THE UNIVERSITY *of* MANCHESTER

July 2002

Computer Science
University of Manchester

A New Clausal Class Decidable by Hyperresolution

Lilia Georgieva, Ullrich Hustadt, and Renate A. Schmidt

Department of Computer Science
University of Manchester
Preprint Series
CSPP-18

A New Clausal Class Decidable by Hyperresolution

Lilia Georgieva

Department of Computer Science, University of Manchester, United Kingdom
Max-Planck-Institut für Informatik, Saarbrücken, Germany
georgiel@cs.man.ac.uk

Ullrich Hustadt

Department of Computer Science, University of Liverpool, United Kingdom
U.Hustadt@csc.liv.ac.uk

Renate A. Schmidt

Department of Computer Science, University of Manchester, United Kingdom
Max-Planck-Institut für Informatik, Saarbrücken, Germany
schmidt@cs.man.ac.uk

July 2002

Abstract

In this paper we define a new clausal class, called $\mathcal{BU}$, which can be decided by hyperresolution with splitting. We also consider the model generation problem for $\mathcal{BU}$ and show that hyperresolution plus splitting can also be used as a Herbrand model generation procedure for $\mathcal{BU}$ and, furthermore, that the addition of a local minimality test allows us to generate only minimal Herbrand models for clause sets in $\mathcal{BU}$. In addition, we investigate the relationship of $\mathcal{BU}$ to other solvable classes.

1 Introduction

In recent work [13, 14] we have considered the fragment GF1$^-$ of first-order logic which was introduced by Lutz, Sattler, and Tobies [22]. GF1$^-$ is a restriction of the guarded fragment which incorporates a variety of modal and description logics via standard or non-standard translations, and can be seen as a generalisation of these logics. In contrast to the guarded fragment [1], GF1$^-$ allows for the development of a space-efficient decision procedure. Under the assumption that either (i) there is a bound on the arity of predicate symbols in GF1$^-$ formulae, or (ii) that each subformula of a GF1$^-$ formula has a bounded number of free variables, the satisfiability problem of GF1$^-$ is PSPACE-complete [22], while under identical assumptions the satisfiability problem of the guarded fragment is EXPTIME-complete [16]. Thus, GF1$^-$ has the same complexity as the modal and description logics it generalises.

In [13] we have shown that hyperresolution plus splitting provides a decision procedure for GF1$^-$. One of the interesting features of GF1$^-$ is that it is one of the few solvable classes where, during the deduction by the resolution decision procedure, derived clauses can contain terms of greater depth than the clauses in the initial set of clauses. In [14] we have shown that a modification of the main procedure of a standard saturation based theorem prover with splitting can provide a polynomial space decision procedure for GF1$^-$. We also describe several solutions to the problem of generating minimal Herbrand models for GF1$^-$.

In [13, 14] we have used structural transformation (or definitional form transformation, cf. e.g. [3, 19]), to transform GF1$^-$ formulae into clausal form. While it is straightforward to give a schematic characterisation of the resulting sets of clauses, it is much more difficult to state the conditions which an arbitrary set of clauses needs to satisfy so that it shares most or all the properties of the clause sets we obtain from the definitional form transformation of GF1$^-$ formulae.

In this paper we define a new clausal class $\mathcal{BU}$ which generalises the set of all clause sets we can obtain from GF1$^-$ via the definitional form transformation. $\mathcal{BU}$ is defined such that hyperresolution plus splitting is still a decision procedure. Since hyperresolution is implemented in many state-of-the-art theorem provers, e.g. Otter, SPASS, and Vampire, this gives a practical decision procedure for the class. We also show that if an input clause set from $\mathcal{BU}$ is not refuted, an adequate representation of a model and of a minimal model of the clausal class can be extracted from the information produced by the prover.

A main motivation for studying classes like GF1$^-$ and $\mathcal{BU}$ is that a variety of expressive modal and description logics can be embedded into them. Expressive modal and description logics have found applications in such varied areas as, for example, verification, program analysis, knowledge representation, deductive data bases and the semantic web. However, there are a number of alternative solvable classes for which the same is true. We will discuss the relationship of $\mathcal{BU}$ to some of these alternative classes.

The paper is organised as follows. Section 2 defines the notation used, some basic concepts and a hyperresolution calculus with splitting. The clausal class $\mathcal{BU}$ is defined in Section 3, and the relationship of $\mathcal{BU}$ to other solvable classes is discussed in Section 4. The applicability of the hyperresolution calculus as a decision procedure for the class, model building by hyperresolution and in particular, minimal model building are investigated in Sections 5 and 6. The final section is the Conclusion.

2 Fundamentals and hyperresolution

Notation. The notational convention is as follows. We use the symbols x, y, z for first-order variables, s, t, u for terms, a, b for constants, f, g, h for functions, P, Q, R for predicate symbols, A for atoms, L for literals, C for clauses, φ, ϕ, ψ, for formulae, and N for sets of clauses.

An over-line indicates a sequence. An *i-sequence* is a sequence with i elements. If $\overline{s}$ and $\overline{t}$ are two sequences of terms and X is a set of terms, then the notation $\overline{s} \subseteq \overline{t}$ ($\overline{s} \subseteq X$) means that every term in $\overline{s}$ also occurs in $\overline{t}$ (X). By definition, $\overline{s} = \overline{t}$ ($\overline{s} = X$) iff $\overline{s} \subseteq \overline{t}$ and $\overline{t} \subseteq \overline{s}$ ($\overline{s} \subseteq X$ and every term in X occurs in $\overline{s}$). The union of the terms in $\overline{s}$ and $\overline{t}$ is denoted by $\overline{s} \cup \overline{t}$. Given a sequence $\overline{s}$ of terms, $\overline{f_{\overline{s}}(\overline{u}_k)}$ denotes a sequence of terms of the form$f_1(\overline{u}_1), \dots, f_k(\overline{u}_k)$, where $\overline{u}_i \subseteq \overline{s}$ for every $1 \leq i \leq k$.

Terms, literals, clauses and orderings. The term depth $dp(t)$ of a term t, is inductively defined as follows. (i) If t is a variable or a constant, then $dp(t) = 1$. (ii) If $t = f(t_1, \dots, t_n)$, then $dp(t) = 1 + max\{ dp(t_i) \mid 1 \leq i \leq n\}$. The term depth of a literal is defined by the maximal term depth of its arguments, and the term depth of a clause is defined by the maximal term-depth of all literals in it.

A *literal* is an atomic formula A (a *positive* literal) or the negation $\neg A$ of an atomic formula A (a *negative* literal). We regard a *clause* as a multiset of literals and consider two clauses C and D to be identical if C can be obtained from D by variable renaming. A *multiset* over a set $\mathcal{L}$ is a mapping C from $\mathcal{L}$ to the natural numbers. We write $L \in C$ if $C(L) > 0$ for a literal L. We use $\perp$ to denote the empty clause. The *set of variables* of a clause C is denoted by $\text{var}(C)$. A *positive* (*negative*) clause contains only positive (negative) literals. The *positive* (*negative*) *part* of a clause is the subclause of all positive (negative) literals. A *split component* of a clause $C \vee D$ is a subclause C such that C and D do not have any variables in common, i.e. are *variable disjoint.* A *maximally split* (or variable indecomposable) clause cannot be partitioned (or split) into subclauses which do not share variables.

A clause C is said to be *range restricted* iff the set of variables of the positive part of C is a subset of the set of variables of the negative part of C. A clause set is range restricted iff it contains only range restricted clauses. This means that a positive clause is range restricted only if it is a ground clause.

A strict partial ordering $\succ$ on a set $\mathcal{L}$ (i.e. an irreflexive and transitive relation) can be extended to an ordering $\succ^{mul}$ on (finite) multisets over $\mathcal{L}$ as follows: $C \succ^{mul} D$ if (i) $C \neq D$ and (ii) whenever $D(x) > C(x)$ then $C(y) > D(y)$, for some $y \succ x$. $\succ^{mul}$ is called the *multiset extension of* $\succ$.

Given an ordering $\succ$ on literals we define a maximal literal in a clause in the standard way: A literal L in a clause C is *maximal* in C, if there is no literal L' in C, for which $L' \succ L$. A literal L is *strictly maximal* in C if it is the only maximal literal in C.

A term, an atom, a literal or a clause is called *functional* if it contains a constant or a function symbol, and *non-functional*, otherwise.

A hyperresolution calculus with splitting. The calculus $\mathsf{R}^{\mathsf{hyp}}$ we focus on in the following consists of two expansion rules:

Deduce: $\dfrac{N}{N \cup \{C\}}$
where C is a resolvent or a factor.

Splitting: $\dfrac{N \cup \{C_1 \vee C_2\}}{N \cup \{C_1\} \quad | \quad N \cup \{C_2\}}$
where C_1 and C_2 are variable disjoint.

The resolution and factoring inference rules are:

Hyperresolution: $$\frac{C_1 \vee A_1 \quad \ldots \quad C_n \vee A_n \quad \neg A_{n+1} \vee \ldots \vee \neg A_{2n} \vee D}{(C_1 \vee \ldots \vee C_n \vee D)\sigma}$$

where (i) σ is the most general unifier such that $A_i\sigma = A_{n+i}\sigma$ for every i, $1 \leq i \leq n$, and (ii) $C_i \vee A_i$ and D are positive clauses, for every i, $1 \leq i \leq n$. The rightmost premise in the rule is referred to as the *negative* premise and all other premises are referred to as *positive* premises.

Factoring: $$\frac{C \vee A_1 \vee A_2}{(C \vee A_1)\sigma}$$

where σ is the most general unifier of A_1 and A_2.

A *derivation* in $\mathsf{R}^{\mathsf{hyp}}$ from a set of clauses N is a finitely branching, ordered tree T with root N and nodes which are sets of clauses. The tree is constructed by applications of the expansion rules to the leaves. We assume that no hyperresolution or factoring inference is computed twice on the same branch of the derivation. Any path $N(= N_0), N_1, \ldots$ in a derivation T is called a *closed branch* in T iff the clause set $\bigcup_{j\geq 0} N_j$ contains the empty clause, otherwise it is called an *open branch*. We call a branch B in a derivation tree *complete* (with respect to $\mathsf{R}^{\mathsf{hyp}}$) iff no new successor nodes can be added to the endpoint of B by $\mathsf{R}^{\mathsf{hyp}}$, otherwise it is called an *incomplete branch*. A derivation T is a *refutation* iff every path $N(= N_0), N_1, \ldots$ in it is a closed branch, otherwise it is called an *open derivation*.

In general, the calculus $\mathsf{R}^{\mathsf{hyp}}$ can be enhanced with standard simplification rules such as tautology deletion and subsumption deletion, in fact, it can be enhanced by any simplification rules which are compatible with a general notion of redundancy [4, 5]. A set N of clauses is *saturated up to redundancy* with respect to a particular refinement of resolution if the conclusion of every inference from non-redundant premises in N is either contained in N, or else is redundant in N. A derivation T from N is called *fair* if for any path $N(= N_0), N_1, \ldots$ in T, with *limit* $N_\infty = \bigcup_{j\geq 0} \bigcap_{k\geq j} N_k$, it is the case that each clause C which can be deduced from non-redundant premises in N_∞ is contained in some N_j. Intuitively, fairness means that no non-redundant inferences are delayed indefinitely. For a finite complete branch $N(= N_0), N_1, \ldots N_n$, the limit N_∞ is equal to N_n.

Theorem 2.1 ([5]). *Let T be a fair $\mathsf{R}^{\mathsf{hyp}}$ derivation from a set N of clauses. Then: (i) If $N(=N_0), N_1, \ldots$ is a path with limit N_∞, then N_∞ is saturated (up to redundancy). (ii) N is satisfiable if and only if there exists a path in T with limit N_∞ such that N_∞ is satisfiable. (iii) N is unsatisfiable if and only if for every path $N(=N_0), N_1, \ldots$ the clause set $\bigcup_{j\geq 0} N_j$ contains the empty clause.*

3 The clausal class $\mathcal{BU}$

The language of $\mathcal{BU}$ is that of first-order clausal logic. Additionally, each predicate symbol P is uniquely associated with a pair (i,j) of non-negative integers, such that if the arity of P is n then $i+j=n$. The pair is called the *grouping* of the predicate symbol. Sometimes the grouping (i,j) of a predicate symbol P is made explicit by writing $P^{(i,j)}$. The notion of grouping is extended to literals in the following way. A literal L is said to satisfy the *grouping condition with respect to the sequences* $\overline{x}$ *and* $\overline{s}$, if $L = (\neg)P^{(i,j)}(\overline{x}, \overline{s})$ or $L = (\neg)P^{(j,i)}(\overline{s}, \overline{x})$, where $\overline{x}$ is an i-sequence of variables, and $\overline{s}$ is either a j-sequence of variables disjoint from $\overline{x}$ or a j-sequence of functional terms, whose arity is greater than zero, of the form $f(\overline{z})$ where $\overline{z} \subseteq \overline{x}$ and $\overline{z}$ is non-empty. Repetitions of variables and terms in any of the sequences are allowed.

Furthermore, an acyclic relation $\succ_d$, called an *acyclic dependency relation*, is defined over the predicate symbols. Let $\succ_d^+$ be the transitive closure of $\succ_d$. Then $\succ_d^+$ is an ordering on predicate symbols. This ordering extends to atoms, literals and clauses by the following definitions. Given two literals $L_1 = (\neg)P_1(\overline{s})$ and $L_2 = (\neg)P_2(\overline{t})$, $L_1 \succ_D L_2$ iff $P_1 \succ_d^+ P_2$. The multiset extension of the ordering $\succ_D$, also denoted by $\succ_D$, defines an ordering on ground clauses. The acyclicity of $\succ_d$ implies that $\succ_D$ is also acyclic.

Given a finite signature Σ such that (i) any predicate symbol has a unique grouping, and (ii) there is an acyclic dependency relationship $\succ_d$ on the predicate symbols in Σ, we define the *class* $\mathcal{BU}$ of clausal sets over Σ as follows.

A clausal set N belongs to $\mathcal{BU}$ if any clause C in N satisfies one of the three conditions below as well as the following. If C is a non-ground and non-positive clause then C is required to contain a strictly $\succ_D$-maximal literal, which is negative and non-functional. This literal is called the *main literal* of the clause. The predicate symbol P of the main literal must either have the grouping $(0,i)$ or $(i,0)$, where i is the arity of P.

Condition 1: C is a non-positive, non-ground and non-functional clause and the following is true.

(a) The union of the variables of the negative part can be partitioned into two disjoint subsets X and Y, at least one of which is non-empty.

(b) For every literal L in C, either the variables of L are (i) subsets of X, or (ii) subsets of Y, or (iii) there are non-empty sequences $\overline{x}$, $\overline{y}$, such that $\overline{x} \subseteq X$, $\overline{y} \subseteq Y$ and L satisfies the grouping condition with respect to $\overline{x}$ and $\overline{y}$.

(c) Either the main literal contains all the variables of the clause, or it contains all the variables from one of the sets X and Y, and there is a negative literal

L whose arguments satisfy (b.iii) and which contains all the variables from Y if the main literal contains all the variables from X, or all the variables from X if the main literal contains all the variables from Y.

Condition 2: C is a non-positive and non-ground functional clause and the following is true.

(a) The main literal of C contains all the variables of C.

(b) For every other literal L in C one of the following is true. (i) L satisfies the grouping condition with respect to two disjoint sequences of variables $\overline{x}$ and $\overline{y}$. (ii) L satisfies the grouping condition with respect to two sequences $\overline{x}$ and $\overline{f_{\overline{x}}(\overline{u}_k)}$, where $\overline{x}$ is a sequence of variables and $\overline{f_{\overline{x}}(\overline{u}_k)}$ is a sequence of terms $f_i(\overline{u}_i)$ such that $\overline{u}_i \subseteq \overline{x}$, and $\overline{u}_i$ is not empty. (iii) The arguments of L are a sequence of functional terms of the form $\overline{f_{\overline{x}}(\overline{u}_i)}$, where $\overline{u}_i \subseteq \overline{x}$, $\overline{u}_i$ is not empty and $\overline{x}$ is the sequence of the arguments of the main literal of the clause.

Condition 3: C is a positive ground unit clause, its arguments are constants and its predicate symbol has grouping $(0, i)$ or $(i, 0)$.

From the definition of $\mathcal{BU}$ it follows that all non-positive clauses are range-restricted. However, in contrast to clausal classes induced by non-classical logics [10] or guarded formulae [12], the presence of a single negative literal which contains all the variables, i.e. a *guard* literal in the clause, is not required. In $\mathcal{BU}$ the guard can be either a single literal, or a pair of negative literals which together contain all the variables of the clause (provided that the grouping restrictions and the ordering conditions are satisfied).

Consider the following clauses.

1. $\neg Q_2(x,y) \vee \neg R_1(y,z) \vee P(z,z)$
2. $\neg P(x,y) \vee \neg Q_3(x) \vee \neg R_2(x,x,y,z)$
3. $\neg P(x,y) \vee \neg Q_2(y,z) \vee \neg R_2(x,x,y,z)$
4. $\neg P(x,y) \vee \neg Q_2(y,z) \vee \neg R_2(x,y,z,x)$
5. $\neg P(x,y) \vee Q_1(x,x,y,f(x,y))$
6. $\neg P(x,y) \vee Q_1(x,x,y,g(y))$
7. $\neg P(x,y) \vee Q_1(x,x,g(y),y)$
8. $\neg P_2(x) \vee P_2(f(x))$

In the case where a clause in a $\mathcal{BU}$ clause set satisfies Condition 1 and there is no single guard literal, then there is a negative literal which satisfies Condition 1.(b.iii) which together with the $\succ_D$-maximal literal contains all the variables of the clause. An example of such a clause is Clause 1. Assume that R_1 is $\succ_d^+$-maximal among the predicate symbols in Clause 1, R_1 has the grouping $(2,0)$ or $(0,2)$, and Q_2 has the grouping $(1,1)$. Then $\neg R_1(y,z)$ is the main literal of Clause 1. The literal $\neg Q_2(x,y)$ satisfies Condition 1.(b.iii) and together $\neg R_1(y,z)$ and $\neg Q_2(x,y)$ contain all the variables of Clause 1. Alternatively, we could assume that Q_2 is the $\succ_d^+$-maximal predicate symbol among the predicate symbols in Clause 1, Q_2 has the grouping $(2,0)$ or $(0,2)$, and R_1 has the grouping $(1,1)$. Now, $\neg Q_2(x,y)$ is the main literal of Clause 1 and $\neg R_1(y,z)$ is the literal satisfying Condition 1.(b.iii) which together with $\neg Q_2(x,y)$ contains all variables of Clause 1. Note that the clause is not loosely guarded, since there is no negative literal containing the variables x and z.

Consider Clause 2. Whether this clause could occur in a $\mathcal{BU}$ clause set depends on the acyclic dependency relation $\succ_d$. If Q_3 is the $\succ_d^+$-maximal predicate symbol in Clause 2, then $\neg Q_3(x)$ is the maximal literal with respect to $\succ_D$ and therefore the main literal of Clause 2. But then there is no negative literal which together with $\neg Q_3(x)$ contains all the variables of the clause, so that the grouping restrictions are satisfied. Thus, under this dependency relation Clause 2 does not belong to any $\mathcal{BU}$ clause set. The same is true if P is the $\succ_d^+$-maximal predicate symbol in Clause 2 which would make $\neg P(x, y)$ the main literal of the clause. For Clause 2 to occur in a $\mathcal{BU}$ clause set, the literal $\neg R_2(x, x, y, z)$, must be the main literal of the clause and to this end, the predicate symbol R_2 has to be $\succ_d^+$-maximal with grouping $(4, 0)$ or $(0, 4)$. The grouping of Q_3 and P is immaterial as well as their relative precedence with respect to $\succ_d$.

But observe that in Clause 3, if $\neg P(x, y)$ is the main literal then the grouping condition must hold for $\neg Q_2(y, z)$ and $\neg R_2(x, x, y, z)$, i.e. the grouping of Q_2 and R_2 must be $(1, 1)$ and $(3, 1)$, respectively in order for the clause to belong to a $\mathcal{BU}$ clause set.

In Clause 4, if $\neg P(x, y)$ is the main literal then the sequence (x, y, z, x) cannot be divided into disjoint non-empty subsequences, because the variable x would appear in both of the subsequences and consequently the clause does not belong to any $\mathcal{BU}$ clause set. Clauses 5 and 6 are examples of clauses which satisfy Condition 2, provided P has grouping $(2, 0)$ or $(0, 2)$, Q_1 has grouping $(3, 1)$, and $P \succ_d Q_1$. Clause 7 on the other hand violates Condition 2 because the $Q_1(x, x, g(y), y)$ literal does not satisfy the grouping condition, regardless of what the grouping of Q_1 may be. Clause 8 violates Condition 2, because it does not contain a strictly maximal negative literal, with respect to any acyclic dependency relation. In general, the acyclicity requirement excludes clauses where the predicate symbol of the main literal occurs both positively and negatively. Thus the transitivity clause and the symmetry clause do not belong to any clause set in $\mathcal{BU}$. Also the reflexivity clause and the seriality clause are excluded from $\mathcal{BU}$ clause sets, because every non-ground clause must contain a negative main literal. Thus, other than irreflexivity $\neg R(x, x)$, none of the standard properties of relations except forms of relational inclusion (e.g. $\neg R_1(x, y) \vee R_2(y, x)$) can be formulated in $\mathcal{BU}$. We comment on this 'apparent' limitation in the next section.

4 Relationships to other solvable classes

One of the main motivations for studying $\mathcal{BU}$ is that a variety of modal and description logics can be embedded into it. Simple examples are the basic multi-modal logic $\mathsf{K}_{(m)}$ and the corresponding description logic $\mathcal{ALC}$ [25]. For example, if we translate formulae of $\mathsf{K}_{(m)}$ into first-order logic and transform the resulting formulae into clausal form using structural transformation, then the clauses we obtain take one of the following forms [10, 18, 19].

$$
\begin{array}{lll}
Q_1(a) & \neg Q_1(x) \vee Q_2(x) & \neg Q_1(x) \vee Q_2(x) \vee Q_3(x) \\
\neg Q_1(x) \vee \neg R(x, y) \vee Q_2(y) & \neg Q_1(x) \vee R(x, f(x)) & \neg Q_1(x) \vee Q_2(f(x))
\end{array}
$$

Furthermore, we can always define an acyclic dependency relation on the predicate symbols in these clauses and associate groupings $(0,1)$ and $(1,1)$ with every unary and binary predicate symbol, respectively, such that the clause set satisfies the conditions for clause sets in $\mathcal{BU}$. Much more expressive logics like the multi-modal logic $K_{(m)}(\cap, \cup, \smile)$ which is defined over families of relations closed under intersection, union, and converse, and the corresponding extension of $\mathcal{ALC}$ can also be embedded into $\mathcal{BU}$.

We have already mentioned in the previous section that clauses expressing most of the standard properties of binary relations like reflexivity, seriality, symmetry, and transitivity cannot occur in $\mathcal{BU}$ clause sets. Consequently, the standard translation of formulae in modal logics extending $\mathsf{K}_{(m)}$ by one or more of the axiom schemata T (reflexivity), D (seriality), B (symmetry) and $\mathsf{4}$ (transitivity) do not result in $\mathcal{BU}$ clause sets. However, for the extensions of $\mathsf{K}_{(m)}$ by any combination of the axiom schemata T, D, and B, a non-standard translation proposed by De Nivelle [9] exists which together with structural transformation allows us to translate formulae of these modal logics into $\mathcal{BU}$ in a satisfiability equivalence preserving way. Although this non-standard translation also provides an alternative approach for the modal logic $\mathsf{K4}_{(m)}$, the resulting clause sets are still not in $\mathcal{BU}$, since it is in general impossible to define the required acyclic dependency relationship. This negative result is not surprising, since tableau decision procedures for $\mathsf{K4}_{(m)}$ require an auxiliary loop checking mechanism besides the tableau expansion rules to ensure termination.

Another example of a reasoning problem in description logics that can be solved by embedding into the class $\mathcal{BU}$ is the satisfiability problem of $\mathcal{ALC}$ concepts with respect to acyclic TBoxes. This problem has recently been shown to be PSPACE-complete [21]. Here the acyclicity of a TBox $\mathcal{T}$ allows us to define an acyclic dependency relation on predicate symbols occurring in the translation of $\mathcal{T}$ such that the conditions for $\mathcal{BU}$ clause sets are satisfied. Note that the standard translation of $\mathcal{T}$, which contains closed first-order formulae, is not in GF1^{-}.

There are a number of other fragments of first-order logic and clausal classes which would cover the same modal and description logics, including the guarded fragment [1], the dual of Maslov's class K [17], and fluted logic [24]. The clausal classes corresponding to the guarded fragment [12] and the dual of Maslov's class K contain only clause sets where every non-constant functional term t contains all the variables of the clause C in which it occurs. Clause 6 on page 5 illustrates that this is not the case for $\mathcal{BU}$ clause sets. Fluted logic requires a certain ordering on variable occurrences which means that a clause like $\neg R(x,y) \vee Q(y,x)$ is not fluted, but could occur in a $\mathcal{BU}$ clause set. On the other hand, we can also give examples for each of these three classes showing that $\mathcal{BU}$ subsumes neither of them. Thus, all four classes are distinct from each other. However, $\mathcal{BU}$ is the only class among them for which a hyperresolution decision procedure is known.

Other syntactically defined clausal classes which are also decidable by hyperresolution include the classes $\mathcal{PVD}$ and $\mathcal{KPOD}$ [11, 20]. For $\mathcal{PVD}$ the syntactic restrictions on the class imply that during a derivation by hyperresolution the depth of a derived clause does not exceed the depth of the corresponding positive premises. An example of a clause set which is in $\mathcal{BU}$ but not in $\mathcal{PVD}$ is $\{\neg Q(x) \vee \neg R(x,y), \neg P(x) \vee R(x,f(x))\}$, while $\{\neg R(x, y) \vee R(y, x)\}$ is an example of a $\mathcal{PVD}$ clause set which is not in $\mathcal{BU}$. For $\mathcal{KPOD}$, like $\mathcal{BU}$, the term depth of derived clauses can increase during the derivation. Essential for $\mathcal{KPOD}$ is the restriction of clauses to Krom form ($|C| \leq 2$), while $\mathcal{BU}$ has no restriction on the number of literals in a clause. On the other hand, $\mathcal{KPOD}$ does not require an acyclic dependency relation on predicate symbols or any grouping restriction. Therefore, $\mathcal{BU}$, $\mathcal{PVD}$, and $\mathcal{KPOD}$ are all distinct from each other.

There are syntactically defined clausal classes which cannot be decided by hyperresolution. One such class is $\mathcal{VAR}1$ [20]. A finite clausal set N belongs to the clausal class $\mathcal{VAR}1$ if and only if it contains clauses with at most one variable. Clearly, there are clause sets in $\mathcal{VAR}1$ which are also in $\mathcal{BU}$, provided that all conditions from the definition of $\mathcal{BU}$ are satisfied and the clause set contains only one variable. A suitable example is the clause set $\{P(a, b),\ \neg P(x, x) \vee Q(x, x, x, f(x, x))\}$. However, there are also clause sets which are in $\mathcal{VAR}1$ but not in $\mathcal{BU}$ and vice versa. Any clause set containing clauses with more than one variable is not in $\mathcal{VAR}1$. On the other hand, for example, any clause set containing only clauses with one variable, but at least one clause where a predicate symbol occurs in a positive as well as in a negative literal, is in $\mathcal{VAR}1$, but not in $\mathcal{BU}$. The class $\mathcal{VAR}1$ can be decided by ordered resolution [11, 20]. However, hyperresolution does not terminate on $\mathcal{VAR}1$.

5 Deciding $\mathcal{BU}$

To decide $\mathcal{BU}$ we use the calculus $\mathsf{R}^{\mathsf{hyp}}$, described in Section 2, which consists of hyperresolution, factoring, and splitting (though factoring is optional). We assume in the following that a hyperresolution inference cannot use a clause C as a positive premise if the splitting rule or, if present, the factoring rule can be applied to C. As usual we make a minimal assumption that no inference rule is applied twice to the same set of premises during a derivation.

For the classes of clause sets we consider in the present paper the positive premises are always ground, in particular, because we use splitting, the positive premises are always ground *unit* clauses, and the conclusions are always positive ground clauses. Crucial for termination is that the unit clauses are always either *uni-node* or *bi-node*. These notions are adapted and extended from similar notions in [22] and [13, 14].

A sequence $\overline{t} = (t_1, \ldots, t_n)$ (or multiset $\{t_1, \ldots, t_n\}$) of ground terms is called a *uni-node* iff all terms in the sequence (or multiset) have the same depth, that is, $dp(t_i) = dp(t_j)$ for every $1 \leq i, j \leq n$. If $\overline{t}$ and $\overline{s}$ are uni-nodes and $\overline{t} \subseteq \overline{s}$, we say $\overline{t}$ is a *uni-node defined over* $\overline{s}$. A sequence $\overline{t} = (t_1, \ldots, t_m)$ (or multiset) is called a *direct successor* of a sequence $\overline{s} = (s_1, \ldots, s_n)$ (or multiset) iff for each t_i, $1 \leq i \leq m$, there is a function symbol f_i such that t_i is of the form $f_i(\overline{u}_i)$, where $\overline{u}_i \subseteq \overline{s}$, and $\overline{u}_i$ is non-empty. A sequence (or multiset) of ground terms is called a *bi-node* (*over* $\{X_1, X_2\}$) iff it can be presented as a union $X_1 \cup X_2$ of two non-empty disjoint uni-nodes X_1 and X_2 such that X_2 is a direct successor of X_1.

A ground literal (unit clause) is a *uni-node* iff the set of its arguments is a uni-node. The empty clause $\bot$ is a special type of uni-node literal (with no direct successors). A ground literal L (unit clause) is a *bi-node* iff the set of its arguments is a bi-node over $\{\overline{s}, \overline{t}\}$ and has the form $L = (\neg)P^{(i,j)}(\overline{s}, \overline{t})$, where $\overline{s}$ is an i-sequence and $\overline{t}$ is a j-sequence of terms. If the latter is true we say L satisfies the *grouping condition with respect to $\overline{s}$ and $\overline{t}$* (this extends the definition in Section 3 to ground literals). Subsequently, when we write $(\neg)P^{(i,j)}(\overline{s}, \overline{t})$ we mean that this literal satisfies the grouping condition with respect to $\overline{s}$ and $\overline{t}$.

The following table gives examples of uni-nodes and bi-nodes.

Uni-nodes:	$\{a, a, b\}$, $\{g(a, b)\}$, $\{g(a, b), f(b, b)\}$
Bi-nodes:	$\{a, b, f(b)\}$, $\{a, b, g(a, b), h(a, b, b)\}$, $\{a, b, f(b), h(b, a, b)\}$

The notions of uni-node and direct successor are more general than the notions defined in [13, 14, 22]. For example, $\{a, b, f(b)\}$ is not a bi-node (nor a uni-node) under the previous definitions. The set $\{a, f(a, b)\}$ is not a bi-node (or a uni-node) under either definitions.

In the rest of the section, assume N is a given (finite) clausal set in $\mathcal{BU}$. The aim is to show that any derivation from N by $\mathsf{R}^{\mathsf{hyp}}$ terminates. The following properties are characteristic about hyperresolution inferences for the class $\mathcal{BU}$:

1. All conclusions are ground.

2. Each of the split components of the derived clauses are ground unit clauses which are either uni-nodes or bi-nodes.

3. Each of the ground unit clauses used as a positive premise produces a bounded number of different conclusions.

These properties are key to the termination proof given below. The first property is easy to see, since any $\mathcal{BU}$ clause set is range restricted and if all positive premises of hyperresolution inference steps are ground, and all non-ground clauses are range restricted, the conclusion of any inference step by $\mathsf{R}^{\mathsf{hyp}}$ is either the empty clause, or a positive ground unit clause, or a positive ground clause which can be split into positive ground unit clauses. The second property is established in Lemma 5.2.2. (This property is the reason for choosing the name $\mathcal{BU}$ for the considered clausal class.) The third property is a consequence of Lemma 5.1.

As factoring is applied only to positive clauses, and positive clauses in any $\mathsf{R}^{\mathsf{hyp}}$ derivation from a $\mathcal{BU}$ clause set are always ground, factoring has the effect of eliminating duplicate literals in ground clauses. For this reason no special consideration is given to factoring inference steps in subsequent proofs.

Given a finite signature the following can be proved using the same argument as in the corresponding lemma for $\mathrm{GF1}^-$ in [13].

Lemma 5.1. *1. The cardinality of any uni-node set is finitely bounded.*

2. Every uni-node has a bounded number of direct successors which are uni-nodes.

3. *For any given uni-node $\overline{s}$, the number of the uni-nodes and bi-nodes that have terms in $\overline{s}$ as elements is finitely bounded.*

Lemma 5.2. *In any $\mathsf{R}^{\mathsf{hyp}}$ derivation from a clause set in $\mathcal{BU}$:*

1. *At least one of the positive premises of any hyperresolution inference step is a uni-node.*

2. *Maximally split conclusions are either uni-nodes or bi-nodes.*

3. *If $P(\overline{s}, \overline{t})$ is a bi-node over $\{\overline{s}, \overline{t}\}$ and occurs in the derivation and $\overline{t}$ is a direct successor of $\overline{s}$, then all terms in $\overline{s}$ have the same depth d and all terms in $\overline{t}$ have the same depth d+1.*

Proof. The proof is by induction. In the first step of any $\mathsf{R}^{\mathsf{hyp}}$ derivation the only possible positive premises are uni-nodes. Since all their arguments are constants, they have the same depth. The induction hypothesis is that the above properties are true for the premises and conclusions of the first n inference steps in any derivation.

Now consider the different inference possibilities in step $n+1$. First, a general observation. The grouping restriction on the main literal of any non-positive clause in a $\mathcal{BU}$ clause set implies that the premise associated with the main literal, generally, we call it the *main premise*, must be a unit clause whose literal has grouping $(0, i)$ or $(i, 0)$. This means the main premise is a uni-node. This proves property 1.

Consider an inference step by hyperresolution involving a non-positive clause C satisfying Condition 1, as negative premise. Assume the main premise is a uni-node of the form $Q(\overline{s})$. If the main literal contains all the variables of the clause then each variable in C is unified with a term from $\overline{s}$. It follows immediately that all other premises and all maximally split conclusions are uni-nodes and the depths of all arguments are the same, since by the induction hypothesis the depth of all arguments in $Q(\overline{s})$ are the same.

Let X and Y be as in Condition 1.(a). Assume the main literal does not contain all the variables of C, instead it contains all the variables from $X (\neq \emptyset)$. Then C has a negative literal that satisfies 1.(b.iii) of the definition of $\mathcal{BU}$ and contains all the variables of Y. Suppose this literal has the form $\neg P^{(i,j)}(\overline{x}, \overline{y})$, where $\overline{x} \subseteq X$ and $\overline{y} = Y$, and the corresponding premise has the form $P(\overline{u}, \overline{t})$. The grouping restriction ensures that the sequences of variables $\overline{x}$ and $\overline{y}$ have the same length as the sequences of terms $\overline{u}$ and $\overline{t}$, respectively. Then $\overline{u} \subseteq \overline{s}$. If $P(\overline{u}, \overline{t})$ is a uni-node then, as above, it is easy to see that all other premises and all maximally split conclusions are uni-nodes and the depths of all arguments are the same. If not, then $P(\overline{u}, \overline{t})$ is a bi-node over $\{\overline{u}, \overline{t}\}$ (by the induction hypothesis and because the grouping associated with a predicate symbol is unique). Hence, $\overline{u}$ and $\overline{t}$ are distinct uni-nodes and $\overline{u}$ (and $\overline{s}$) is a direct successor of $\overline{t}$, or vice versa. As X and $\overline{y}$ together cover all the variables of C, all other premises and all maximally split conclusions are either uni-nodes defined over $\overline{s}$ or $\overline{t}$ (more precisely, uni-nodes of the form $P(\overline{w})$ where $\overline{w} \subseteq \overline{s}$ or $\overline{w} \subseteq \overline{t}$), or they are bi-nodes over $\{\overline{w}, \overline{v}\}$, where $\overline{w} \subseteq \overline{s}$ and $\overline{v} \subseteq \overline{t}$. This proves property 2. As $\overline{s}$ is a direct successor of $\overline{t}$, or vice versa, the difference in depth between terms in $\overline{u}$ (or $\overline{s}$) and terms in $\overline{t}$ is one. Property 3 follows directly.

Consider the resolution steps, where the negative premise is a clause C which satisfies Condition 2. This means that the negative premise is a non-ground functional clause and the main literal contains all the variables of C.

Assume that the positive premise corresponding to the main literal is the uni-node $Q(\overline{s})$. Then each variable in C will be unified with a term from $\overline{s}$. By Condition 2.(b) we have that either (i) every other literal L in C satisfies the grouping condition with respect to two disjoint sequences of variables $\overline{x}$ and $\overline{y}$, or (ii) has the form $(\neg)P^{(i,j)}(\overline{x}, \overline{f_{\overline{x}}(\overline{u}_j)})$ where $\overline{x}$ is a sequence of variables and $\overline{f_{\overline{x}}(\overline{u}_j)}$ is a sequence of j terms $f_l(\overline{u}_l)$, $1 \leq l \leq j$, such that $\overline{u}_l \subseteq \overline{x}$, and $\overline{u}_l$ is not empty or (iii) has the form $(\neg)P^{(i,j)}(\overline{f_{\overline{x}}(\overline{u}_{i+j})})$, where for every $1 \leq l \leq i{+}j$, $\overline{u}_l \subseteq \overline{x}$, $\overline{u}_l$ is non-empty, and $\overline{x}$ is a sequence of the variables of the main literal of C.

Consider case (i). The sequences of variables $\overline{x}$ and $\overline{y}$ in L are instantiated with terms from $\overline{s}$, because the main literal contains all the variables of the clause and consequently terms from $\overline{s}$ instantiate all variables. Therefore the argument set of any maximally split conclusion derived by instantiating the arguments of such literal is a sequence of terms $\overline{u}$, where $\overline{u} \subseteq \overline{s}$. We know that the sequence of terms $\overline{s}$ is a uni-node. By assumption all the terms in $\overline{s}$ have the same depth. Therefore a split conclusion corresponding to L is a uni-node and its arguments are of the same depth.

Consider case (ii). The argument set of the split conclusion which is an instance of a literal with variable as well as non-variable arguments has the form $P(\overline{w}, \overline{f_{\overline{w}}(\overline{t})})$, where $\overline{w} \subseteq \overline{s}$ so that $\overline{t} \subseteq \overline{w}$. By definition this means that $\overline{f_{\overline{w}}(\overline{t})}$ is a direct successor of $\overline{w}$. Therefore conclusions of the form $P(\overline{w}, \overline{f_{\overline{w}}(\overline{t})})$ are bi-nodes. Because $\overline{w}$ and $\overline{t}$ are subsequences of $\overline{s}$, they have the same depth d, and all the terms in $\overline{f_{\overline{w}}(\overline{t})}$ have depth $d{+}1$.

Consider Case (iii). The arguments of every functional literal of the form $(\neg)P^{(i,j)}(\overline{f_{\overline{w}}(\overline{u})})$ after instantiation with a subsequence of the arguments $\overline{w} \subseteq \overline{s}$ form a uni-node. We have that the arguments of the derived clause of the form $P^{(i,j)}(\overline{f_{\overline{w}}(\overline{u})})$, where $\overline{u} \subseteq \overline{w}$, form a uni-node which is a direct successor of the uni-node $\overline{w}$. As before because $\overline{f_{\overline{w}}(\overline{u})})$ is a direct successor of $\overline{w}$, and it has the same depth as $\overline{s}$, and the difference in depth between the terms $\overline{f_{\overline{w}}(\overline{u})})$ and $\overline{s}$ is one. □

The importance of the acyclic dependency relationship on the predicate symbols for decidability will become apparent in the proof of Lemma 5.3. By the definition of $\mathcal{BU}$ the negative premise of a hyperresolution inference step always contains a main literal, which is strictly maximal with respect to the ordering $\succ_D$. Hence a non-empty conclusion is always smaller than the main premise resolved with the main literal. However, for termination this property is not sufficient. We also need the following result.

Lemma 5.3. *There is a bound on the term depth of any clause in a $\mathsf{R}^{\mathsf{hyp}}$ derivation from N belonging to $\mathcal{BU}$.*

Proof. In order to establish this bound we define a complexity measure μ on ground unit clauses by:

$$\mu(C) = \begin{cases} Q & \text{if } C = Q(\overline{s}) \text{ and } Q \text{ has grouping } (0,i) \text{ or } (i,0), \\ P & \text{otherwise, where } P \text{ is the predicate symbol of the main premise} \\ & \text{with which } C \text{ was derived.} \end{cases}$$

Thus the complexity measure of any non-empty, ground unit clause is determined by a predicate symbol with grouping $(0, i)$ or $(i, 0)$. By definition the complexity measures are related by $\succ_d$. This is an acyclic relationship, which can always be linearised. Suppose therefore that $\succ_c$ is an arbitrary total ordering on the predicate symbols in N with grouping $(0, i)$ or $(i, 0)$ and $\succ_d \subseteq \succ_c$.

Note that the complexity measure of a clause C depends on the way C has been derived. For example, consider the clause set $\{Q_1(a), Q_2(a), \neg Q_1(x) \vee R(x, f(x)), \neg Q_2(x) \vee R(x, f(x))\}$. There are two ways to derive the unit clause $R(a, f(a))$. $\mu(R(a, f(a))) = Q_1$ if $R(a, f(a))$ is derived from $Q_1(a)$ and $\neg Q_1(x) \vee R(x, f(x))$ and $\mu(R(a, f(a))) = Q_2$ if $R(a, f(a))$ is derived from $Q_2(a)$ and $\neg Q_2(x) \vee R(x, f(x))$. Since we are considering clause sets, not clause multisets, at any time there can only be one copy of a ground clause in the clause set and we assume that it is known how this copy has been derived. Under this assumption the complexity measure μ is well-defined.

The proof is by induction on the ordering $\succ_c$ over predicate symbols in N with grouping $(0, i)$ or $(i, 0)$. Technically, let the enumeration be $Q_1 \succ_c Q_2 \succ_c \cdots \succ_c Q_n$ where n is the number predicate symbols in N with grouping $(0, i)$ or $(i, 0)$. Let $\overline{\mu}$ be a function from ground unit clauses to $\{1, \ldots, n\}$ such that $\overline{\mu}(C) = i$, provided $\mu(C) = Q_i$. In a sense $\overline{\mu}$ preserves μ, but reverses the ordering. We prove the following property is true for every non-empty, ground unit clause C derived from N.

$$dp(C) \leq \begin{cases} \overline{\mu}(C) & \text{if } C = Q(\overline{s}) \text{ and } Q \text{ has grouping } (0,i) \text{ or } (i,0), \\ \overline{\mu}(C) + 1 & \text{otherwise.} \end{cases} \tag{†}$$

Initially the only ground unit clauses in N are those which have depth 1. For any of these clauses C, $dp(C) \leq \overline{\mu}(C)$. Suppose the induction hypothesis is: Let D be an arbitrary ground unit clause in the derivation and suppose (†) is true for all ground unit clauses C in the derivation with greater measure, i.e. $\mu(C) \succ_D \mu(D)$.

Assume D is a maximally split conclusion of an inference with the main premise $Q(\overline{s})$ and negative premise C. Then $Q(\overline{s}) \succ_D D$ and the inductive hypothesis applies to $Q(\overline{s})$. (a) D can either have a predicate symbol with grouping $(0, i)$ or $(i, 0)$. Then $\overline{\mu}(Q(\overline{s})) < \overline{\mu}(D)$ as Q is greater than any other predicate symbol in C. (b) Otherwise, by the definition of μ, $\mu(Q(\overline{s})) = Q = \mu(D)$ as Q is the predicate symbol of the main premise with which D is derived, and consequently, $\overline{\mu}(Q(\overline{s})) = \overline{\mu}(D)$. So, in either case, i.e. for any D, we have the property:

$$\overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D). \tag{‡}$$

Now suppose C is a clause which satisfies Condition 1. If the main literal contains all the variables of C, then D is a uni-node, $P'(\overline{w})$, say, where $\overline{w} \subseteq \overline{s}$ (by the same argument as in the proof of Lemma 5.2). Thus, $dp(D) = dp(Q(\overline{s}))$ since $dp(\overline{w}) = dp(\overline{s})$, $dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s}))$ by the inductive hypothesis, and $\overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ by (‡). Consequently, $dp(D) \leq \overline{\mu}(D)$. Hence, (†) holds for D in this case.

If the main literal in C does not contain all the variables, w.l.o.g. assume the main literal contains all the variables from $X(\neq \emptyset)$ and there is a negative literal in C which satisfies 1.(b.iii) and contains all the variables of Y, where X and Y are defined as in 1.(a) of the definition of $\mathcal{BU}$. Suppose this literal has the form $\neg P^{(i,j)}(\overline{x}, \overline{y})$ where $\overline{x} \subseteq X$ and $\overline{y} = Y$, and the corresponding premise is $P^{(i,j)}(\overline{u}, \overline{t})$. Hence, $\overline{u} \subseteq \overline{s}$.

If $P(\overline{u}, \overline{t})$ is a uni-node then $dp(\overline{u}) = dp(\overline{s}) = dp(\overline{t})$. D is an instance of a positive literal L in C and by Condition 1.(b) all variables of D are in $X \cup Y$. So, all the arguments of D are among the arguments of $Q(\overline{s})$ and $P(\overline{u}, \overline{t})$. Therefore, $dp(D) = dp(P(\overline{u}, \overline{t})) = dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ (as above, by the inductive hypothesis and (‡)). Hence, (†) holds in this case.

If $P^{(i,j)}(\overline{u}, \overline{t})$ is a bi-node, and $\overline{u}$ is a direct successor of $\overline{t}$, then $dp(\overline{u}) = dp(\overline{t}) + 1$. By Condition 1.(b), if D is a uni-node, then all arguments of D are either all among $\overline{s}$ or all among $\overline{t}$. In the first case because $dp(\overline{u}) = dp(\overline{s})$, $dp(D) = dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ (again, by the inductive hypothesis and (‡)). Similarly, when D is a uni-node over $\overline{t}$, $dp(D) = dp(Q(\overline{s})) - 1$, because $dp(\overline{s}) = dp(\overline{u}) = dp(\overline{t}) + 1$. Then $dp(D) \leq \overline{\mu}(Q(\overline{s})) - 1 \leq \overline{\mu}(D) - 1$. This implies that $dp(D) \leq \overline{\mu}(D)$. Otherwise, D is a bi-node over $\{\overline{w}, \overline{v}\}$ where $\overline{w} \subseteq \overline{s}$ and $\overline{v} \subseteq \overline{t}$ (by the same argument as in the proof of Lemma 5.2). Then $dp(D) = dp(Q(\overline{s})) \leq \overline{\mu}((Q(\overline{s})) \leq \overline{\mu}(D)$. This proves (†).

If, on the other hand, $\overline{t}$ is a direct successor of $\overline{u}$, then $dp(\overline{t}) = dp(\overline{u}) + 1 = dp(\overline{s}) + 1$. Similarly as above, D is either a uni-node over $\overline{s}$, a uni-node over $\overline{t}$, or a bi-node over $\{\overline{w}, \overline{v}\}$ where $\overline{w} \subseteq \overline{s}$ and $\overline{v} \subseteq \overline{t}$. Then $dp(D) = dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ in the first case. In the second and third case, $dp(D) = dp(Q(\overline{s})) + 1 \leq \overline{\mu}(Q(\overline{s})) + 1 \leq \overline{\mu}(D) + 1$. This proves (†).

Now assume that C is a clause which satisfies Condition 2. By the definition of $\mathcal{BU}$ the main literal contains all the variables of C. Because of Condition 2.(b) we have three separate cases for the form of the maximally split conclusion D of an inference with the main premise $Q(\overline{s})$ and the negative premise C. Namely (i) D is a uni-node, say, $P'(\overline{w})$, where $\overline{w} \subseteq \overline{s}$, (ii) D is a bi-node over $\{\overline{w}, \overline{v}\}$, where $\overline{w} \subseteq \overline{s}$, $\overline{v} \subseteq \overline{f_{\overline{w}}(\overline{u})}$, where $\overline{u} \subseteq \overline{w}$, and (iii) D is a uni-node, say, $P'(\overline{u})$ where $\overline{u}$ is a direct successor of $\overline{s}$, and (by the same argument as in the proof of the second case of Lemma 5.2).

Assume that D is of the form (i). Then Q is the maximal predicate symbol with respect to $\succ_d^+$ and therefore we have that $Q(\overline{s}) \succ_D D$. We have that $dp(\overline{w}) = dp(\overline{s})$, i.e. $dp(D) = dp(Q(\overline{s}))$, $dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s}))$ by the inductive hypothesis, and $\overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ by (‡). Therefore we have that $dp(D) \leq \overline{\mu}(D)$ and (†) holds for D in this case.

Next assume that D is of the form (ii), i.e. D is a bi-node of the form $P(\overline{w}, \overline{f_{\overline{w}}(\overline{t})})$. Because $\overline{f_{\overline{w}}(\overline{t})}$ is a direct successor of $\overline{w}$ and $\overline{w} \subseteq \overline{s}$, we have that $dp(D) = dp(Q(\overline{s})) + 1$. By the inductive hypothesis and (‡) we have that $dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$ and consequently $dp(D) \leq \overline{\mu}(D) + 1$. Therefore (†) holds.

Finally, assume that D is of the form (iii). Then by definition because $\overline{u}$ is a direct successor of $\overline{s}$ we have that $dp(\overline{u}) = dp(\overline{s}) + 1$. Therefore $dp(D) = dp(Q(\overline{s})) + 1$. By the induction hypothesis we have that $dp(Q(\overline{s})) \leq \overline{\mu}(Q(\overline{s}))$. By (‡) we have that $\overline{\mu}(Q(\overline{s})) \leq \overline{\mu}(D)$. Therefore $dp(D) = dp(Q(\overline{s})) + 1 \leq \overline{\mu}(Q(\overline{s})) + 1 \leq \overline{\mu}(D) + 1$. This proves (†).

□

Example 5.1. *As an illustration consider the following clause set.*

$$\begin{array}{rl} \{(1) & Q(a,a), \\ (2) & \neg Q(x,x) \vee P(x,x,f(x,x)) \\ (3) & R(a,b), \\ (4) & \neg R(x,y) \vee Q(x,y) \vee P(x,y,f(x,y))\} \end{array}$$

It belongs to $\mathcal{BU}$ provided that $R \succ_d Q \succ_d P$ and (i) the grouping associated with the predicate symbols R is either $(0,2)$ or $(2,0)$; (ii) the grouping associated with the predicate symbol Q is also either $(0,2)$ or $(2,0)$, and (iii) the grouping associated with the predicate symbol P is $(2,1)$.

The following (non-redundant) conclusions can be derived by $\mathsf{R}^{\mathsf{hyp}}$ from the clause set.

$$\begin{array}{ll} (1,2) & P(a,a,f(a,a)) \\ (3,4) & Q(a,b) \vee P(a,b,f(a,b)) \end{array}$$

$$\begin{array}{ccc} & \textit{Split} & \\ \swarrow & & \searrow \\ (3,4)\ \ Q(a,b) & & (3,4)\ \ P(a,b,f(a,b)) \end{array}$$

The complexity measure $\mu(C)$ on all ground positive unit clauses C is then defined as follows.

$$\begin{array}{rcl} \mu(R(a,b)) & = & R \\ \mu(Q(a,b)) & = & Q \\ \mu(P(a,b,f(a,b)) & = & R \\ \mu(Q(a,a)) & = & Q \\ \mu(P(a,a,f(a,a)) & = & Q \end{array}$$

Given that the ordering among predicate symbols R and Q is $R \succ_d Q$, we let $\overline{\mu}(Q(a,b)) = 2$ and $\overline{\mu}(R(a,b)) = 1$. The complexity measure $\overline{\mu}(C)$ as defined in Lemma 5.3 gives us the following term depth bounds.

$$\begin{array}{rcl} dp(R(a,b)) & \leq & 1 \\ dp(Q(a,b)) & \leq & 2 \\ dp(P(a,b,f(a,b))) & \leq & \overline{\mu}(R(a,b)) + 1 \leq 2 \\ dp(Q(a,a)) & \leq & 1 \\ dp(P(a,a,f(a,a))) & \leq & \overline{\mu}(Q(a,a)) + 1 \leq 2 \end{array}$$

Theorem 5.1 (Termination, soundness, completeness). *Let N be a finite set of $\mathcal{BU}$ clauses. Then:*

1. *Any $\mathsf{R}^{\mathsf{hyp}}$ derivation from N terminates.*

2. *If T is a fair derivation from N then: (i) If $N(=N_0), N_1, \ldots$ is a path with limit N_∞, N_∞ is saturated up to redundancy. (ii) N is satisfiable if and only if there exists a path in T with limit N_∞ such that N_∞ is satisfiable. (iii) N is unsatisfiable if and only if for every path $N(=N_0), N_1, \ldots$ the clause set $\bigcup_j N_j$ contains the empty clause.*

This theorem subsumes corresponding results for GF1$^-$ [13] and the first-order fragment encoding the extended modal logic $K_{(m)}(\cap, \cup, \smile)$ [10, 19].

The calculus $\mathsf{R}^{\mathsf{hyp}}$ (with optional factoring) is the simplest calculus with which the class $\mathcal{BU}$ can be decided. In practice, one wants to maximize the efficiency. For this purpose, the result above permits the use of any refinement and simplification rules based on the resolution framework of [4]. The result also permits the use of stronger versions of the splitting rule which ensure the branches in a derivation tree are disjoint. Such splitting rules can lead to earlier closure of branches.

6 Minimal Herbrand model generation

It is well-known that hyperresolution, like tableaux methods, can be used to construct models for satisfiable formulae [11] and minimal Herbrand models for satisfiable formulae and clausal classes [2, 6].

A *Herbrand interpretation* is a set of ground atoms. By definition a ground atom A is *true* in an interpretation H if $A \in H$ and it is *false* in H if $A \notin H$, $\top$ is true in all interpretations and $\bot$ is false in all interpretations. A literal $\neg A$ is true in H iff A is false in H. A conjunction of two ground atoms A and B is true in an interpretation H iff both A and B are true in H and respectively, a disjunction of ground atoms is true in H iff at least one of A or B is true in the interpretation. A clause C is true in H iff for all ground substitutions σ there is a literal L in $C\sigma$ which is true in H. A set N of clauses is true in H iff all clauses in N are true in H. If a set N of clauses is true in an interpretation H then H is referred to as a *Herbrand model* of N. H is a *minimal Herbrand model* for a set N of clauses iff H is a Herbrand model of N and for no Herbrand model H' of N, $H' \subset H$ holds.

For $\mathcal{BU}$ (more generally, range restricted clauses), the procedure $\mathsf{R}^{\mathsf{hyp}}$ implicitly generates Herbrand models. If $\mathsf{R}^{\mathsf{hyp}}$ terminates on a clause set N in $\mathcal{BU}$ without having produced the empty clause then a model can be extracted from any open branch in the derivation. The model is given by the set of ground unit clauses in the limit of the branch, i.e. the clause set at the leaf of the branch.

Bry and Yahya [8] have proved the following even stronger result: For every minimal model H of a satisfiable, range restricted clause set N, there exists a branch in the $\mathsf{R}^{\mathsf{hyp}}$ derivation tree for N, such that the set of ground unit clauses in the limit of the branch coincides with the ground atoms in H. Since by definition every clause set in $\mathcal{BU}$ is range restricted, this result also applies to $\mathcal{BU}$ clause sets.

Consequently, if we want to turn $\mathsf{R}^{\mathsf{hyp}}$ into a procedure which generates only minimal models for satisfiable clause sets in $\mathcal{BU}$, it is sufficient to modify the calculus in a way that eliminates all those branches of a derivation that would generate non-minimal models. In [14] we have discussed various ways of how this can be achieved, including (i) an approach which extends $\mathsf{R}^{\mathsf{hyp}}$ by a *model constraint propagation rule*, (ii) a modification of the extension of $\mathsf{R}^{\mathsf{hyp}}$ by the model constraint propagation rule which replaces the splitting rule by a *complement splitting rule* and investigates the derivation tree in a particular order [8], and finally, (iii) a variant of Niemelä's groundedness test [23] which tests the minimality of a model locally for each branch by invoking another theorem proving derivation. We have compared the worst case space requirements of these approaches for clause sets associated with $\mathrm{GF1}^-$ formulae and concluded that Niemelä's groundedness test has the best worst case space requirement among the three approaches for $\mathrm{GF1}^-$ formulae [14]. The minimal Herbrand model generation procedure which utilizes Niemelä's groundedness test has the best worst case space requirement among the three approaches for $\mathcal{BU}$ clause sets as well.

The groundedness test is based on the following observation. Given a (finite) set H of ground atoms (or positive unit clauses) define: $\neg H = \{\neg A \mid A \in H\}$ and $\overline{H} = \bigvee_{A \in H} \neg A$. Let N be a set of clauses and U the set of all atoms over the Herbrand universe of N. Let H be a finite Herbrand model of N. Then H is a minimal Herbrand model of N iff $MMT(N, H) = N \cup \neg(U - H) \cup \{\overline{H}\}$ is unsatisfiable. This model minimality test is called *groundedness test*. Thus, we can use $\mathsf{R}^{\mathsf{hyp}}$ to enumerate all models of a $\mathcal{BU}$ clause set N and also use $\mathsf{R}^{\mathsf{hyp}}$ to test each model H for minimality by testing $MMT(N, H)$ for unsatisfiability. This approach has also been applied in [2, 7], for ground clause logic.

A problem in applying the groundedness test to $\mathcal{BU}$ is that the set U of all atoms over the Herbrand universe of a $\mathcal{BU}$ clause set N is usually infinite. Consequently, $\neg(U - H)$ and $MMT(N, H)$ are usually infinite sets of clauses. However, in the case of an $\mathsf{R}^{\mathsf{hyp}}$ derivation from $MMT(N, H)$, we observe the clauses in $\neg(U - H)$ have only the effect of deriving a contradiction for any clause set N' derivable from N which contains a positive unit clause not in H. Since H itself is finite, this effect is straightforward to implement. A detailed presentation of the approach and an algorithmic description is given in [14].

Theorem 6.1. *Let N be a clausal set in $\mathcal{BU}$. Let N_∞ be the limit of any branch B in an $\mathsf{R}^{\mathsf{hyp}}$ derivation tree with root N and let H be the set of all positive ground unit clauses in N_∞. Then, the satisfiability of $MMT(N, H)$ can be tested in finite time and H is a minimal model of N iff $MMT(N, H)$ is unsatisfiable.*

7 Conclusion

The definition of the class $\mathcal{BU}$ attempts to capture characteristic properties for ensuring decidability by hyperresolution (or if the reader prefers hypertableaux or ground tableaux calculi, which are closely related, see [13, 14]), while permitting term depth growth during the inference process. $\mathcal{BU}$ covers many familiar description logics and the corresponding extended propositional modal logics, for example the description logic $\mathcal{ALC}$ with inverse roles, conjunctions and disjunctions of roles and the corresponding modal logics below $K_{(m)}(\cap, \cup, \smile)$. Although recent results (see e.g. [10, 11, 12, 17, 18, 19, 24]) show that ordered resolution is the more powerful method when decidability is an issue, an advantage of hyperresolution is that it can be used for Herbrand model generation without the need for extra machinery, except when we want to generate minimal Herbrand models for which a modest extension is needed (cf. Section 6).

An open question is the complexity of the decision problem of $\mathcal{BU}$. One of the advantages of GF1$^-$ compared to other solvable classes such as the guarded fragment or fluted logic is the low complexity of its decision problem, which is PSPACE-complete. Intuitively, due to the more restricted form of bi-nodes in GF1$^-$ it is possible to investigate bi-nodes independently of each other. For details see [14]. In contrast, the very general definition of bi-nodes given in this paper makes it difficult to establish whether the same approach is possible for $\mathcal{BU}$.

Acknowledgements

We thank the referees of the short version of this paper [15] for helpful comments and suggestions. The work is supported by research grants GR/M36700 and GR/R92035 from the EPSRC, UK.

References

[1] H. Andréka, J. van Benthem, and I. Németi. Back and forth between modal logic and classical logic. *Bull. IGPL*, 3(5):685–720, 1995.

[2] C. Aravindan and P. Baumgartner. Theorem proving techniques for view deletion in databases. *J. Symbolic Computat.*, 29(2):119–147, 2000.

[3] M. Baaz, U. Egly, and A. Leitsch. Normal form transformations. In *Handbook of Automated Reasoning*, pp. 273–333. Elsevier, 2001.

[4] L. Bachmair and H. Ganzinger. Resolution theorem proving. In *Handbook of Automated Reasoning*, pp. 19–99. Elsevier, 2001.

[5] L. Bachmair, H. Ganzinger, and U. Waldmann. Superposition with simplification as a decision procedure for the monadic class with equality. In *Proc. KGC'93*, vol. 713 of *LNCS*, pp. 83–96. Springer, 1993.

[6] P. Baumgartner, P. Fröhlich, U. Furbach, and W. Nejdl. Semantically guided theorem proving for diagnosis applications. In *Proc. IJCAI'97*, pp. 460–465. Morgan Kaufmann, 1997.

[7] P. Baumgartner, J. Horton, and B. Spencer. Merge path improvements for minimal model hyper tableaux. In *Proc. TABLEAUX'99*, vol. 1617 of *LNAI*. Springer, 1999.

[8] F. Bry and A. Yahya. Positive unit hyperresolution tableaux for minimal model generation. *J. Automated Reasoning*, 25(1):35–82, 2000.

[9] H. de Nivelle. Translation of S4 into GF and 2VAR. Manuscript, 1999.

[10] H. de Nivelle, R. A. Schmidt, and U. Hustadt. Resolution-based methods for modal logics. *Logic J. IGPL*, 8(3):265–292, 2000.

[11] C. G. Fermüller, A. Leitsch, U. Hustadt, and T. Tammet. Resolution decision procedures. In *Handbook of Automated Reasoning*, pp. 1791–1849. Elsevier, 2001.

[12] H. Ganzinger and H. de Nivelle. A superposition decision procedure for the guarded fragment with equality. In *Proc. LICS'99*, pp. 295–303. IEEE Computer Society Press, 1999.

[13] L. Georgieva, U. Hustadt, and R. A. Schmidt. Hyperresolution for guarded formulae. To appear in *J. Symbolic Computat.*

[14] L. Georgieva, U. Hustadt, and R. A. Schmidt. Computational space efficiency and minimal model generation for guarded formulae. In *Proc. LPAR 2001*, vol. 2250 of *LNAI*, pp. 85–99. Springer, 2001.

[15] L. Georgieva, U. Hustadt, and R. A. Schmidt. A new clausal class decidable by hyperresolution. To appear in *Proc. CADE'2002*, *LNAI*, Springer, 2002.

[16] E. Grädel. On the restraining power of guards. *J. Symbolic Logic*, 64:1719–1742, 1999.

[17] U. Hustadt and R. A. Schmidt. Maslov's class K revisited. In *Proc. CADE-16*, vol. 1632 of *LNAI*, pp. 172–186. Springer, 1999.

[18] U. Hustadt and R. A. Schmidt. Issues of decidability for description logics in the framework of resolution. In *Automated Deduction in Classical and Non-Classical Logics*, vol. 1761 of *LNAI*, pp. 191–205. Springer, 2000.

[19] U. Hustadt and R. A. Schmidt. Using resolution for testing modal satisfiability and building models. In *SAT 2000: Highlights of Satisfiability Research in the Year 2000*, pp. 459–483. IOS Press, Amsterdam, 2000.

[20] A. Leitsch. Deciding clause classes by semantic clash resolution. *Fundamenta Informaticae*, 18:163–182, 1993.

[21] C. Lutz. Complexity of terminological reasoning revisited. In *Proc. LPAR'99*, vol. 1705 of *LNAI*, pp. 181–200. Springer, 1999.

[22] C. Lutz, U. Sattler, and S. Tobies. A suggestion of an n-ary description logic. In *Proc. DL'99*, pp. 81–85. Linköping University, 1999.

[23] I. Niemelä. A tableau calculus for minimal model reasoning. In *Proc. TABLEAUX'96*, vol. 1071 of *LNAI*, pp. 278–294. Springer, 1996.

[24] R. A. Schmidt and U. Hustadt. A resolution decision procedure for fluted logic. In *Proc. CADE-17*, vol. 1831 of *LNAI*, pp. 433–448. Springer, 2000.

[25] M. Schmidt-Schauß and G. Smolka. Attributive concept descriptions with complements. *J. Artificial Intelligence*, 48:1–26, 1991.