Elementary Canonical Formulae: A Survey on Syntactic, Algorithmic, and Model-theoretic Aspects

Willem Conradie, Valentin Goranko and
Dimiter Vakarelov

ABSTRACT. In terms of validity in Kripke frames, a modal formula expresses a universal monadic second-order condition. Those modal formulae which are equivalent to first-order conditions are called *elementary*. Modal formulae which have a certain persistence property which implies their validity in all canonical frames of modal logics axiomatized with them, and therefore their completeness, are called *canonical*. This is a survey of a recent and ongoing study of the class of elementary and canonical modal formulae. We summarize main ideas and results, and outline further research perspectives.

1 Introduction

1.1 Elementary canonical formulae

We study modal formulae ϕ which are:

(i) elementary (first-order definable):

- *locally,* if there is a first-order formula $\alpha(x)$ such that for every frame F and $w \in F$:

$$F, w \models \phi \text{ iff } F \models_{FO} \alpha(w).$$

- or, *globally,* if there is a first-order sentence α such that for every frame F:

$$F \models \phi \quad \text{iff} \quad F \models_{FO} \alpha.$$

  Clearly, every locally elementary formula is globally so, too. The converse does not hold, as we will see later.

Advances in Modal Logic, Volume 5.

(ii) canonical: informally, that means valid in the canonical frames of all modal logics in which such formula is an axiom. This property is important, because it implies *frame completeness* of logics axiomatized with such formulae.

Formally, we define canonicity in a somewhat stronger, but more uniform and precise way, as *persistence* with respect to a suitable class of general frames containing all canonical general frames of the language. In the standard (polyadic) modal languages, these are the *descriptive frames* (see e.g. [2]), and we identify canonicity with persistence with respect to such frames ($\mathcal{D}$-persistence). However, we note that if the language contains special sorts, such as nominals, or the logics admit special inference rules, the notion of canonicity accordingly changes.

While the class[1] of globally elementary and canonical formulae properly extends the class of locally elementary and canonical ones (see examples in [3]), these two classes behave very similarly, and hereafter we will concentrate mainly on the latter one. By the Fine-van Benthem theorem (see [3]), an elementary modal formula is canonical *iff* the modal logic axiomatized by that formula is complete, so the elementary and canonical formulae are precisely the elementary and complete ones.

The elementary and canonical formulae axiomatize many important modal logics and are of particular practical interest, as they lend themselves to the computational tools developed for first-order logic.

Examples of locally elementary canonical formulae include:

- every valid modal formula;
- the axioms for most well-known logics, incl. T, B , $K4$, $S4$, $S5$,...
- all Sahlqvist formulae [39], and many more...

Non-examples include:

- any formula axiomatizing an incomplete logic, e.g. van Benthem's formula $\Box\Diamond\top \to \Box(\Box(\Box p \to p) \to p)$, which is elementary, but not complete, hence not canonical.
- Fine's formula $\Diamond\Box(p \vee q) \to \Diamond(\Box p \vee \Box q)$, which is canonical, but not elementary.

[1]Here and further we often use the term 'class' when referring to sets of specially defined formulae, not in set-theoretic sense, but as a stylistic way of emphasizing their importance and internal structure.

- McKinsey's formula $\Box\Diamond p \to \Diamond\Box p$ and the Gödel-Löb formula $\Box(\Box p \to p) \to \Box p$, which, although complete, are neither first-order definable, nor canonical.

Hereafter, 'elementary' modal formula will usually mean a globally elementary one, unless otherwise specified in the context.

The sets of locally and globally elementary canonical formulae are not recursive (see [3]) hence the problem arises to establish useful characterizations and to identify rich natural subclasses of effectively recognizable such formulae. Our study follows three main threads of obtaining such characterizations: *syntactic, algorithmic, and model-theoretic,* which are discussed in the subsequent sections. Below we summarize the main issues and results.

1.2 Syntactic classes of elementary canonical formulae

The best-known class of elementary canonical formulae, which was also the starting point of this study, is the class of *Sahlqvist formulae* [39]. While bearing a clear semantic motivation, these formulae are defined purely syntactically, and that syntactic definition is only a lower approximation of the underlying semantic idea (of *minimal valuations*, see [3]). The syntactic definition is extremely fragile, as it does not withstand even simple boolean transformations, or even substitutions changing the polarity of propositional variables. It has, therefore, become customary to tacitly consider the Sahlqvist formulae closed under such simple transformations. On the other hand (see [7]), axiomatic equivalence to a Sahlqvist formula is not decidable, and hence it would be unreasonable to close the class of Sahlqvist formulae under such equivalences. Thus, the notion of Sahlqvist formulae has become fuzzy, and the question '*What is a Sahlqvist formula?*' has gained increasing pertinence.

In [23] and [25] we have extended the Sahlqvist formulae to the class of *inductive formulae* in arbitrary polyadic languages, also generalized for hybrid modal languages in [24]. These are still syntactically defined elementary canonical formulae, and their first-order equivalents are still computed by means of minimal, first-order definable valuations which enable elimination of the second-order predicate variables. These minimal valuations are defined inductively, in an order determined by certain syntactic dependencies between the propositional variables within the formula. Like Sahlqvist formulae, the syntactic shape of inductive formulae is rather vulnerable to otherwise inessential transformations, and thus the question 'What is an inductive formula?' remains actual.

In our study we analyze the possibilities to extend the class of syntactically determined elementary canonical formulae:

- by extending further the syntactic definition of inductive formulae;

- by adding and refining a *pre-processing* phase, in attempt to transform the formula into an inductive formula, while preserving its frame condition.
- by closing the class of inductive formulae under suitable equivalences preserving the elementary canonical formulae, to larger, still effectively recognizable classes.

Also, we develop purely syntactic procedures for computation of the first-order equivalents of effectively defined classes of elementary canonical formulae. For instance, in [24] we present such method for inductive formulae in temporal (more generally, reversive) languages with nominals.

1.3 Algorithmic approach to elementary canonical formulae

A natural extension of the syntactic approach aims at development of algorithms which identify elementary canonical formulae, and thus produce *effectively enumerable* classes of such formulae. Such algorithmically definable classes need not be decidable, but they are much less tied-up with the syntactic shape of the formulae, and the algorithmic approach penetrates deeper into the semantic nature of the elementary canonical formulae.

To establish first-order definability of a modal formula amounts to elimination of the monadic second-order quantifiers occurring in its standard translation. Therefore, prime candidates for algorithms producing elementary canonical formulae are the two currently developed and implemented algorithms for second-order quantifier elimination, viz. SCAN and DLS (see [34]). Both are provably correct and incomplete and, while seemingly based on different ideas and with quite distinct computational behavior, none of them is stronger than the other for that task. It has been proved in [27] that SCAN succeeds for all Sahlqvist formulae. Moreover, this holds for all inductive formulae, and the same applies to DLS (see [12]).

In [10] we have developed a new algorithm, SQEMA, for computing first-order equivalents of modal formulae and have proved the canonicity of all formulae on which it succeeds.

1.4 Model-theoretic aspects of elementary canonical formulae

This direction of research aims at characterizing the elementary canonical formulae of a given modal language in practically more useful model-theoretic terms. A typical such characterization (as a sufficient condition) is a suitable notion of persistence, e.g.: with respect to all descriptive frames, in the case of standard polyadic modal languages; with respect to all discrete frames, in the case of hybrid modal languages with nominals; with respect to all refined frames, in the case of logics with additional non-orthodox rules of inference.

Various other persistence properties have emerged as useful tools for model theoretic analysis and classification of elementary canonical, and related, formulae. For instance, as established by van Benthem in [3], the modal formulae amenable to the method of substitutions turn out to be precisely those persistent with respect to the general frames in which all parametrically first-order definable sets are admissible. Also, in [25] we have introduced a new notion of persistence which separates the Sahlqvist formulae from the inductive ones, and have proved that, up to local equivalence in all discrete general frames in reversive languages with nominals, the inductive, pure (not containing propositional variables), and locally discrete-persistent formulae, coincide, thus delineating a very large and natural class of elementary and 'discretely-canonical' formulae in such languages.

The persistence properties of the elementary canonical formulae have a distinct topological nature, first identified in [40] and used in [41] to give a uniform proof of first-order definability and canonicity of Sahlqvist formulae. We have continued and extended that analysis, and used topological arguments to establish first-order definability and canonicity of inductive formulae in [25], and of the formulae on which the algorithm SQEMA succeeds in [10].

2 Preliminaries

In this paper we assume that the reader has basic familiarity with syntax and semantics of modal logic, some useful references on which include [3], [2], and [6]. For the reader's convenience, we briefly recall some important facts related to general frames and persistence.

2.1 General frames

For technical simplicity, we will only consider a basic monadic modal language. For treatment of general polyadic languages see [23], [25], as well as [24] for hybrid polyadic languages. For general background on model theory of modal logic see [2], [6], [8], and [28].

Given a Kripke frame $F = \langle W, R \rangle$, a **general frame over** F is a structure $\mathfrak{F} = \langle W, R, \mathbb{W} \rangle$ expanding F with a modal algebra $\mathbb{W}$ of **admissible subsets** of W, closed under all Boolean and modal operators, i.e., $\mathbb{W}$ is a modal subalgebra of $\langle \mathcal{P}(W); \cap, -, \Box, \varnothing \rangle$, where $\Box X = \{x \in W \mid \forall y(Rxy \to y \in X)\}$. The operator $\Diamond$ is defined dually: $\Diamond X = \{x \in W \mid \exists y(Rxy \wedge y \in X)\}$.

A valuation V in the frame F is **admissible** in $\mathfrak{F}$ if $V(p) \in \mathbb{W}$ for every variable p. A **Kripke model over** $\mathfrak{F}$ is any Kripke model $M = \langle F, V \rangle$ with a valuation V admissible in $\mathfrak{F}$.

Local (at a state) and **global validity of a modal formula in a general frame** is defined as truth at the state (resp. validity) in every admissible model over that general frame.

Every general frame $\mathfrak{F} = \langle W, R, \mathbb{W} \rangle$ defines a topology $T(\mathfrak{F})$ on W with $\mathbb{W}$ as a base of clopen sets, i.e. the closed sets of the topology are precisely all intersections of admissible sets.

2.2 Some important classes of general frames

A general frame $\mathfrak{F} = \langle W, R, \mathbb{W} \rangle$ is:

- **differentiated** if for every $x, y \in W$, if $x \neq y$ then $x \in X$ and $y \notin X$ for some $X \in \mathbb{W}$ or, equivalently, if $T(\mathfrak{F})$ is Hausdorff.
- **tight** if for any $x, y \in W$, Rxy iff $\forall Y \in \mathbb{W}(y \in Y \Rightarrow x \in \Diamond(Y))$, or, equivalently, if R is point-closed, i.e. $R(\{x\})$ is closed for every $x \in W$, where $R(X)$ denotes the set of all successors of points in X.
- **refined** if it is differentiated and tight.
- **compact** if every family of admissible sets in $\mathfrak{F}$ with FIP has a non-empty intersection, or, equivalently, if $T(\mathfrak{F})$ is compact.
- **discrete**, if $\{u\} \in \mathbb{W}$ for every $u \in W$.
- **elementary**, if every subset of W, which is parametrically first-order definable in the first-order language for Kripke frames, is admissible.
- **descriptive** if it is refined and compact.

The class of all differentiated (resp. tight, refined, discrete, elementary, descriptive) general frames will be denoted by $\mathcal{DF}$ (resp. $\mathcal{T}, \mathcal{R}, \mathcal{DI}, \mathcal{E}, \mathcal{D}$).

Some relationships between these classes (see [28]):

$$\mathcal{E} \subsetneqq \mathcal{DI} \subsetneqq \mathcal{R} = \mathcal{DF} \cap \mathcal{T}; \quad \mathcal{D} \subsetneqq \mathcal{R}; \quad \mathcal{D} \nsubseteq \mathcal{DI} \nsubseteq \mathcal{D}.$$

2.3 Persistence and canonicity

Let $\mathcal{C}$ be any class of general frames. A modal formula is **locally $\mathcal{C}$-persistent**, if for every general frame $\mathfrak{F} = \langle F, \mathbb{W} \rangle \in \mathcal{C}$, and $w \in F$: $\mathfrak{F}, w \models \phi$ implies $F, w \models \phi$; ϕ is **$\mathcal{C}$-persistent**, if for every such general frame, $\mathfrak{F} \models \phi$ implies $F \models \phi$.

Clearly, local persistence implies persistence, but the converse does not always hold. If we denote by $\mathcal{C}^p$ the set of all $\mathcal{C}$-persistent formulae, we have the following (see [28]):

$$\mathcal{DF}^p \cap \mathcal{T}^p = \mathcal{R}^p \subsetneqq \mathcal{DI}^p \subsetneqq \mathcal{E}^p; \quad \mathcal{R}^p \subsetneqq \mathcal{D}^p; \quad \mathcal{DI}^p \nsubseteq \mathcal{D}^p \nsubseteq \mathcal{E}^p.$$

The same relationships hold for the local persistences.

How are persistence and canonicity related? Descriptive frames typically appear as the *canonical general frames* of every normal modal logic without any special inference rules. Thus, all $\mathcal{D}$-persistent formulae are valid in the underlying canonical Kripke frames, and hence they axiomatize Kripke complete logics. For that reason the $\mathcal{D}$ -persistent formulae are often (incl. in this study) identified with canonical formulae.

However, in hybrid logics with nominals, or in logics with special additional ('context', or 'non-orthodox') rules of inference, $\mathcal{D}$ -persistent formulae *need not* be canonical, because the canonical general frames for such logics are only *discrete* (for hybrid logics, see [2]) or *refined* (in logics with additional rules).

Furthermore, $\mathcal{DI}$-persistent formulae have the important property of remaining canonical when added as axioms to hybrid logics with nominals, while $\mathcal{R}$-persistent formulae remain canonical not only in the presence of other axioms, but even if additional rules of inference of the type mentioned above are added to the axiomatic system. Thus, the right notion of canonicity in such languages is $\mathcal{DI}$-persistence, resp. $\mathcal{R}$-persistence.

3 Syntactic classes of elementary canonical formulae

3.1 The starter: Sahlqvist formulae

After the introduction of Kripke semantics for modal logics, a quest for general completeness results ensued, which culminated in Sahlqvist's theorem [39]. Sahlqvist proved two notable facts about a large, syntactically defined class of modal formulae, now called Sahlqvist formulae: the *first-order correspondence:* that they all define first-order conditions on Kripke frames and these conditions can be effectively "computed" from the modal formulae; and the *canonicity:* that all these formulae are valid in their respective canonical frames, and hence axiomatize completely the classes of frames satisfying their corresponding first-order conditions.

DEFINITION 1. In a fixed standard modal language ML we define the following syntactic classes of formulae.

- **Positive and negative formulae** are defined as usual.
- A **boxed atom** is a formula $\Box_1 \dots \Box_n p$ where $\Box_1, \dots, \Box_n$ is a (possibly empty) string of unary boxes and p is a propositional variable.
- A **Sahlqvist antecedent** is a formula constructed from boxed atoms and negative formulae by applying conjunctions, disjunctions, and diamonds.
- A **definite Sahlqvist antecedent** is a Sahlqvist antecedent obtained without applying disjunctions.

- A **(definite) Sahlqvist implication** is a formula $A \to P$ where A is a (definite) Sahlqvist antecedent and P is a positive formula.

- A **Sahlqvist formula** is a formula obtained from Sahlqvist implications by freely applying conjunctions, disjunctions, and boxes.

  We note that every Sahlqvist implication is tautologically equivalent to a formula of the type $\neg A$ where A is a Sahlqvist antecedent, and therefore every Sahlqvist formula is semantically equivalent to a negated Sahlqvist antecedent, too.

  The set of Sahlqvist formulae of ML will be denoted by SF(ML), or just SF if the language is clear from the context.

Some examples: $\Diamond\Box p \to \Box p$ and $\Box((\Box(\Diamond\neg p \vee \Diamond\Box\neg q) \wedge \Diamond\Box p) \to \Box\Diamond\Box(p \vee \Diamond\Box q))$ are Sahlqvist formulae, but not $\Box\Diamond p \to p$ or $\Box(p \vee q) \to (p \vee q)$. Even the K-axiom $\Box(p \to q) \to (\Box p \to \Box q)$, or its equivalent $\Box p \wedge \Box(\neg p \vee q) \to \Box q$ are (syntactically) not Sahlqvist formulae.

THEOREM 2 (Sahlqvist, 1973). *All Sahlqvist formulae are elementary and canonical.*

For more on Sahlqvist's theorem, including a proof and related results, see [3], [41], [2], [6], [33], [30], [16]. A generalization of Sahlqvist's theorem for polyadic languages has been proposed in [38].

What makes Sahlqvist formulae tick? The characteristic semantic feature of the Sahlqvist formulae, which is in the heart of *Sahlqvist-van Benthem substitution method* (see [3], [2]), is the existence of a *minimal valuation* for the occurring propositional variables, which makes the antecedent true. The minimal valuations have the following property:

If a Sahlqvist formula is valid for the minimal valuation in any given frame, then it is valid for every valuation on that frame.

Thus, the idea of the method of substitutions, applied to Sahlqvist formulae, is to compute the minimal valuations from the antecedent of the standard translation, and then to substitute them in the consequent. The result of that substitution is an equivalent formula where the second-order predicates in the standard translation are eliminated.

Example: Take $\theta = \Box p \to \Box\Box p$.

Then $ST(\theta)(x) = \forall y(Rxy \to Py) \to \forall y(Rxy \to \forall z(Ryz \to Pz))$. The minimal valuation of P satisfying the antecedent is $P = \{y \mid Rxy\}$. Substituting in the consequent yields $\forall y(Rxy \to \forall z(Ryz \to Rxz))$, i.e. transitivity.

Furthermore, the minimal valuations for Sahlqvist formulae are:

- *first-order definable,* whence the first-order definability of Sahlqvist formulae;
- *closed sets* in the topological spaces generated by the admissible sets in descriptive frames, whence the canonicity of Sahlqvist formulae can be derived.

How far does the class of Sahlqvist formulae stretch? On one hand, it is quite large but on the other, being syntactically defined, it is unstable even under Boolean transformations, so usually some *pre-processing* is needed to transform (if possible) a formula into that shape. For instance:

- $\neg p \to \neg\Box p$ is not a Sahlqvist formula but becomes one after a trivial Boolean transformation. Likewise for the contradictory formula $(\Box(\Box p \to p) \to \Box p) \wedge \neg(\Box(\Box p \to p) \to \Box p)$.
- $\Box(p \to q) \to (\Box p \to \Box q)$ becomes a Sahlqvist formula after a Boolean transformation and substitution of $\neg q$ for q. Likewise for $\Box\Diamond p \to p$ and $\Box(p \vee q) \to (p \vee q)$.
- $p \wedge \Box(\Diamond p \to \Box q) \to \Diamond\Box q$ is not a Sahlqvist formula, nor is it reducible to one with such 'simple' syntactic transformations. Yet, it is an elementary canonical formula and determines the same frame condition as the Sahlqvist formula $p \to \Diamond(\Diamond p \vee \Box\bot)$.

Sahlqvist formulae do not cover, in any reasonable sense, all elementary canonical formulae. For instance:

- $p \wedge \Box(\Diamond p \to \Box q) \to \Diamond\Box\Box q$ is an elementary canonical formula but not a Sahlqvist formula, nor is it reducible to one. In fact, it does not determine the same frame condition as any Sahlqvist formula in the basic modal language, as proved in [25]. Still, in the basic tense language it is equivalent to the Sahlqvist formula $p \to FGGP(Fp \wedge Pp)$.
- Likewise, $(\Box\Diamond p \to \Diamond\Box p) \wedge (\Box p \to \Box\Box p)$ is an elementary canonical formula [3], but not frame equivalent to any Sahlqvist formula.
- $\Diamond\Box(p \vee q) \to \Diamond(\Box p \vee \Box q)$, $\Box\Diamond p \to \Diamond\Box p$, and $\Box(\Box p \to p) \to \Box p$ are not Sahlqvist formulae, because they are neither elementary nor canonical [3].

A formula which is not in SF can possibly still be reduced to an equivalent Sahlqvist formula, in one or another sense. So, *what should we call a Sahlqvist formula?* To address this question, we first need to digress from it and discuss in more detail the various natural notions of equivalence arising in modal logic.

3.2 A hierarchy of equivalences

DEFINITION 3. Modal formulae A and B are:

- **tautologically equivalent (TAU)** if $A \leftrightarrow B$ is a propositional tautology.
- **semantically equivalent (SEM)** if $A \leftrightarrow B$ is a valid modal formula, i.e. A and B are valid at the same states of every Kripke model.
- **model-equivalent (MOD)** if valid in the same Kripke models.
- **locally equivalent (LOC)** if valid at the same states of every general frame.
- **algebraically equivalent (ALG)** if valid in the same modal algebras, equivalently, in the same general frames.
- **locally frame-equivalent (LFR)** if valid at the same states in every frame.
- **frame-equivalent (FR)** if valid in the same Kripke frames.
- **axiomatically equivalent (AX)** if the logics $\mathbf{K}+A$ and $\mathbf{K}+B$ have the same theorems. Equivalently, if $\mathbf{K}+A \vdash B$ and $\mathbf{K}+B \vdash A$.

We want to close the class of Sahlqvist formulae under as strong as possible equivalences, so as to preserve its effectiveness. Note, that **AX**, **LFR** and **FR** are not decidable [3]. Moreover, **AX** is not decidable even on the class SF [7]. Thus, we cannot safely close the class of Sahlqvist formulae under **AX**, if we want to preserve its effectiveness. Decidable equivalence closures of SF are currently under investigation.

3.3 Beyond Sahlqvist formulae: monadic inductive formulae

Here we present an extension of the class of Sahlqvist formulae, introduced for arbitrary polyadic languages in [23], [25].

DEFINITION 4. Let ML be a fixed monadic (multi-)modal language and $\#$ be a symbol not in ML. We define **box-forms of** $\#$ as follows:

- $\#$ is a box-form of $\#$.
- If $\mathbf{B}(\#)$ is a box-form of $\#$, then $\Box\mathbf{B}(\#)$ is a box-form of $\#$, for any box $\Box$ in ML.
- If $\mathbf{B}(\#)$ is a box-form of $\#$, and A is a positive formula, then $A \rightarrow \mathbf{B}(\#)$ is a box-form of $\#$;

Thus, box-forms are, up to semantic equivalence, of the type

$$\Box_1(A_1 \to \Box_2(A_2 \to \ldots \Box_n(A_n \to \#) \ldots))$$

where $\Box_1, \ldots, \Box_n$ are compositions of boxes in $\mathrm{ML}(\tau)$ and $A_1, \ldots, A_n$ are positive formulae.

A **box-formula of** p is the result $\mathbf{B}(p)$, of substitution of p for $\#$ in any box-form $\mathbf{B}(\#)$. The last occurrence of the variable p is the **head** of $\mathbf{B}(p)$ and every other occurrence of a variable in $\mathbf{B}(p)$ is **inessential** there.

DEFINITION 5. A **(monadic) regular formula** is any modal formula built from positive formulae and negated box-formulae by applying conjunctions, disjunctions, and boxes.[2]

The **dependency digraph** of a set of box-formulae $\mathcal{B} = \{\mathbf{B}_1(p_1), \ldots, \mathbf{B}_n(p_n)\}$ is a digraph $G = \langle V, E\rangle$, where $V = \{p_1, \ldots, p_n\}$ is the set of heads in $\mathcal{B}$, and $p_i E p_j$ iff p_i occurs as an inessential variable in a box from $\mathcal{B}$ with a head p_j; in such case we say that p_j **depends** on p_i. A digraph is called **acyclic** if it does not contain oriented cycles (including loops).

A **monadic inductive formula** is a monadic regular formula for which the dependency digraph of the set of all box-formulae occurring in it as subformulae, is acyclic.

EXAMPLE 6. The formula

$$D = p \wedge \Box(\Diamond p \to \Box q) \to \Diamond\Box\Box q \equiv \neg p \vee \neg\Box(\Diamond p \to \Box q) \vee \Diamond\Box\Box q$$

is an inductive formula, obtained as a disjunction of the negated box-formulae $\neg p$ and $\neg\Box(\Diamond p \to \Box q)$, and the positive formula $\Diamond\Box\Box q$. The dependency digraph of D over the set of heads $\{p, q\}$ has only one edge, from p to q.

Sahlqvist formulae are a simple particular case of inductive formulae, where all box-formulae are just boxed atoms $\Box_1 \ldots \Box_n p$, and hence the dependency digraph has no arcs at all. In fact, the class SF can be substantially generalized simply by replacing in the definition of classical monadic (multi-)modal Sahlqvist formulae 'boxed atoms' by 'box-formulae', and further requiring that the set of all box-formulae occurring as subformulae in the antecedent is *independent*, i.e. they all have different heads, and no head occurs inessentially in any of them. For instance,

$$\Diamond(\Box(\Box\Diamond q \to \Box\Box p_1) \wedge \Box\Box(\Diamond\Box q \to \Box(\Diamond q \to p_2))) \to \Diamond(p_1 \wedge \Box(\Diamond p_2 \vee q))$$

is not a Sahlqvist formula, but a simply generalized one.

[2]Just like a Sahlqvist formula, where the boxed atoms are now any box-formulae.

THEOREM 7 ([23, 25]). *All monadic inductive formulae are locally elementary and canonical.*

Proof. (Sketch) The first-order equivalents of inductive formulae can be computed by the method of substitutions, just like for Sahlqvist formulae, but inductively, following a partial ordering $\prec$ induced by the dependency digraph. More specifically, we first compute the minimal valuations of the variables which are not heads of box-subformulae. They only occur positively in the formula, so their minimal valuations are $\varnothing$. Then we proceed with the head variables in the box-subformulae, beginning with those which do not depend on any variables (i.e. the sources in the dependency graph). Thus, step by step we compute the minimal valuations of all head variables which only depend on variables whose valuations have already been computed. The acyclicity of the dependency graph of the inductive formula guarantees the successful completion of that procedure.

The canonicity follows similar lines, but needs some topological arguments. Let $A = A(q_1, \ldots, q_n)$ be a monadic inductive formula, $\mathfrak{F} = \langle \mathbf{F}, \mathbb{W} \rangle$ be a descriptive general frame such that $\mathfrak{F} \models A$, and V_m be the minimal valuation for $q_1, \ldots, q_n$. It suffices to prove that $\mathbf{F}, V_m \models A$.

Problem: *the minimal valuation need not be admissible* in $\mathfrak{F}$, so we cannot claim that $\mathfrak{F}, V_m \models A$. However, it suffices to show the following:

(C1) V_m is closed i.e. an intersection of admissible valuations.

(C2) For every closed valuation U in $\mathfrak{F}$ and a positive formula P, $U(P) = \bigcap_{U \preceq V} V(P)$ where the intersection ranges over all admissible valuations V which extend U.

(C1) is proved by $\prec$ -induction for every $V_m(q_j)$. (C2) is proved by structural induction on positive formulae, where the crucial step is Esakia's lemma (see e.g. [6], [2]) which essentially claims that infinite intersections of admissible sets in descriptive frames distribute over $\Diamond$, thus implying that $\Diamond$ is a closed operator in every topology over a descriptive frame. ■

REMARK 8. The conditions (C1) and (C2) in the proof above hold trivially for Kripke frames (i.e. full general frames), which allows for simultaneous treatment of first-order definability and canonicity of inductive formulae, in the spirit of Sambin and Vaccaro's proof [41].

EXAMPLE 9. The local first-order correspondent of the formula $D = \neg p \vee \neg \Box(\Diamond p \to \Box q) \vee \Diamond \Box \Box q$ is computed as follows. Since $p \prec q$, we first compute $V_m(p) = \{w\}$ where w denotes the current state in a frame with domain W. Then $V_m(q)$ is the minimal subset $Q(w)$ of W such that

$w \in \Box(\Diamond\{w\} \to \Box Q(w))$. This is equivalent to $\Diamond^{-1}\{w\} \in \Diamond\{w\} \to \Box Q_w$, i.e. $\Diamond^{-1}\{w\} \cap \Diamond\{w\} \subseteq \Box Q(w)$, i.e. $\Diamond^{-1}(\Diamond^{-1}\{w\} \cap \Diamond\{w\}) \subseteq Q(w)$. Thus, $V_m(q) = \Diamond^{-1}(\Diamond^{-1}\{w\} \cap \Diamond\{w\})$ and the (set-theoretic record of the) local first-order equivalent of D at w is

$$w \in \Diamond\Box\Box\Diamond^{-1}(\Diamond^{-1}\{w\} \cap \Diamond\{w\}).$$

This condition corresponds to the local first-order formula

$$\mathrm{FO}(D)(w) = \exists y(Rwy \wedge \forall z(R^2yz \to \exists u(Rwu \wedge Ruw \wedge Ruz))).$$

As mentioned earlier, the formula D is not frame equivalent to any Sahlqvist formula in the basic modal language (see section 5.1). In particular, we note that $\mathrm{FO}(D)$ is not a Kracht formula (see [32]). For more details on computing first-order equivalents of inductive formulae, see [23], [25].

3.4 Inductive formulae in polyadic modal languages

We now outline the generalization of monadic inductive formulae to arbitrary polyadic languages, introduced in [23], [25].

First, note that the inductive formulae *are not implications (like Sahlqvist implications), but composite polyadic boxes of special shape*. In order to extend the inductive formulae to polyadic modal languages we will adopt a somewhat non-orthodox view on these languages, by treating conjunctions and disjunctions as modal operators, and allowing compositions of modal operators, in PDL style. This treatment flattens the structure of polyadic modal formulae and makes their syntactic classification simpler.

Purely modal polyadic languages.

DEFINITION 10. A purely modal polyadic language $\mathcal{L}_\tau$ contains a countably infinite set propositional variables VAR, negation $\neg$, and a **modal similarity type** τ consisting of a set of **basic modal terms** (modalities) with pre-assigned finite arities, including a 0-ary modality ι_0, a unary one ι_1, and a binary one ι_2.

The intuition behind the 3 distinguished modalities above: ι_0 will be interpreted as the constant $\top$ and its dual as $\bot$; ι_1 will be the self-dual identity; ι_2 will be $\vee$ and its dual — $\wedge$.

DEFINITION 11. By simultaneous mutual induction we define the set of **modal terms** $MT(\tau)$ and their **arity function** ρ, and the set of **(purely) modal formulae** $MF(\tau)$ as follows:

(MT i) Every basic modal term is a modal term of the predefined arity.

(MT ii) Every constant formula (having no variables) is a 0-ary modal term.

(MT iii) If $n > 0$, $\alpha, \beta_1, \ldots, \beta_n \in MT(\tau)$ and $\rho(\alpha) = n$, then $\alpha(\beta_1, \ldots, \beta_n) \in MT(\tau)$ and $\rho(\alpha(\beta_1, \ldots, \beta_n)) = \rho(\beta_1) + \ldots + \rho(\beta_n)$.

Modal terms of arity 0 will be called **modal constants**.

(MF i) Every propositional variable is a modal formula.

(MF ii) Every modal constant is a modal formula.

(MF iii) If A is a formula then $\neg A$ is a formula;

(MF iv) If $A_1, \ldots, A_n$ are formulae, α is a modal term and $\rho(\alpha) = n > 0$, then $[\alpha](A_1, \ldots, A_n)$ is a modal formula.

Note that all formulae in a purely modal language are literals, boxes, or diamonds (negations of boxes). For technical purposes we extend the series of ι's with n-ary modalities ι_n: inductively as follows: $\iota_{n+1} = \iota_2(\iota_1, \iota_n)$ for $n > 1$.

Some notation on formulae:
$\langle\alpha\rangle(A_1, \ldots, A_n) := \neg[\alpha](\neg A_1, \ldots, \neg A_n)$; $\top := \iota_0, \bot := \neg\iota_0$;
$A \vee B := [\iota_2](A, B)$, $A \wedge B := \langle\iota_2\rangle(A, B)$, and respectively
$A_1 \vee \ldots \vee A_n := [\iota_n](A_1, \ldots, A_n)$, $A_1 \wedge \ldots \wedge A_n := \langle\iota_n\rangle(A_1, \ldots, A_n)$;
$A \to B := \neg A \vee B$; $A \leftrightarrow B := (A \to B) \wedge (B \to A)$.

For instance, the formula $D = p \wedge \Box(\Diamond p \to \Box q) \to \Diamond\Box\Box q$, after elimination of $\to$ and $\wedge$ becomes $\neg p \vee \neg\Box(\Box\neg p \vee \Box q) \vee \Diamond\Box\Box q$, which is represented in the polyadic language as:

$$D = [\iota_3](\neg p, \neg[\alpha(\iota_2(\alpha, \alpha))](\neg p, q), \langle\alpha\rangle[\alpha][\alpha]q),$$

where $[\alpha]$ corresponds to $\Box$.

Positive and negative occurrences of variables and positive and negative formulae are defined as usual.

Let us fix an arbitrary purely modal language $\mathcal{L}_\tau$. The semantics of $\mathcal{L}_\tau$ is a straightforward combination of the standard Kripke semantics for polyadic modal languages and PDL-type polymodal languages, taking into account the fact that conjunctions and disjunctions are now treated as modalities. This is accomplished by using the $(n+1)$-ary *identity relation* as the accessibility relation corresponding to $[\iota_n]$. Also, the notions of general frames and truth and validity in them generalize in a predictable way. (For details, see [23], [25].) The **standard translation** ST extends to polyadic languages with the clauses:

- $ST(\sigma) = R_\sigma(x)$ for every modal constant σ;
- $ST([\alpha](A_1, \ldots, A_n) = \forall \overline{y}(R_\alpha x y_1 \ldots y_n \to \bigvee_{i=1}^n ST(A_i)(y_i/x))$

Note that the propositional logical connectives $\wedge, \vee, \to$, as defined above, have their standard semantic interpretation. Therefore, the purely modal polyadic languages and the traditional ones are equally expressive.

Polyadic inductive formulae.

Given a purely modal polyadic language $\mathcal{L}_\tau$, an **essentially box-formula** in it is a modal formula of one of the following two types:

- **Headless boxes**, of the form $B = [\beta](N_1, \ldots, N_m)$, where β is any m-ary (composite) modal term, for $m \geq 1$, and $N_1, \ldots, N_m$ are negative formulae.

- **Headed boxes**, of the form $B = [\beta](p, N_1, \ldots, N_m)$, where β is any $(m+1)$-ary (composite) modal term, for $m \geq 0$, and $N_1, \ldots, N_m$ are negative formulae. The variable p is called the **head** of the box (here the head is put on the first place only for convenience of notations). In particular, p and $[\beta]p$ for any unary modal term β, are headed boxes.

All variables in an essentially box-formula except for the head of the formula (if any) are called **inessential variables** in that formula.

A **regular (polyadic) formula** is any modal constant (a 0-ary modal term) or a formula $A = [\alpha](\neg B_1, \ldots, \neg B_n)$ where α is an n -ary modal term and $B_1, \ldots, B_n$ are essentially box-formulae. The **dependency digraph** of A is a digraph $G = \langle \mathbf{V}_A, \mathbf{E}_A \rangle$ where $\mathbf{V}_A = \{p_1, \ldots, p_n\}$ is the set of heads in A, and $p_i \mathbf{E}_A p_j$ iff p_i occurs as an inessential variable in a formula from $B_1, \ldots, B_n$ with a head p_j.

A (**polyadic) inductive formula** is any regular formula A with an acyclic dependency digraph.[3] Note that the class of polyadic inductive formulae contains all monadic ones. In particular,
$D = [\iota_3](\neg p, \neg[\alpha(\iota_2(\alpha, \alpha))](\neg p, q), \langle\alpha\rangle[\alpha][\alpha]q)$ is a polyadic inductive formula. The class of polyadic inductive formulae can be further closed under conjunctions.

THEOREM 12 ([23],[25]). *All polyadic inductive formulae are locally elementary and canonical.*

The proof extends the one for monadic inductive formulae with the due technical overhead, but without essential conceptual complications.

[3] In [23] these were called 'polyadic Sahlqvist formulae'.

EXAMPLE 13. Computing the first-order equivalent to the polyadic inductive formula $B = [\mathbf{3}](\neg[\mathbf{1}]p, \neg[\mathbf{2}](\neg p, q), \langle\mathbf{1}\rangle[\mathbf{1}]q)$: the dependancy graph has one arc, $p \prec q$, so we first compute $V_m(p) = R_1(y_1)$. Then $V_m(q) = \{z | \exists s(R_2 y_2 sz \wedge R_1 y_1 s)\}$. Finally, $FO(B) =$
$\forall xy_1y_2y_3(R_{\mathbf{3}}xy_1y_2y_3 \to \exists v(R_{\mathbf{1}}y_3v \wedge \forall w(R_{\mathbf{1}}vw \to \exists s(R_{\mathbf{2}}y_2sw \wedge R_{\mathbf{1}}y_1s))))$.

Note that, once $V_m(p)$ is determined, $[\mathbf{2}](\neg p, q)$ can be regarded as a *unary box*: $[\alpha](q) = [\mathbf{2}](\neg V_m(p), q)$ where $\alpha = \mathbf{2}(\neg V_m(p), \iota_1)$ is a *unary parametrized modal term*, the relation of which can be accordingly computed: $R_\alpha xy$ iff $\exists s(R_2 xsy \wedge V_m(p)(s))$. This trick is essential in the proof of canonicity.

3.5 Inductive formulae in hybrid and reversive modal languages

Given a modal similarity type τ we extend the modal language $\mathcal{L}_\tau$ by adding *nominals* and the *universal modality* (see e.g. [18] or [2]), as well as inverse (residual) modalities, to obtain $\mathcal{L}^{u;\mathbf{n}}_{\tau,r}$. We now briefly consider inductive formulae in such languages, as developed in [24].

A **pure formula** in (a sublanguage of) $\mathcal{L}^{u;\mathbf{n}}_{\tau,r}$ is a formula that contains no propositional variables. Note that that *every pure formula is locally first-order definable.*

The definition of modal terms in $\mathcal{L}^{u;\mathbf{n}}_{\tau,r}$ extends the original one with the clause: *Every pure formula is a* 0*-ary modal term*, i.e. modal terms can be parameterized with pure formulae.

Inductive polyadic formulae in $\mathcal{L}^{u;\mathbf{n}}_{\tau,r}$ are defined as in purely modal polyadic languages, but on the extended set of modal terms.

A modal polyadic language is **reversive**[4] if, together with every n-ary modal term α it contains its 'inverses' $\alpha^1, \ldots, \alpha^n$ where for each $k = 1, \ldots, n$:

$$xR_{\alpha^k}y_1 \ldots y_k \ldots y_n \text{ iff } y_kR_\alpha y_1 \ldots x \ldots y_n.$$

In fact, it suffices to require this condition for the *basic* modal terms from the signature. An example of a reversive language is the basic tense language.

THEOREM 14 ([24]). *Every inductive formula in a reversive language with nominals,* $A = [\alpha](\neg H_1, \ldots, \neg H_n, P_1, \ldots, P_k)$*, is axiomatically equivalent*[5] *to a pure formula*

$$A^\circ = [\alpha](\neg c_1, \ldots, \neg c_n, Q_1, \ldots, Q_k)$$

where $c_1, \ldots, c_n$ *are nominals and* $Q_1, \ldots, Q_k$ *are obtained by means of effectively computable pure substitutions.*

---

[4] These are similar to Venema's *versatile* languages [45], but not quite the same.

[5] In an axiomatic system with additional rules for the nominals (see [18, 2]).

Note that the corresponding pure formula of a given Sahlqvist formula A codes the intended first-order equivalent of A.

Recall, that the right notion of canonicity in languages with nominals is $\mathcal{DI}$-persistence, so we will hereafter refer to this notion as *discrete canonicity.*

COROLLARY 15. *Every inductive formula in a reversive language with nominals is elementary and discretely canonical.*

Thus, proving the analogue of Sahlqvist's theorem for inductive formulae in reversive hybrid language with universal modality becomes merely a syntactic exercise.

3.6 Pushing the limits of the syntactic approach

Sahlqvist and inductive formulae do not exhaust the shapes of elementary canonical formulae. Other syntactic classes of such formulae include:

- All formulae of modal depth 1: van Benthem has classified them and proved their FO definability in [3]. Their canonicity can be verified by considering all cases.
- Consider *modal reduction principles* $M_1 p \to M_2 p$, where M_1 and M_2 are strings of boxes and diamonds. Again, van Benthem [3] has identified the first-order definable ones, and they are all easily seen to be canonical.
- *All* modal reduction principles on transitive frames.
- *Complex formulae,* see [44]. Example:

  $$\Diamond\Box(p \lor q) \land \Diamond\Box(p \lor \neg q) \land \Diamond\Box(\neg p \lor q) \to POS(p \to q, p \leftrightarrow q, p \land q)$$

  for any positive formula $POS(p_1, p_2, p_3)$.

  These are not inductive, but can be converted into inductive formulae by means of rather intricate substitutions.

All these syntactic classes are unstable under inessential transformations. For instance, the inductiveness can foolishly fail, e.g. in $\Box(p \leftrightarrow q) \to q$. What more can be done to extend the scope of the syntactic approach? Here are some further ideas:

- *Pre-processing* (see [24],[25]), using Boolean and modal equivalences, suitable substitutions, e.g. changing polarities or the special substitutions for complex formulae, normal forms, etc.

- In the definitions of Sahlqvist and inductive formulae, ' *positive*' and '*negative*' formulae can be replaced respectively by *upwards* and *downwards monotone*. By Lyndon's monotonicity theorem for modal logic (see e.g. [37]), such replacements preserve the formulae up to semantical equivalence, and hence preserve first-order definability and canonicity. Note that testing monotonicity of a modal formula is *decidable:* $B(p)$ is upwards monotone iff $\models B(p \wedge q) \rightarrow B(p)$ for q not occurring in $B(p)$. So, the definitions can be amended without loss of effectivity (though, at the expense of increased complexity).

- The definitions of both Sahlqvist and inductive formulae can be further extended by closing under effective equivalences, e.g. under semantic equivalence.

All these techniques push the limits of the syntactic approach farther. Still, it has firm boundaries, as it only produces *decidable* (usually, of fairly low complexity) classes of elementary canonical formulae. So, let us try something stronger...

4 Algorithmic approach to elementary canonical formulae

A natural strengthening of the syntactic approach is to develop algorithms that generate or identify elementary canonical formulae. Such algorithms need not be complete, i.e. successful for all elementary canonical formulae, but should always produce a correct result, if any, and thus define (recursively enumerable) classes of elementary canonical formulae. The roots of such an algorithm can be found in the *method of substitutions*, which originated from Sahlqvist's paper and was independently developed by van Benthem [4], see also [3]. That method was further sophisticated and extended by Simmons [43]. In particular, Simmons presented it in an explicitly algorithmic form which can be regarded as the first algorithm for producing elementary canonical formulae[6]. Simmons' algorithm works on a larger set of formulae, including non-elementary ones which have equivalents in FOL + Henkin quantifiers, such as all modal reduction principles. Note that the Sahlqvist-van Benthem substitution method works successfully only on formulae with fixed syntactical shape, such as Sahlqvist's formulae. In [23, 25] the method has been extended to work on inductive modal formulae, defining the appropriate substitutions by induction on the order determined by

[6] Though, Simmons' algorithm uses Skolemization, but does not involve a mechanism for unskolemization.

the dependency graph, but it still works only on formulae of the precise syntactic shape. The same applies for Simmons' method.

REMARK 16. The substitution method and Simmons' algorithm only establish first-order definability of modal formulae, but not their canonicity. However, if properly restricted and precisely specified, they can be shown to produce canonical formulae, by suitably modifying the proof of Sahlqvist's theorem.

Later in this section we show how the Sahlqvist-van Benthem substitution method can be considerably extended, by introducing the algorithm SQEMA [10] which works on arbitrary modal formulae and when it renders a successful result, it can be obtained by a sequence of suitable substitutions, computed in the course of the work of the algorithm. Before that, we present two other existing algorithms that can be used for computing first-order equivalents of modal formulae.

4.1 First-order definability as second-order quantifier elimination

Recall that the local validity of a modal formula $\phi = \phi(p_0, \ldots, p_n)$ in a pointed Kripke frame $(\mathfrak{F}, w)$ is expressed as

$$\mathfrak{F}, w \models \phi \text{ iff } \mathfrak{F}, w \models \forall P_0 \ldots \forall P_n \mathrm{ST}(\phi)(w/x),$$

where $\mathrm{ST}(\phi)(x)$ is the standard translation of ϕ over the free variable x.

Respectively, the global validity is expressed as

$$\mathfrak{F} \models \phi \text{ iff } \mathfrak{F} \models \forall P_0 \ldots \forall P_n \forall x \mathrm{ST}(\phi)(x).$$

Thus, the search for a local or global first-order equivalent of ϕ can be thought of as an attempt to *eliminate the universally quantified second-order variables* $P_0, \ldots, P_n$ and obtain a first-order formula equivalent to $\forall P_0 \ldots \forall P_n \forall x \mathrm{ST}(\phi)$. Sometimes it is more convenient to eliminate existentially quantified second-order variables. Then, the negation $\neg \forall P_0 \ldots \forall P_n \forall x \mathrm{ST}(\phi)$ is taken, and the resulting first-order formula is negated again.

Currently, there are two developed algorithms for second-order quantifier elimination: SCAN and DLS. They are both implemented and available online, and can be used to compute first-order equivalents of modal formulae.

SCAN

SCAN was developed in 1992 by Gabbay and Ohlbach [17]. Its current implementation, available online at http://www.mpi-sb.mpg.de/units/ag2/projects/SCAN/index.html, is based on the theorem prover OTTER. SCAN works on skolemized and clausified existentially quantified second-order formulae, and attempts to reduce them to equivalent first-order ones by generating sufficiently many logical consequences, and eventually keeping from the resulting set of formulae only those in which no second-order variables occur. As an input SCAN takes second-order formulae of the form $\exists Q_1 \dots \exists Q_k\ \psi$, where Q_i are predicate variables and ψ is a first-order formula. The algorithm involves three stages:

(i) transformation to clausal form and Skolemization;

(ii) a special kind of *constraint resolution (C-resolution)*, involving a *purity deletion rule* allowing one to delete 'used up' clauses.

(iii) reverse Skolemization (unskolemization), if possible.

SCAN can fail to produce a first-order equivalent of an input formula for one of two reasons: either (i) the C-resolution stage fails to terminate due to looping, or (ii) the C-resolution terminates, yielding a set of clauses in which the specified second-order variables are eliminated, but for which the Skolemization cannot be reversed.

SCAN can be used to compute the first-order equivalent of a modal formula by running it on the negation of its standard translation.

THEOREM 17 ([27]). SCAN *is successful on all Sahlqvist formulae.*[7]

That result can be supplemented with the following.

THEOREM 18 ([12]). SCAN *is successful on all polyadic inductive formulae.*

The latter does not formally subsume the former, as the proof that SCAN succeeds on a conjunction of Sahlqvist formulae is technically involved, while this step is avoided in the case of inductive formulae.

We conjecture that all modal formulae on which SCAN succeeds are canonical. This conjecture can be proved under some idealizing assumptions about SCAN, consistent with its specification in [14]. The difficulty in proving it for the actual implementation of the algorithm is that it does not match precisely the specification. We venture an even stronger conjecture, viz. that all modal formulae on which SCAN succeeds are locally equivalent to inductive formulae. A currently open question is if the class of modal formulae on which SCAN succeeds is decidable. Our conjecture is 'no'.

[7] This result holds under the assumption that SCAN uses inner Skolemization. In fact, the current implementation of SCAN does not always unskolemize successfully when run on Sahlqvist formulae, because it does not always employ inner Skolemization.

DLS

DLS was originally introduced by A. Szalas in 1993 and further developed by Doherty, Lukaszewics, and Szalas [13]. Its original implementation is available online at http://www.ida.liu.se/labs/kplab/projects/dls/, and a new one is currently being tested. DLS works on existentially quantified second-order formulae and always terminates, by either producing a first-order equivalent, or reporting failure. It is based on applying, after suitable preprocessing including Skolemization, the following lemma due to W. Ackermann:

LEMMA 19 ([1]). *For any first-order formula A not containing the predicate P and a first-order formula B, the following hold:*

$$\exists P(\forall \overline{x}(A(\overline{x}) \to P(\overline{x})) \wedge B(P)) \;\equiv\; B(A/P) \;\textit{(Downwards-Ackermann)},$$

if B is negative in P, and respectively,

$$\exists P(\forall \overline{x}(P(\overline{x}) \to A(\overline{x})) \wedge B(P)) \;\equiv\; B(A/P) \;\textit{(Upwards-Ackermann)},$$

if B is positive in P, where $B(A/P)$ is the result of uniform substitution of all occurrences of P in B by $A(\overline{x})$, with the arguments of each particular occurrence of P each time substituted for $\overline{x}$ in $A(\overline{x})$.

We note that the lemma can be strengthened by replacing 'positive' and 'negative', by 'upward monotone' and 'downward monotone' respectively, as these semantic properties are, in fact, used in the proof.

THEOREM 20 ([12]). DLS *is successful on all conjunctions of polyadic inductive formulae. In particular,* DLS *is successful on all Sahlqvist formulae.*

In fact, it turns out that DLS does not have to Skolemize on the translations of inductive formulae. We furthermore conjecture that for every modal formula on which DLS succeeds, it can succeed without skolemization.

We also claim that all modal formulae on which DLS succeeds are canonical, but, as with SCAN, the difficulty in proving such claim lies in the fact that the available specification of DLS in [29] is only partial, and the actual implementation does not match it precisely[8].

Finally, we conjecture, as for SCAN, that all modal formulae on which DLS succeeds are locally equivalent to inductive formulae.

Comparing SCAN *and* DLS

The constraint resolution rule of SCAN is based on a particular case of Ackermann's lemma. However, DLS does not subsume SCAN because it

[8] In fact, we have discovered a bug in the original implementation of DLS, which consists in reports of success ('true') in some cases where the algorithm should not succeed, and the formula to which it is applied is not equivalent to 'true'.

does not apply Ackermann's lemma repeatedly on the same variable, and does not use a purity deletion rule. Moreover, the C-resolution rule is not equivalence preserving.

From our practical experience with both algorithms, we find that SCAN is generally more flexible and syntax-tolerant (but easier to fool into looping) as it works on a low level, with formulae decomposed into a simple (clausal) form, and with simple rules (constraint resolution and factorization) applied repeatedly. On the other hand, DLS is more rigid and syntax-dependent, as it works on a high level, with only one, 'macro' rule (Ackermann's lemma).

In particular, neither of the implemented algorithms subsumes the other, but it seems that SCAN is generally more successful on modal formulae. For instance, it succeeds on the formula

$$\Box(p \leftrightarrow q) \rightarrow q,$$

on which DLS fails.

On the other hand, SCAN loops on the formula

$$\Box(p \vee \Box\neg p) \rightarrow \Diamond(p \wedge \Diamond\neg p),$$

(an example due to Szalas) on which, theoretically, DLS succeeds.[9]

We currently have no examples of *modal formulae* in the basic modal language, on which SCAN *loops* while DLS succeeds, but such examples can be constructed if the universal modality with its standard semantics is added to the language.

Ackermann's lemma and the method of substitutions.

We have obtained (see [10]) the following *modal version of Ackermann's lemma:*

LEMMA 21 (Ackermann, modal version). *For any modal formula A not containing p and a modal formula B, the following hold:*[10]

$$\exists p([\mathbf{u}](A \rightarrow p) \wedge B(p)) \equiv B(A/p) \text{ (Modal Downward-Ackermann),}$$

if B is negative (or stronger, downward monotone) in p, and respectively

$$\exists p([\mathbf{u}](p \rightarrow A) \wedge B(p)) \equiv B(A/p), \text{ (Modal Upward-Ackermann),}$$

if B is positive (or stronger, upward monotone) in p, where $B(A/p)$ is the result of uniform substitution of all occurrences of p in B by A, and $\equiv$ denotes local equivalence.

[9] But, the current implementation does not.

[10] These equivalences can be interpreted as follows: the right-hand side gives a condition for existence of a solution in p of the 'modal equation' on the left-hand side.

Indeed, under the conditions above, e.g. for the modal upwards-Ackermann lemma, the following holds: for every Kripke model M and $w \in M$, $M, w \models B(A/p)$ iff there is a model M' possibly differing from M only at the valuation of p, such that $M', w \models [\mathbf{u}](p \to A) \wedge B(p)$.

Note the contrapositive form of the downward Ackermann lemma, after replacing $\neg B$ with B:

$$\forall p([\mathbf{u}](A \to p) \to B(p)) \equiv B(A/p),$$

for any modal formula A not containing p, and a modal formula B which is upward monotone in p. This equivalence can be interpreted as follows: $[\mathbf{u}](A \to p) \to B(p)$ is valid in a given frame iff $B(P)$ is true for the '*minimal*' valuation satisfying the antecedent, viz. A. *This is precisely the technical idea at the heart of the substitution method of Sahlqvist and van Benthem!*

4.2 SQEMA: a new algorithm for computing elementary canonical formulae

In [10] we have introduced SQEMA: an algorithm for **S**econd-order **Q**uantifier **E**limination in **M**odal formulae, using **A**ckermann's lemma. It has the following basic features:

- Combines ideas from both DLS and SCAN and uses the modal version of Ackermann's lemma to eliminate the existentially quantified propositional variables.
- Works directly on (negated) modal formulae and decomposes them into sets of modal implications, called 'equations'.
- Does not introduce Skolem functions, but only Skolem constants, as nominals.
- Preserves formulae up to local frame equivalence.
- When successful, eventually produces a pure modal formula in a language, possibly extending the original one with nominals and inverse (reversive) modalities. The standard translation of this formula produces the corresponding first-order condition of the original formula.

The core algorithm.

Here we will present the algorithm on languages with unary modalities only. For the general case, see [12] and [11].

The input of SQEMA is a modal formula ϕ.

Step 1 Negate ϕ, eliminate $\rightarrow$ and $\leftrightarrow$, and rewrite in negation normal form. Then distribute diamonds and conjunctions over disjunctions as much as possible. The algorithm now proceeds on each disjunct ψ, separately, as follows:

Step 2 Rewrite as $i \rightarrow \psi$, where i is a fixed nominal, reserved to name the initial state. This is the only initial equation.

Step 3 Eliminate every variable p in which the system is monotone (upwards or downwards), by replacing it with $\top$ or $\perp$.

Step 4 If there are propositional variables remaining in equations of the system, choose to eliminate one, say p, the elimination of which has not been attempted yet.

If all remaining variables have been attempted and Step 5 has failed, backtrack and attempt another order of elimination.

If all orders of elimination and all remaining variables have been attempted and step 5 has failed, report failure.

If all propositional variables have been eliminated from the system, proceed to Step 6.

Step 5 The goal now is, by applying the transformation rules listed below, to rewrite the system of equations so that the Ackermann-rule becomes applicable with respect to the chosen variable p in order to eliminate it. Thus, the current goal is to transform the system into one in which every equation is either negative in p, or of the form $\alpha \rightarrow p$, with p not occurring in α, i.e. to 'extract' p and 'solve' for it.

If this fails, backtrack, change the polarity of p by substituting $\neg p$ for it everywhere, and attempt again to prepare for the Ackermann-rule.

If this fails again, or after the completion of this step, return to Step 4.

Step 6 If this step is reached it means that all propositional variables have been successfully eliminated from all systems resulting from the input formula. What remains now is to return the desired first-order equivalent. In each system, take the conjunction of all equations to obtain a formula pure, and form the formula $\forall \overline{y} \exists x_0 \mathrm{ST}(\neg \mathsf{pure})$, where $\overline{y}$ is the tuple of all occurring variables corresponding to nominals, but with y_i (corresponding to the designated current state nominal i) left free if the local correspondent is to be computed. Then take the conjunction of these translations over the systems on all disjunctive branches. For motivation of the correctness of this translation the reader is referred to the examples in the following subsection as well as the correctness proof in [10].

Return the result, which is the (local) first-order condition corresponding to the input formula.

The transformation rules

I. Rules for the logical connectives:

$\wedge$-rule: $\dfrac{\beta \to \gamma \wedge \delta}{\beta \to \gamma, \beta \to \delta}$ $\beta \to \gamma \wedge \delta$ $\Downarrow$ $\beta \to \gamma, \beta \to \delta$	$\Diamond$-rule: $j \to \Diamond\gamma$ $\Downarrow$ $j \to \Diamond k,\ k \to \gamma$ where k is a new nominal.
Left-shift $\vee$-rule: $\beta \to \gamma \vee \delta$ $\Downarrow$ $(\beta \wedge \neg\gamma) \to \delta$	Right-shift $\vee$-rule: $(\beta \wedge \neg\gamma) \to \delta$ $\Downarrow$ $\beta \to \gamma \vee \delta$
Left-shift $\Box$-rule: $\gamma \to \Box\delta$ $\Downarrow$ $\Diamond^{-1}\gamma \to \delta$	Right-shift $\Box$-rule: $\Diamond^{-1}\gamma \to \delta$ $\Downarrow$ $\gamma \to \Box\delta$

We will write Rjk as an abbreviation of $j \to \Diamond k$.

II. Auxiliary propositional rules:

1. Commutativity and associativity of $\wedge$ and $\vee$ (tacitly used).
2. Replace $\gamma \vee \neg\gamma$ with $\top$, and $\gamma \wedge \neg\gamma$ with $\bot$.
3. Replace $\gamma \vee \top$ with $\top$, and $\gamma \vee \bot$ with γ.
4. Replace $\gamma \wedge \top$ with γ, and $\gamma \wedge \bot$ with $\bot$.
5. Replace $\gamma \to \bot$ with $\neg\gamma$ and $\gamma \to \top$ with $\top$.
6. Replace $\bot \to \gamma$ with $\top$ and $\top \to \gamma$ with γ.

III. Polarity switching rule: Switch the polarity of every occurrence of a chosen variable p within the current system.

IV. Ackermann rule:

$$\left\Vert\begin{array}{l} \alpha_1 \to p, \\ \ldots \\ \alpha_n \to p, \\ \beta_1(p), \\ \ldots \\ \beta_m(p), \end{array}\right. \quad \Rightarrow \quad \left\Vert\begin{array}{l} \beta_1[(\alpha_1 \vee \ldots \vee \alpha_n)/p], \\ \ldots \\ \beta_m[(\alpha_1 \vee \ldots \vee \alpha_n)/p]. \end{array}\right.$$

where p does not occur in $\alpha_1, \ldots, \alpha_n$ and each β_i is negative in p.[11]

4.3 Examples

The best way to get a feel of the workings of the algorithm is perhaps to consider an example or two. (For more, see [10].)

EXAMPLE 22. We take as input the formula $\Diamond\Box p \to \Box\Diamond p$.

The initial system of equations is

$$\| i \to (\Diamond\Box p \wedge \Diamond\Box\neg p)$$

Applying the $\wedge$-rule gives

$$\left\| \begin{array}{l} i \to \Diamond\Box p \\ i \to \Diamond\Box\neg p \end{array} \right.$$

Applying the $\Diamond$-rule to the first equation yields:

$$\left\| \begin{array}{l} Rij \\ j \to \Box p \\ i \to \Diamond\Box\neg p \end{array} \right.$$

and then applying the Left-shift $\Box$-rule:

$$\left\| \begin{array}{l} Rij \\ \Diamond^{-1} j \to p \\ i \to \Diamond\Box\neg p \end{array} \right.$$

The Ackermann rule is now applicable, yielding the system

$$\left\| \begin{array}{l} Rij \\ i \to \Diamond\Box\neg(\Diamond^{-1} j) \end{array} \right.$$

Taking the conjunction of the equations gives

$$Rij \wedge (i \to \Diamond\Box\neg(\Diamond^{-1} j)).$$

Negating we obtain

$$Rij \to (i \wedge \Box\Diamond\Diamond^{-1} j),$$

which, translated, becomes

$$\forall y_j \exists x_0 [Ry_i y_j \to (x_0 = y_i) \wedge \forall y (Rx_0 y \to \exists u (Ryu \wedge \exists v (Rvu \wedge v = y_j)))],$$

[11] As already discussed, this rule can be strengthened by replacing 'negative' with 'downwards monotone', but this brings a higher complexity price.

and simplifies to

$$\forall y_j[Ry_iy_j \to \forall y(Ry_iy \to \exists u(Ryu \wedge Ry_ju))]$$

defining the Church-Rosser property, as expected.

EXAMPLE 23. We take as input the (non-inductive) formula

$$\Box(\Box p \leftrightarrow q) \to p$$

on which both SCAN and DLS fail.

This yields the initial equation

$$\Big\| \; i \to \Box((\Diamond\neg p \vee q) \wedge (\neg q \vee \Box p)) \wedge \neg p$$

Choose q to eliminate first. Applying the $\wedge$-rule and the Left-shift $\Box$-rule:

$$\left\| \begin{array}{l} \Diamond^{-1}i \to (\Diamond\neg p \vee q) \\ \Diamond^{-1}i \to (\neg q \vee \Box p) \\ i \to \neg p \end{array} \right.$$

Applying the Left Shift $\vee$-rule to the first equation yields

$$\left\| \begin{array}{l} (\Diamond^{-1}i \wedge \Box p) \to q \\ \Diamond^{-1}i \to (\neg q \vee \Box p) \\ i \to \neg p \end{array} \right.$$

to which the Ackermann-rule is applicable with respect to q. This gives

$$\left\| \begin{array}{l} \Diamond^{-1}i \to (\neg\Diamond^{-1}i \vee \neg\Box p \vee \Box p) \\ i \to \neg p \end{array} \right. .$$

The first equation is now a tautology and may be removed, yielding the system

$$\Big\| \; i \to \neg p$$

in which p may be replaced by $\bot$ since it occurs only negatively, resulting in the system

$$\Big\| \; \top \quad .$$

Negating we obtain $\bot$.

Some results and comments on SQEMA*.*

THEOREM 24 ([10]).

1. SQEMA *is sound: if successful, it produces a first-order formula locally frame equivalent to the input modal formula.*
2. SQEMA *is successful on all conjunctions of inductive formulae. In particular,* SQEMA *is successful on all Sahlqvist formulae.*
3. *All modal formulae on which* SQEMA *succeeds are canonical.*

Note that the original Sahlqvist's theorem and its extension to inductive formulae now follow from the results above.

Again, we conjecture that all modal formulae on which SQEMA succeeds are locally equivalent to inductive formulae.

How does SQEMA compare to SCAN and DLS on modal formulae? We believe that it is stronger than both, but this claim, if correct, can only be proved if precise descriptions of the specifications and implementations of SCAN and DLS are available.

Finally, we note that SQEMA is amenable to various extensions, e.g. with a recursive version of the Ackermann rule, which enables computation of correspondents of modal formulae in FO+LFP, see [11],[26].

4.4 The power and limits of the algorithmic approach

The algorithmic approach is certainly more powerful as a generator of elementary canonical formulae than the syntactic approach. It produces effectively enumerable classes of elementary canonical formulae, which in general need not be decidable.

The different algorithms discussed here: method of substitutions, Simmons' algorithm, SCAN, DLS, and SQEMA, have different computing powers and scope of applicability on modal formulae. Yet, we believe that the algorithmic approach, if developed to its full potential, will generate a natural class of *algorithmically elementary canonical formulae.* The major challenge of this research area is to develop such an optimal algorithm.

5 Model-theoretic aspects of elementary canonical formulae

In this section we briefly discuss semantic characterizations of elementary canonical formulae. The main model-theoretic tool we use for such characterizations is *persistence.* While canonicity *is defined* in terms of persistence, first-order definability can only be approximated in such a way. The approximation which we discuss here, due to van Benthem [3], defines a large and natural class of elementary modal formulae.

5.1 Sahlqvist formulae and ample-persistence

How can one prove that a given elementary canonical formula is *not* equivalent to a Sahlqvist formula in any reasonable sense? Here is a method, introduced in [25], based on a special kind of persistence.

DEFINITION 25. A general frame $\langle W, R, \mathbb{W}\rangle$ is **ample** if for every $w \in W$ and $n \in \mathbf{N}$, $R^n(w) = \{u \mid wR^n u\} \in \mathbb{W}$.

Note that every ample general frame is discrete, for $R^0(w) = \{w\}$.

DEFINITION 26. A modal formula A is **locally a-persistent** if it is locally persistent with respect to every ample general frame, i.e. for every such frame $\mathfrak{F} = \langle F, \mathbb{W}\rangle$, where $F = \langle W, R\rangle$, and $w \in W$,

$$\mathfrak{F}, w \models A \text{ iff } F, w \models A.$$

The following can be proved by inspection of the minimal valuations corresponding to Sahlqvist formulae.

LEMMA 27. *Every Sahlqvist formula in the basic modal language is locally* a-*persistent.*

PROPOSITION 28 ([25]). *The inductive formula*

$$D = p \wedge \Box(\Diamond p \to \Box q) \to \Diamond\Box\Box q$$

is not (even globally) a-*persistent.*

COROLLARY 29. *The formula D is not frame equivalent to any Sahlqvist formula in the basic modal language.*

5.2 van Benthem formulae and the limits of the substitutions method

Let FO be the first-order language for Kripke frames, and $\beta(\overline{x})$ be a FO-formula with unary predicates $P_1, \ldots, P_n$, such that the variables $\overline{x}$ do not occur bound in β and the variables $z_1, \ldots, z_k, y$ do not occur in β at all.

A **universally parameterized** FO-**substitution instance** of β is any FO-formula

$$\forall z_1 \ldots \forall z_k \beta[\sigma_1/P_1, \ldots, \sigma_n/P_n]$$

obtained from β by selecting FO-formulae $\sigma_i = \sigma_i(\overline{x}, z_1, \ldots, z_k, y)$ for $i = 1, \ldots, n$, substituting $\sigma_i[x/y]$ for every occurrence of $P_i x$, and then universally quantifying over $z_1, \ldots, z_k$.

Let $\Theta(\beta)$ be the set of all universally parametrised FO-substitution instances of β.

DEFINITION 30. A modal formula $\phi = \phi(p_1, \ldots, p_n)$ is a **van Benthem formula** if

$$\Theta(\mathrm{ST}(\phi; x_0)) \models \forall P_1 \ldots \forall P_n \mathrm{ST}(\phi; x_0).$$

We let VB denote the class of van Benthem formulae (defined slightly differently by van Benthem himself in [3], as the class M_1^{sub}).

THEOREM 31 ([28]. Essentially first proved in [3]). *A modal formula is locally $\mathcal{E}$-persistent iff it is a van Benthem formula.*

Since, by compactness, all van Benthem formulae are locally first-order definable, we obtain the following.

COROLLARY 32. *Every locally $\mathcal{E}$-persistent modal formula is locally elementary.*

Some burning questions arise now:

- Is every van Benthem formula canonical ($\mathcal{D}$-persistent)?

  Sadly, no: van Benthem's incomplete formula $\mathbf{vB} = \Box\Diamond\top \to \Box(\Box(\Box p \to p) \to p)$ (see [3] or [2]) is a counter-example.

- Is every elementary canonical formula a van Benthem formula?

  Even more sadly, no: $(\Box p \to \Box\Box p) \wedge \Box(\Box p \to \Box\Box p) \wedge (\Box\Diamond p \to \Diamond\Box p)$ is a locally elementary canonical formula, but not locally $\mathcal{E}$-persistent [3].

Still, van Benthem formulae (in his own words) 'neatly delimit the range of the method of substitutions', and provide a natural and important upper bound for the class of elementary canonical formulae.

THEOREM 33 ([4]). *The set VB is recursively enumerable.*

Thus, there is an algorithm generating all van Benthem formulae, and essentially based on the method of substitutions. It is a natural challenge to develop a practical one.

An even more challenging question is whether the set of canonical van Benthem formulae is recursively enumerable, and if so, to construct a generating algorithm for it.

5.3 Elementary canonical formulae and persistence in reversive languages with nominals

The leading problem of our model-theoretic approach to elementary canonical formulae is to characterize them in terms of a natural persistence property. We do not have (yet) a solution to this problem for the basic modal language, but we do for '*rich enough*' languages, viz. reversive languages with nominals. Recall again, that the natural notion of canonicity in languages with nominals is ‘discrete canonicity’, i.e. $\mathcal{DI}$-persistence.

THEOREM 34 ([25]). *For every modal formula A in a reversive language with nominals, the following are equivalent:*

1. *A is locally $\mathcal{DI}$-persistent.*

2. *A is locally equivalent to an inductive formula.*

3. *A is locally equivalent in the class of discrete frames to a pure formula.*

5.4 Topological perspective on elementary canonical formulae

Following topological ideas going back to Sambin and Vaccaro [41], we show in [25] how first-order definability and canonicity of inductive formulae can be established in a uniform way. In a similar way we give a simultaneous proof of the correctness and the canonicity of the algorithm SQEMA in [10]. Here are the key points of these arguments. The formulae for which first-order definability and canonicity (persistence) is to be established, are transformed into ‘simple’ ones, for which these properties are immediate. Typically, these are pure formulae in a reversive hybrid extension of the original language. Such transformation can be semantic (as is the case of inductive formulae in [25]), deductive (inductive formulae in hybrid languages, in [24]), or algorithmic (the formulae on which SQEMA succeeds, in [10]), but in any case they preserve the desired properties. For first-order definability, such preservation is proved by a direct semantic argument on Kripke frames, but for the general frames over which the persistence is to be proved (e.g. descriptive frames), the argument involves suitable topological closure properties of the modal operators in the extended language, considered as operators in the topologies on the general frames of the original language. These closure properties guarantee that the formulae under consideration (Sahlqvist, inductive, SQEMA) allow the semantic argument proving preservation on Kripke frames to be simulated for them on e.g. descriptive frames, thus implying $\mathcal{D}$-persistence. Proving the desired topological behavior in all cases we have studied crucially depends on the effective (syntactic, or algorithmic) nature of these formulae.

This topological approach is still open to further development, and the main aim of that approach is to find a sufficiently general argument which applies to all elementary canonical formulae, regardless of their syntactic features.

6 Concluding remarks: closing about elementary canonical formulae

Each of the syntactic, algorithmic, and model-theoretic approaches provides a hierarchy of approximations of the class of elementary canonical formulae, but none of them seems to yield both a practical and precise characterization yet. It is currently unknown if such characterization, better than the definition itself, exists at all. In particular, we do not know the complexity of the class of elementary canonical formulae, nor that of the class of first-order formulae definable by elementary canonical modal formulae.

We note that there are interesting and important cases of canonical modal formulae that are not elementary, see e.g. [19], [46]. Moreover, it has recently been established in [21] that a canonical modal logic need not be complete with respect to any elementary class of frames. Thus, first-order definability and canonicity are not as closely related as it was been conjectured by Fine in the 1970's. It is therefore natural to extend this study along each of these properties separately. In that respect, we should also mention the *algebraic* approach to canonicity, developed by Jónsson who gave an algebraic proof of Sahlqvist's theorem in [30].

Finally, we should mention an important family of modal formulae and logics axiomatized with such formulae, for which elementariness and canonicity coincide. These are the *subframe formulae and logics* introduced by Fine [15] and further extended to *cofinal subframe formulae and logics* by Zakharyaschev [48, 6][12].

With each finite transitive general frame $\mathfrak{F}$ one can associate (see [48, 6]) a *subframe formula* $\alpha(\mathfrak{F}, \emptyset)$, and a *cofinal subframe formula* $\alpha(\mathfrak{F}, \emptyset, \bot)$, such that any transitive general frame $\mathfrak{G}$ refutes $\alpha(\mathfrak{F}, \emptyset)$ (resp. $\alpha(\mathfrak{F}, \emptyset, \bot)$) if and only if $\mathfrak{G}$ is subreducible (resp. cofinally subreducible) to $\mathfrak{F}$ by way of a bounded morphism. A normal extension of $K4$ is a *subframe logic* if it can be axiomatized over $K4$ by a set of subframe formulae $\{\alpha(\mathfrak{F}_i) : i \in I\}$, for some family of transitive general frames $\{\mathfrak{F}_i : i \in I\}$. *Cofinal subframe logics* are defined similarly.

It turns out that, on transitive frames, a cofinal subframe formula is elementary iff it is $\mathcal{D}$-persistent; for subframe formulas these are equivalent to $\mathcal{R}$-persistence, as well. The same equivalences apply to (cofinal) subframe logics.

Similar results were established by Wolter [47] for modal formulae preserved in subframes, and normal modal logics L characterized by classes of (general) frames closed under taking subframes. For these, elementariness, $\mathcal{D}$-persistence, and $\mathcal{R}$-persistence coincide.

[12] Note that the term 'canonical formulae' in [48] has different meaning from the commonly used one in the present paper.

Acknowledgments

V. Goranko acknowledges the financial support provided by the National Research Foundation of South Africa and the Faculty of Science at Rand Afrikaans University. Part of this work was completed during the visit of Goranko and Conradie to the Department of Computer Science at the University of Manchester, the financial support for which was provided by a research grant from the British Research Council for Science and Engineering. We are grateful to Renate Schmidt for obtaining that grant and organizing our visits, and for some useful comments on SQEMA. D. Vakarelov was partially supported by the EU COST project 274 TARSKI and the project RILA-12 sponsored by the Bulgarian Ministry of Science and Education. We thank an anonymous reader for some useful comments and references.

BIBLIOGRAPHY

[1] Ackermann, W., Untersuchung über das Eliminationsproblem der mathematischen Logic. *Mathematische Annalen,* 110:390-413, 1935.

[2] Blackburn, P., M. de Rijke, Y. Venema, **Modal Logic**, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 2001.

[3] van Benthem J.F.A.K., **Modal Logic and Classical Logic**, Bibleapolis, 1985.

[4] van Benthem J.F.A.K., **Modal Correspondence Theory**, Ph.D. Thesis, Mathematisch Instituut & Instituut voor Grondslagenonderzoek, Univ. of Amsterdam, 1976.

[5] van Benthem J.F.A.K., Minimal predicates, fixed points, and definability, J. Symb. Logic, to appear.

[6] Chagrov, A. and M Zakharyaschev, **Modal Logic**, Clarendon Press, Oxford, 1997.

[7] Chagrov, A. and M. Zakharyaschev, Sahlqvist formulae are not so elementary, **Logic Colloquium'92**, L. Csirmaz, D. Gabbay and M. de Rijke (eds.), CSLI Publications, Stanford, 1995, 61-73.

[8] Chagrov, A., F. Wolter, and M. Zakharyaschev, Advanced modal Logic, in: Handbook of Philosophical Logic, 2nd edition, vol 3, Kluwer, 2001, 83–266.

[9] Conradie, W., Decidable Equivalences of Modal Formulae, 2005, in preparation.

[10] Conradie, W., V. Goranko, and D. Vakarelov: Algorithmic correspondence and completeness in modal logic. I. The algorithm SQEMA, 2004, submitted.

[11] Conradie, W., V. Goranko, and D. Vakarelov: Algorithmic correspondence and completeness in modal logic. II. Extensions of the algorithm SQEMA, 2005, in preparation.

[12] Conradie, W. and V. Goranko: Algorithmic Classes of Elementary Canonical Formulae, 2005, in preparation.

[13] Doherty, P, W. Lukaszewics, and A. Szalas. Computing circumscription revisited: A reduction algorithm, *Journal of Automated Reasoning*, 18(3):297–336, 1997.

[14] Engel, T. Quantifier Elimination in Second-Order Predicate Logic, Diploma Thesis, Univ. of Saarbruecken, 1996.

[15] Fine, K., Logics containing K4, part II. *Journal of Symbolic Logic, 50:619-651, 1985.*

*[16] Gabbay, D., I. Hodkinson, and M. Reynolds, **Temporal Logic: Mathematical Foundations and Computational Aspects**, Vol. 1, Clarendon Press, Oxford, 1994.*

[17] Gabbay, D. and H.-J. Ohlbach, Quantifier elimination in second-order predicate logic, South African Computer Journal, vol. 7, 1992, 35-43.

[18] Gargov, G. and V. Goranko, Modal Logic with Names, Journal of Philosophical Logic, 22/6 (1993), pp. 607-636.

[19] Ghilardi, S. and G. Meloni, Constructive canonicity in non-classical logics, Annals of Pure and Applied Logic, 86 (1997), pp. 1-32.

[20] Goldblatt, R., *First-order definability in modal logic*, *Journal of Symbolic Logic*, 40(1975), 35-40.

[21] Goldblatt, R., I. Hodkinson, and Y. Venema, *Erdös graphs resolve Fine's canonicity problem*, Bull. of Symb. Logic, 10 (2), 2004, 186-208.

[22] Goranko, V., *Axiomatizations with Context Rules of Inference in Modal Logic*, *Studia Logica*, 61(2), 1998, 179-197.

[23] Goranko, V. and D. Vakarelov, *Sahlqvist formulae Unleashed in Polyadic Modal Languages*, **Advances in Modal Logic, vol. 3**, World Scientific, Singapore, 2002, pp. 221-240.

[24] Goranko, V. and D. Vakarelov, *Sahlqvist formulae in Hybrid Polyadic Modal Languages*, *J. of Logic and Computation*, vol. 11(5), 2001, 737-754.

[25] Goranko, V. and D. Vakarelov, *Elementary Canonical Formulae: Extending Sahlqvist's Theorem*, 2003, to appear in Annals of Pure and Applied Logic.

[26] Conradie, W., Goranko, V. and D. Vakarelov, *Computing equivalents of modal formulae in FO(LFP)*, in preparation, 2005.

[27] Goranko, V., U. Hustadt, R. Schmidt, and D. Vakarelov. *SCAN is complete for all Sahlqvist formulae*, in: Relational and Kleene-Algebraic Methods in Computer Science (Proc. of RelMiCS 7). LNCS 3051, Springer, 2004, 149-162.

[28] Goranko, V., and M. Otto, *Model Theory of Modal Logic*, in: Handbook of Modal Logic, Kluwer, 2005, to appear.

[29] Gustafsson, J., *An Implementation and Optimization of an Algorithm for Reducing Formulae in Second-Order Logic*. Technical Report LiTH-MAT-R-96-04. Dept. of Mathematics, Linkoping University, Sweden, 1996.

[30] Jónsson, B., *On the Canonicity of Sahlqvist Identities*, *Studia Logica*, 53 (1995), pp. 473-491.

[31] Jónsson, B. and A. Tarski, *Boolean Algebras with Operators, Part 1*, *American J. of Mathematics* , 73 (1952), 891-939.

[32] Kracht, M., *How Completeness and Correspondence Theory Got Married*, in: **Diamonds and Defaults**, M. de Rijke (ed.), Kluwer, Synthese Library, 1993, 175-214.

[33] Kracht, M., **Tools and Techniques in modal Logic**, Elsevier, Amsterdam, 1999.

[34] Nonnengart, A., H. J. Ohlbach, and A. Szalas. *Quantifier elimination for second-order predicate logic*, in: **Logic, Language and Reasoning: Essays in honour of Dov Gabbay**, Part I, H. J. Ohlbach and U. Reyle (eds.), Kluwer, 1997.

[35] Nonnengart, A. and A. Szalas. *A fixpoint approach to second-order quantifier elimination with applications to correspondence theory*, in: **Logic at Work, Essays dedicated to the memory of Helena Rasiowa**, E. Orlowska (ed.), Springer Physica-Verlag, 1998, pp. 89-108.

[36] de Rijke, M., *How not to generalize Sahlqvist's Theorem*, Technical Note, ILLC, 1992.

[37] de Rijke, M., **Extending Modal Logic**, Ph.D. thesis, ILLC, University of Amsterdam, ILLC Dissertation Series 1993-4, 1993.

[38] de Rijke, M. and Y. Venema, *Sahlqvist's Theorem For Boolean Algebras with Operators with an Application to Cylindric Algebras*, *Studia Logica*, 54 (1995), 61-78.

[39] Sahlqvist, H., *Correspondence and completeness in the first and second-order semantics for modal logic*, **Proc. of the 3rd Scandinavial Logic Symposium, Uppsala 1973**, S. Kanger (ed.), North-Holland, Amsterdam, 1975, 110-143.

[40] Sambin, G. and V. Vaccaro, *Topology and Duality in Modal Logic*, *Annals of Pure and Applied Logic*, 37(1988), 249-296.

[41] Sambin, G. and V. Vaccaro, *A New Proof of Sahlqvist's Theorem on Modal Definability and Completeness*, *Journal of Symbolic Logic*, 54(1989), 992-999.

[42] Szałas, A., *On the correspondence between modal and classical logic: An automated approach*, 3(6):605–620, 1993.

[43] Simmons, H., *The Monotonous Elimination of Predicate Variables*. *Journal of Logic and Computation*, 4(1):23–63, 1994.

[44] *Vakarelov, D., Modal Definability in Languages with a Finite Number of Propositional Variables and a New extension of the Sahlqvist's Class, Advances in Modal logic, vol. 4, 499-518.*

[45] *Venema, Y., Derivation rules as anti-axioms in modal logic, Journal of Symbolic Logic,* ***58*** *(1993), 1003– 1034.*

[46] *Venema, Y., Canonical pseudo-correspondence, in:* ***Advances in Modal Logic, vol. 2****, M. Kracht, M. de Rijke, H. Wansing, and M. Zakharyaschev (eds.), CSLI Publications, Stanford, 2000, 421-430.*

[47] *Wolter, F., The structure of lattices of subframe logics, Annals of Pure and Applied Logic, 86 (1997), 47-100.*

[48] *Zakharyaschev, M. V., Canonical formulas for K4. Part II: Confinal subframe logics, Journal of Symbolic Logic, 61:421-449, 1996.*

Willem Conradie

Department of Mathematics and Statistics, University of Johannesburg
PO Box 524, Auckland Park 2006, Johannesburg, South Africa

wec@rau.ac.za

Valentin Goranko

Department of Mathematics and Statistics, University of Johannesburg
PO Box 524, Auckland Park 2006, Johannesburg, South Africa

vfg@rau.ac.za

Dimiter Vakarelov

Department of Mathematical Logic with Laboratory for Applied Logic,
Faculty of Mathematics and Computer Science, Sofia University
blvd James Bouchier 5, 1126 Sofia, Bulgaria

dvak@fmi.uni-sofia.bg