a refutation—and is thus redundant—or it reduces the number of connections of at least one component in the decomposition of a minimal comp-subgraph of the given graph. This is our answer to the fundamental question posed in the introduction.

Note that our proof of the theorem focused on one particular minimal comp-graph G in the initial graph, which may contain further minimal comp-graphs other than G. In fact, the actual proof of the given matrix (i.e., the empty clause) may result from some of these other graphs while resolutions upon connections in G may even delay the proof's completion. The point of our argument is that any graph like G is a guarantor for eventually producing a proof unless some other part completes this job earlier.

The fairness condition involved in the theorem cannot be dispensed with, as was demonstrated in Eisinger [1989], with an example that admits an infinite derivation if fairness is not required.

5. *Strong Completeness of General cg-Resolution*

As with all connection-based proof calculi, the distinctive features of cg-resolution are propositional by nature. Any such calculus becomes a first-order one by adding unification and the generation of clause variants. In consequence of this neat separation, it is also straightforward to lift a completeness proof from the ground level to the general level which justifies this standard two-step proof technique. For the case of our strong completeness result, this lifting will be carried out in the present section.

There is an even more elegant way, due to Herbrand [1930], which allows to separate unification completely from the lifting proof.[3] Instead of performing unification at each proof step, the set of atoms involved is partitioned such that two atoms end up in the same part if and only if they become unified in the course of the proof. This way the first-order atoms can be treated as propositional variables and the additional proof efforts needed for lifting Theorem 4.3 become marginal. We begin by stating the corresponding version of Herbrand's theorem, essentially Herbrand's Property A theorem [Herbrand 1930], in a clausal version, and a definition required for it, both borrowed from Robinson [1995] (see also Robinson [1968]).

Definition 5.1. A pair $(R, \mathcal{P})$ is an *abstract refutation* of a set S of clauses if the following conditions hold.

[3] The first author owes this important aspect of the proof below to Alan Robinson.

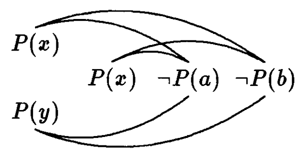

FIG. 12. The connection graph for the expanded matrix.

(1) R is a set of variants of clauses in S;
(2) $\mathcal{P}$ is a unifiable partition of the set of atoms that occur in clauses in R;
(3) every assignment of truth values to the atoms that occur in clauses in R either assigns different values to two atoms which lie in the same part of $\mathcal{P}$, or falsifies some clause in R, or both.

THEOREM 5.2. *A set of clauses is unsatisfiable iff it has an abstract refutation.*

Next, we recall cg-resolution for the general level [Kowalski 1975]. It is defined exactly like resolution (see, for example, Bibel [1993]) except that potential kernels of resolution steps, that is, pairs of literals with unifiable atoms and different signs, are handled explicitly the way specified in Definition 2.1.3. Complementarity is generalized to the first-order level in the usual way (see, for example, Section 3.3 in Bibel [1993]). It syntactically characterizes unsatisfiability as on the ground level.

We point out two important aspects that are not present at the ground level. A clause may contain different but unifiable literals which upon a resolution step merge into a single one in the resolvent by using an appropriate substitution for the unification of the kernel atoms. Whether or not such a merge is done is a strategical decision in the search for a proof. If the chosen strategy takes a wrong decision, the proof process needs to backtrack and revise it. An alternative solution to this explicit backtracking technique is an expansion of the given matrix by substituting each (given or generated) clause by all its possible *factors* (as they are called in Kowalski [1975]). Formally these notions are defined as follows.

Definition 5.3. Let c be a clause, that is, a set of (first-order) literals, and $c' \subseteq c$ a unifiable subset of its literals, that is, there exists a substitution σ and a literal L such that $c'\sigma = \{L\}$. Then $c\sigma$ is called a *factor* of c. The result of replacing each clause in a matrix by all its different factors (modulo renaming) is called the *expanded* matrix.

Any clause is also one of its factors. A simple example for illustrating both notions is the matrix $\{\{P(x), P(y)\}, \{\neg P(a)\}, \{\neg P(b)\}\}$. In general, each clause has a trivial factor which is the clause itself. In the present example only the first clause has a nontrivial factor which is $\{P(x)\}$ (and no others since $\{P(y)\}$ is identical with its modulo renaming). Hence, the connection graph consisting of the expanded matrix along with its full mating is the one shown in Figure 12.

The concept of an abstract refutation deals with this issue by partitioning the set of atoms. For our current example, one possible abstract refutation partitions the set of atoms into the two parts $\{P(x), P(y), P(a)\}$ and $\{P(b)\}$. Once the partition is determined the remaining proof may be carried out in a purely propositional way. In fact, we might even replace every atom in a part by one and the same propositional variable. In the case of our example, this replacement would lead to the formula $\{\{X\}, \{\neg X\}, \{\neg Y\}\}$. In view of this possible replacement, connections in the matrix of an abstract refutation are pairs of *sets of* literals rather than pairs of literals. For instance, with the partition in the present example, $\{\{P(x), P(y)\}, \{\neg P(a)\}\}$ is a generalized connection in this sense. Under this view, a matrix along with a partition may be seen as a purely propositional graph.

In order to maintain that different clauses used in a resolution proof are variable disjoint, all variables are renamed in the parent clauses in the course of determining the resolvent. This stipulation also serves to enable the use of a clause in different variants (which is equivalent with providing more than one variants of a given clause at the outset as done in an abstract refutation). Because of this possibility, connections may link two different literals within the same clause, which is not possible on the ground level. Resolving upon such a link means determining the resolvent by resolving upon a connection between two different variants of the clause in the usual way. Thus, a connection graph on the first-order level encodes an arbitrary number of variants of its clauses along with the corresponding copies of the connections. In consequence we generalize the notion $G \leq H$ for two connection graphs $G = (M, C)$ and $H = (N, D)$ from Definition 2.1.3 to mean M is a submatrix of N' and $C \subseteq D'$ whereby N' may contain for each clause in N an arbitrary number of variants and D' contains the corresponding copies of connections from D. With these preparations, we are now ready to prove the theorem of this section.

THEOREM 5.4. *cg-resolution is strongly complete.*

PROOF. The proof proceeds in strict analogy with the one for Theorem 4.3 to which the reader is referred for what follows: We are given an arbitrary unsatisfiable matrix represented as a complementary connection graph H with a full mating. The theorem then claims that any fair maximal derivation $\Delta = ((N_i, D_i))_{i \in I}$ (see Definition 2.2.1, which also applies to the general case) starting with H is a (finite) refutation.

Since the matrix of H is unsatisfiable, Herbrand's Theorem 5.2 provides us with an abstract refutation $(M, \mathcal{P})$. We may thereby assume that M is a *minimal* complementary matrix. Since the mating D_0 of H is full, $G = (M, C_M) \leq (N_0, D_0) = H_0 = H$ holds whereby the generalized version of $\leq$ defined above applies. So we have exactly the same situation as at the beginning of the proof for Theorem 4.3.

The main idea of the proof is to view G as a ground graph and project this "ground view" into the given derivation of H. By "grounding" the entire proof, in this way, the proof of Theorem 4.3 does in fact cover the general case. The essence of this projection has the two main aspects of generalized connections on the one hand, and of unwrapping the clause variants out of the given matrix in accordance with the assumed abstract refutation on the other, both discussed before stating the theorem. Given that this way we are actually dealing with the ground case, we may proceed literally as in the proof of Theorem 4.3, which will not be repeated here. Rather, we restrict ourselves to a few comments on important proof aspects.

Under the view of G as a propositional graph with generalized connections corresponding to propositional connections, Theorem 3.1 can be applied as is, that is, without any generalization. Hence, $\mu(G)$ is also well defined. In consequence, the two different sequences of (propositional) graphs (one starting with G, the other with its decomposition), as illustrated in Figure 11, play exactly the same role as before.

Further, in order to maintain the correspondence between H and G, and the graphs derived from them, we stipulate that the resolution steps in the given derivation use the same variable names in clause variants as the corresponding ones in G (in cases where corresponding ones exist). This is without loss of generality, since the variable names used for the variants in Herbrand's theorem as well as for generating resolvents is arbitrary anyway. With this provision, not only the correspondence between the clauses in H and G, and the graphs derived from them, but also between their connections is the same as on the ground level with one exception. Namely, the variant clauses from H become unwrapped as the derivation proceeds while those from G are already present as they are on the ground level. QED.

6. *Discussion and Summary*

In this paper, we first presented a result in which any minimal unsatisfiable matrix may be decomposed into (exponentially many) regular matrices, each of which is easy to prove. This result has provided a new technique for proving completeness of resolution refinements.

We have applied this technique to establish strong completeness of cg-resolution on the ground level. Using a characterization of unsatisfiability due to Herbrand, exactly the same proof could be used to generalize this completeness result to the first-order level. It is now guaranteed that any sequence of cg-resolution steps will eventually yield the empty clause, provided the initial formula is unsatisfiable and each (initial or inherited) connection has a finite chance to be selected as the kernel of a cg-resolution step.

If the initial formula is satisfiable, then the process may never terminate even on the ground level as demonstrated in Eisinger [1989] with a simple example mentioned at the end of Section 4. This property does not affect theorem-proving applications on the first-order level, since those proof procedures fail to be decision procedures anyway. But one might still think of some modification of cg-resolution such that it becomes a decision procedure on the ground level, an issue not further pursued here.

We feel that our result concerning the decomposition of unsatisfiable formulas might have applications other than the one presented in this paper. It might also shed new light on the unsatisfiability problem both in view of proof strategies and of complexity considerations. The proof task may now be formulated as follows: Given a matrix, find a transformation to the "closest" regular graph. As we have seen in Section 2, regular graphs may be proved in linear time. So the kernel of the theorem problem lies in the task to bridge the "gap" between general matrices and regular graphs which is a new way of viewing this problem. Let us illustrate this with the picture shown in Figure 13.

The figure shows a (full and a carved-up) globe (which is actually a graph net) each point of which represents a (minimal) unsatisfiable formula in n variables. The north pole represents the complete matrix and the south pole the matrix consisting of the empty clause. The nodes on the longitudes of the globe

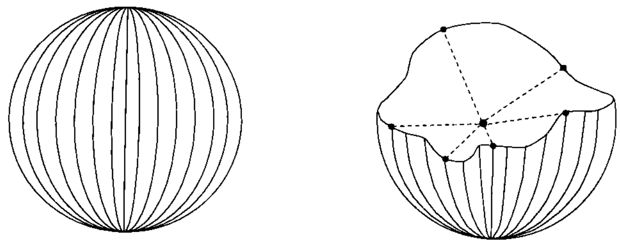

FIG. 13. An illustration of the space of unsatisfiable formulas.

correspond to the regular unsatisfiable graphs, each longitude representing a particular ordering of the variables.[4] The distance of any point on a longitude to the south pole is linear in the size of the formula, that is, these longitudes provide fast tracks indeed. These longitudes form the $n!$ axes of the coordinate system. Any point in the globe[5] has then $n!$ "coordinates" representing their regular components as illustrated in the second picture. Points which are related by the resolution operation are connected by an edge.

Refinements such as linear resolution or model elimination follow quite different strategies and, therefore, fail to prove regular graphs in linear time like π-regular reductions, which we consider a major weakness of these refinements. In fact, for some regular graphs, these refinements fare exponentially badly. They are not *adequate* in the sense of this term introduced in Bibel [1991]. With our result, there is ample room for improvement of such refinements so that they overcome this particular weakness. This cannot be achieved however in a naïve way as mentioned at the end of Section 3. A first (independent) step into such a direction was recently done with the paper [Klingenbeck and Hähnle, 1994] introducing the concept of refutation graphs that are related with the π-regular graphs in this paper. With Gabbay [1991], we share the motivation to aim at reductions (cf. our π-regular reductions which lead to "diminishing resources" in the sense of that paper).

A particularly interesting question in the context of complexity considerations is whether the addition of the inverse resolution rule, that is, sort of an extension rule, might lead to shorter proofs. Recall, from the proof of Theorem 3.1, that we used this sort of an extension rule. Speaking in terms of the picture, this amounts to the question of whether the path from some point in the globe to the south pole might be shortened by taking "upward detours". This kind of an extension rule is different from the one introduced in Tseitin [1968/1970], which indeed has such an effect, but is a stronger rule allowing for abbreviations. Our conjecture is that the answer to the question is a negative one in the case of inverse resolution.

This answer would amount to a special (and negative) answer for the P = NP question. Since we know that there are many points in the globe from which the resolution path to the south pole requires exponentially many steps [Haken 1984; Urquhart 1987; Chvátal and Szemerédi 1988], we could conclude that there are no shorter paths whatsoever for these. This answer is special since the edges in

[4] To be precise these longitudes are actually graphs not just lines.

[5] One might as well consider the space outside (rather than inside) of the globe for this illustration. The advantage would be that the distance of two graphs could be modeled more closely in terms of the metric distance in the space.

the graph are restricted to resolution (or inverse resolution) steps while other inferential steps might be possible.

It also seems now possible to explain in a coherent and plausible way why there are many examples (including the pigeonhole formulas) for which resolution fares exponentially badly as just mentioned. For them, the "gap" is bound to be exponentially large. Similarly, by studying the Horn approximation [Selman and Kautz 1993], one might consider "regular approximation." Finally, it seems possible to characterize a wider class of formulas that can be solved in polynomial time, namely exactly that class for which only polynomially many permutations are sufficient for the decomposition of the formulas.

ACKNOWLEDGMENTS. E. Eder participated in the initial phase of writing-up the proof. Thereby, he forged many of the mathematical concepts and definitions assembled mostly in Section 2. Otherwise, the paper is due to W. Bibel.

A crucial bit of the proof idea emerged during an early morning breakfast of the first author with Alan Robinson which took place at the symposium honoring the late Woody Bledsoe on occasion of his 70th birthday. Upon this author's pertinent question (stated again in the Introduction), Alan taught him the fact that some semantic tree reduces after any useful resolution step. Alan interacted again during his stay as a Humboldt awardee at the Technical University of Darmstadt and contributed the idea of using Herbrand's property A for lifting the proof for the ground case to the first-order level in a particularly elegant and simple way. We take it as an honor and as a telling symbol that these two honorable fathers of our field of automated deduction influenced the presented result in significant ways. Since circumstances prevented the first author from contributing to the Festschrift honoring Alan Robinson at the occasion of his sixtieth birthday, we take the liberty to dedicate this paper to him.

Special thanks are due to David Plaisted who has helped to improve a previous version with numerous constructive comments and questions. Thanks are also due to M. Bormann, U. Egly, S. Hölldobler, A. Pettorossi, C. Weidenbach, and anonymous referees for their helpful suggestions.

REFERENCES

ANDERSON, R., AND BLEDSOE, W. W. 1970. A linear format for resolution with merging and a new technique for establishing completeness. *J. ACM 17*, 3 (July), 525–534.

BIBEL, W., AND EDER, E. 1993. Methods and calculi for deduction. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 1, chap. 3. D. M. Gabbay, C. J. Hogger, and J. A. Robinson, eds. Oxford University Press, Oxford, England, pp. 71–193.

BIBEL, W. 1981a. On matrices with connections. *J. ACM 28*, 4 (Oct.), 633–645.

BIBEL, W. 1981b. On the completeness of connection graph resolution. In *German Workshop on Artificial Intelligence*. J. Siekman, ed. Informatik Fachberichte 47. Springer, Berlin, Germany, pp. 246–247.

BIBEL, W. 1983. Matings in matrices. *Commun. ACM 26*, 11 (Nov.), 844–852.

BIBEL, W. 1991. Perspectives on automated deduction. In *Automated Reasoning: Essays in Honor of Woody Bledsoe*. R. S. Boyer, ed. Kluwer Academic, Utrecht, Germany, pp. 77–104.

BIBEL, W. 1993. *Deduction: Automated Logic*. Academic Press, Orlando, Fla.

BROWN, F. M. 1976. Notes on chains and connection graphs. Personal notes. University of Edinburgh, Edinburgh, Scotland.

CHANG, C. L., AND LEE, R. C. T. 1973. *Symbolic Logic and Mechanical Theorem Proving*. Academic Press, Orlando, Fla.

CHVÁTAL, V., AND SZEMERÉDI, E. 1988. Many hard examples for resolution. *J. ACM 35*, 4 (Oct.), 759–768.

EISINGER, N. 1989. Completeness, Confluence, and Related Properties of Clause Graph Resolution. Ph.D. dissertation, Universität Kaiserslautern.

GABBAY, D. M. 1991. Algorithmic proof with diminishing resources. In *Proceedings of CSL'90*. E. Börger, H. Kleine Büning, M. M. Richter, and W. Schönfeld, eds. Lecture Notes in Computer Science, vol. 533. Springer, Berlin, Germany, pp. 156–173.

GOLDFARB, W. D., ED. 1971. *J. J. Herbrand—Logical writings*. Reidel, Dordrecht, Germany.

HAKEN, A. 1984. The intractability of resolution. Ph.D. dissertation. Univ. Illinois, Urbana, Ill.

HERBRAND, J. J. 1930. Recherches sur la théorie de la démonstration. *Travaux Soc. Sciences et Lettres Varsovie, Cl. 3* (*Mathem., Phys.*), pp. 128. (English translation in Goldfarb [1971].)

KOWALSKI, R., AND HAYES, P. J. 1969. Semantic trees in automated theorem proving. In *Machine Intelligence 4*, B. Meltzer and D. Michie, eds. Edinburgh University Press, Edinburgh, Scotland, pp. 87–101. (Also published in Siekmann and Wrightson [1983], pp. 217–232.)

KOWALSKI, R. 1975. A proof procedure using connection graphs. *J. ACM 22*, 4 (Oct.), 572–595.

KLINGENBECK, S., AND HÄHNLE, R. 1994. Semantic tableaux with ordering restrictions. In *Automated Deduction—CADE-12*. Alan Bundy, ed. Springer, Berlin, Germany, pp. 708–722.

NILSSON, N. J. 1980. *Principles of Artificial Intelligence*. Tioga, Palo Alto, Calif.

ROBINSON, J. A. 1965. A machine-oriented logic based on the resolution principle. *J. ACM 12*, 23–41.

ROBINSON, J. A. 1968. The generalized resolution principle. In *Machine Intelligence 3*, Donald Michie, ed. Edinburgh University Press, Edinburgh, Scotland, pp. 77–93. (Also published in Siekmann and Wrightson [1983], pp. 135–151.)

ROBINSON, J. A. 1995. A sketch of the Property A view of things. Technical Univ., Darmstadt, Germany.

SELMAN, B., AND KAUTZ, H. 1993. Domain-independent extensions to gsat: Solving large structured satisfiability problems. In *IJCAI-93—Proceedings of the 13th International Joint Conference on Artificial Intelligence*. Ruzena Bajcsy, ed. Morgan-Kaufmann, pp. 290–295.

SIEKMANN, J., AND WRIGHTSON, G., EDS. 1983. *Automation of Reasoning—Classical Papers on Computational Logic 1967–1970*, vol. 2. Springer, Berlin, Germany.

SMOLKA, G. 1982. Completeness of the connection graph proof procedure for unit refutable clause sets. In *Proceedings of GWAI-82*. Informatik Fachberichte, vol. 58. Springer-Verlag, Berlin, Germany, pp. 191–204.

TSEITIN, G. S. 1968/1970. On the complexity of derivation in propositional calculus. In *Studies in Constructive Mathematics and Mathematical Logic, Part II*. A. O. Slisenko, ed. Seminars in Mathematics, V. A. Steklov Mathematical Institute, vol. 8. Leningrad, Russia, pp. 234–259. (English translation: Consultants Bureau, New York, 1970, pp. 115–125.) (Also published in Siekmann and Wrightson [1983], pp. 466–483.)

URQUHART, A. 1987. Hard examples for resolution. *J. ACM 34*, 1 (Jan.), 209–219.

RECEIVED NOVEMBER 1994; REVISED SEPTEMBER 1996; ACCEPTED NOVEMBER 1996