Conservative extensions in modal logic

S. Ghilardi, C. Lutz, F. Wolter and
M. Zakharyaschev

ABSTRACT. Every modal logic L gives rise to the consequence relation $\varphi \models_L \psi$ which holds if, and only if, ψ is true in a world of an L-model whenever φ is true in that world. We consider the following algorithmic problem for L. Given two modal formulas φ_1 and φ_2, decide whether $\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in the sense that whenever $\varphi_1 \wedge \varphi_2 \models_L \psi$ and ψ does not contain propositional variables not occurring in φ_1, then $\varphi_1 \models_L \psi$. We first prove that the conservativeness problem is coNExpTime-hard for all modal logics of unbounded alternatives (which have rooted frames with more than N successors of the root, for any $N < \omega$). Then we show that this problem is (i) coNExpTime-complete for **S5** and **K**, (ii) in ExpSpace for **S4** and (iii) ExpSpace-complete for **GL.3** (the logic of finite strict linear orders). The proofs for **S5** and **K** use the fact that these logics have uniform interpolants of exponential size.

1 Introduction

A theory T_2 is said to be a *conservative extension* of a theory T_1 if any consequence of T_2, which only uses symbols from T_1, is a consequence of T_1 as well. This notion plays an important role in mathematical logic and the foundations of mathematics. For example, the result that Bernays–Gödel set theory BG (or BGC) is a conservative extension of Zermelo–Fraenkel set theory ZF (or ZFC) implies the relative consistency of BG(C): if ZF(C) is consistent then BG(C) is also consistent.

Rather surprisingly, in modal logic the notion of conservative extension has hardly been investigated. Indeed, modal theories—similarly to first-order theories—have become fundamental tools for representing various domains. For example, in epistemic logic, modal theories represent the (possibly introspective) knowledge of an agent; in temporal logic, theories serve as specifications of concurrent systems; in description logic, theories (called TBoxes) are ontologies used to fix the terminology of an application domain, etc. In all these examples, the notion of conservative extension can be used to compare different theories and derive important information about their relation to each other: for instance, a temporal specification T_2 can be regarded as a 'safe' refinement of another temporal specification T_1 if, and only if, T_2 is a conservative extension of T_1 (see, e.g., [15]). A description logic ontology T_2 is a 'safe' extension of another description logic ontology T_1 if, and only if, T_2 is a conservative extension of T_1 (see [1, 6]).

One of the main reasons for using modal logic instead of full first-order logic in the applications above is that reasoning in modal logic is quite often decidable. To employ the notion of conservative extension for modal logics,

Advances in Modal Logic, Volume 6. © *2006, the authors.*

it is therefore crucial to analyse the algorithmic problem of deciding whether one modal theory is a conservative extension of another modal theory.

In this paper, we investigate the notion of conservative extension for a number of basic modal logics and, in particular, determine the computational complexity of the conservativeness problem for these logics.

In modal logic, the notion of conservative extension depends on the consequence relation we are interested in. Of particular importance are the 'local consequence relation' according to which a formula φ follows from a formula ψ if φ is true *in every world* where ψ is true, and the 'global consequence relation' according to which φ follows from ψ if φ is true *everywhere in a model* whenever ψ is true everywhere in this model (see, e.g., [10]). In this paper, we concentrate on the local consequence relation. Some information about the global consequence relation is provided in the final section.

We begin by showing that deciding non-conservativeness is NExpTime-hard for all modal logics of unbounded alternatives (which have rooted frames with more than N successors of the root, for any $N < \omega$). This result covers almost all standard modal logics, for example, **K**, **S4**, **S5**, and **S4.3**. Thus, deciding conservativeness turns out to be much harder than deciding satisfiability. We also observe that for tabular modal logics (see, e.g., [2, 18]) non-conservativeness is $\mathrm{NP}^{\mathrm{NP}}$-complete, which coincides with the complexity of non-conservativeness in classical propositional logic [11]. The proof of this result and many other proofs in this paper are based on some elementary facts connecting conservativeness with bisimulations.

Next, to warm up, we consider the modal logic **S5** and show that in this case non-conservativeness is NExpTime-complete by proving that one can construct a uniform interpolant of exponential size in exponential time (in the size of a given formula) and using the fact that **S5**-satisfiability is decidable in NP. This proof is based on a general result connecting conservativeness with uniform interpolation (see [12, 17] for a discussion of this variant of interpolation).

A slightly different technique is used to prove that for **K** non-conservativeness is NExpTime-complete. Here we employ a result from [14] according to which there exist uniform interpolants for **K** of (only) exponential size, and then provide a direct algorithm deciding non-conservativeness without computing the uniform interpolant.

After that we consider the non-conservativeness problem for **S4** and establish an ExpSpace upper bound. As this upper bound (probably) does not match the NExpTime lower bound, we leave the exact complexity as an open problem. The logic **S4** does not have uniform interpolation [8], and therefore a 'direct' algorithm had to be found. Similar arguments show that non-conservativeness is decidable in ExpSpace for **K4**, **Grz** and **GL**.

Finally, we prove that conservativeness is ExpSpace-complete for **GL.3**, the logic of finite strict linear orders. Here we again give direct proofs for both lower and upper bounds. Similar proofs show ExpSpace-completeness of conservativeness for **K4.3** and **S4.3**.

In many cases, full proofs are omitted. The reader can find them in [7].

2 Preliminaries

We consider the language $\mathcal{ML}$ of propositional unimodal logic with countably many propositional variables $p_1, p_2, \dots$, the Booleans $\wedge$ and $\neg$, and the modal operator $\Box$. Other Boolean operators and the modal diamond $\Diamond$ are defined as usual. Given an $\mathcal{ML}$-formula φ, we denote by $var(\varphi)$ the set of propositional variables occurring in φ.

A (*Kripke*) *frame* $\mathfrak{F} = (W, R)$ is a nonempty set W of points (or worlds) with a binary relation R on it. A (*Kripke*) *model* $\mathfrak{M} = (\mathfrak{F}, \mathfrak{V})$ consists of a frame and a *valuation* $\mathfrak{V}$ giving truth-values to propositional variables in the worlds of W. The *satisfaction relation* '$(\mathfrak{M}, w) \models \varphi$' between *pointed models* $(\mathfrak{M}, w)$ (where $w \in W$) and $\mathcal{ML}$-formulas φ is defined as usual. (If $\mathfrak{M}$ is clear from the context, instead of $(\mathfrak{M}, w) \models \varphi$ we often write $w \models \varphi$.) A formula φ is said to be *valid* in a frame $\mathfrak{F}$ if $(\mathfrak{M}, w) \models \varphi$ holds for every model $\mathfrak{M}$ based on $\mathfrak{F}$ and every point w in it.

Consider now a Kripke complete normal modal logic L (i.e., a subset L of $\mathcal{ML}$ for which there exists a class $\mathcal{F}$ of frames such that L is the set of all formulas that are valid in every $\mathfrak{F} \in \mathcal{F}$). The *local consequence relation* '$\varphi_1 \models_L \varphi_2$' for L is defined as follows: $\varphi_1 \models_L \varphi_2$ holds if, and only if, for every pointed model $(\mathfrak{M}, w)$ based on a frame for L, we have $(\mathfrak{M}, w) \models \varphi_2$ whenever $(\mathfrak{M}, w) \models \varphi_1$.

Given a Kripke complete normal modal logic L and two $\mathcal{ML}$-formulas φ_1 and φ_2, we say that $\varphi_1 \wedge \varphi_2$ is a *conservative extension of* φ_1 in L if, for every $\psi \in \mathcal{ML}$ with $var(\psi) \subseteq var(\varphi_1)$, $\varphi_1 \wedge \varphi_2 \models_L \psi$ implies $\varphi_1 \models_L \psi$.

If $\varphi_1 \wedge \varphi_2$ is *not* a conservative extension of φ_1 in L, then there is a formula ψ with $var(\psi) \subseteq var(\varphi_1)$ such that $\varphi_1 \wedge \psi$ is satisfiable in a model based on a frame for L, while $\varphi_1 \wedge \varphi_2 \wedge \psi$ is not satisfiable in any such model. In this case we call ψ a (non-conservativeness) *witness formula* (or simply a *witness*) for the pair (φ_1, φ_2) in L.

The notion of conservative extension turns out to be closely connected with the notion of uniform interpolation. We remind the reader that a modal logic L is said to have *uniform interpolation* if, for every formula φ and every finite set $\mathbf{p}$ of variables, there exists a formula $\exists_L \mathbf{p}.\varphi$ such that

- $var(\exists_L \mathbf{p}.\varphi) \subseteq var(\varphi) \setminus \mathbf{p}$,
- $\varphi \models_L \exists_L \mathbf{p}.\varphi$, and
- $\varphi \models_L \psi$ implies $\exists_L \mathbf{p}.\varphi \models_L \psi$, for every formula ψ with $var(\psi) \cap \mathbf{p} = \emptyset$.

LEMMA 1. *If L has uniform interpolation, then $\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in L iff $\varphi_1 \models_L \exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2)$, where $\mathbf{p} = var(\varphi_2) \setminus var(\varphi_1)$.*

Proof. Suppose that $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1. Take a formula ψ with $var(\psi) \subseteq var(\varphi_1)$ such that $\varphi_1 \wedge \varphi_2 \models_L \psi$ and $\varphi_1 \not\models_L \psi$. Then we must have $\exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2) \models_L \psi$, from which $\varphi_1 \not\models_L \exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2)$.

Conversely, suppose $\varphi_1 \not\models_L \exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2)$. But then $\varphi_1 \wedge \varphi_2$ cannot be a conservative extension of φ_1 because $\varphi_1 \wedge \varphi_2 \models_L \exists \mathbf{p}.(\varphi_1 \wedge \varphi_2)$. ■

This lemma suggests the following procedure for deciding the conservativeness problem for a modal logic L with uniform interpolation: given φ_1 and φ_2, construct $\exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2)$ with $\mathbf{p} = var(\varphi_2) \setminus var(\varphi_1)$ and then check whether $\varphi_1 \models_L \exists_L \mathbf{p}.(\varphi_1 \wedge \varphi_2)$. Below, we will follow this approach for the modal logic **S5**. However, many standard modal logics, such as **S4**, **K4** or **S4.3**, do not have uniform interpolation [8]. In these cases we will provide direct proofs.

An important notion that can be used for analysing conservative extensions (as well as uniform interpolation) is the standard bisimulation between Kripke models [9]: for a finite set $\mathbf{p}$ of propositional variables and two models $(\mathfrak{M}_1, w_1)$ and $(\mathfrak{M}_2, w_2)$, a $\mathbf{p}$-*bisimulation* between $(\mathfrak{M}_1, w_1)$ and $(\mathfrak{M}_2, w_2)$ is a relation between the two models satisfying the standard bisimulation conditions for the variables in $\mathbf{p}$. If $(\mathfrak{M}_1, w_1)$ and $(\mathfrak{M}_2, w_2)$ are $\mathbf{p}$-bisimilar, then we write $(\mathfrak{M}_1, w_1) \leftrightarrows_{\mathbf{p}} (\mathfrak{M}_2, w_2)$. The first important property of bisimilar models is that if $(\mathfrak{M}_1, w_1) \leftrightarrows_{\mathbf{p}} (\mathfrak{M}_2, w_2)$ then $(\mathfrak{M}_1, w_1) \models \varphi$ iff $(\mathfrak{M}_2, w_2) \models \varphi$, for all formulas φ with $var(\varphi) \subseteq \mathbf{p}$. To discuss the second important property, we remind the reader that a frame $\mathfrak{F} = (W, R)$ (and a model based on $\mathfrak{F}$) is said to be m-*transitive*, for $m \geq 1$, if whenever $uRx_1R \dots Rx_mRv$ then there exist $k < m$ and points $y_1, \dots, y_k \in W$ such that $uRy_1R \dots Ry_kRv$ (in this sense, standard transitivity is 1-transitivity). A Kripke complete normal modal logic L is called m-*transitive* if the frames validating it are m-transitive.

We will be using the following property of m-transitive models: for every finite set $\mathbf{p}$ of variables and every finite pointed m-transitive model $(\mathfrak{M}, w)$, one can construct a formula $\chi_{\mathbf{p}}(\mathfrak{M}, w)$ containing only variables from $\mathbf{p}$ such that, for every pointed m-transitive model $(\mathfrak{M}', w')$,

$$(\mathfrak{M}', w') \models \chi_{\mathbf{p}}(\mathfrak{M}, w) \quad \text{iff} \quad (\mathfrak{M}, w) \leftrightarrows_{\mathbf{p}} (\mathfrak{M}', w').$$

$\chi_{\mathbf{p}}(\mathfrak{M}, w)$ is called the *characteristic formula* for $(\mathfrak{M}, w)$ and $\mathbf{p}$. Notice that $\chi_{\mathbf{p}}(\mathfrak{M}, w)$ is uniquely determined modulo equivalence in the minimal m-transitive modal logic.

LEMMA 2. *For every m-transitive modal logic L with the finite model property, the following conditions are equivalent:*

- *$\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in L,*
- *for every finite model $(\mathfrak{M}, w)$ based on a frame for L, if $(\mathfrak{M}, w) \models \varphi_1$ then there exists a finite $var(\varphi_1)$-bisimilar model $(\mathfrak{M}', w')$ based on a frame for L and such that $(\mathfrak{M}', w') \models \varphi_2$.*

Moreover, if $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1 in L, then there exists a finite pointed model $(\mathfrak{M}, w)$ based on a frame for L such that $\chi_{var(\varphi_1)}(\mathfrak{M}, w)$ is a witness for (φ_1, φ_2) in L.

As a first application of Lemma 2, one can prove the following result for tabular modal logics. (Recall that a modal logic L is called *tabular* if L is the logic of a finite set of finite frames [2].)

THEOREM 3. *The non-conservativeness problem for each tabular logic L is* $\mathrm{NP}^{\mathrm{NP}}$*-complete.*

Intuitively, given formulas φ_1 and φ_2 the $\mathrm{NP}^{\mathrm{NP}}$-algorithm guesses a finite model based on a frame for L and satisfying φ_1, and then calls an NP-oracle to verify that no $var(\varphi_1)$-bisimilar model based on a frame for L satisfies φ_2. A full proof can be found in [7].

3 The NExpTime lower bound

Say that a Kripke complete modal logic L is of *unbounded alternatives* if, for every $N < \omega$, there exist a Kripke frame $\mathfrak{F} = (W, R)$ for L and a point $w \in W$ such that the number of R-successors of w is at least N, or $|\{v \in W \mid wRv\}| \geq N$, to be more precise. Many standard modal logics such as **K**, **K4**, **S4**, **GL**, **S4.3**, **S5** are clearly of unbounded alternatives—unlike the logics $\mathbf{Alt}_n$.

Given a model $\mathfrak{M}$ and a set $\mathbf{q}$ of variables, we call a model $\mathfrak{M}'$ a $\mathbf{q}$-variant of $\mathfrak{M}$ if $\mathfrak{M}'$ can be obtained from $\mathfrak{M}$ by changing the valuation of some variables in $\mathbf{q}$ (but nothing else).

THEOREM 4. *Let L be a Kripke complete normal modal logic of unbounded alternatives. Then the conservativeness problem for L is co*NExpTime*-hard.*

Proof. The proof is by reduction of the complement of the well-known NExpTime-complete $2^n \times 2^n$-*bounded tiling problem* (see, e.g., [16]): given $n < \omega$, a finite set $\mathcal{T}$ of tile types and a $t_0 \in \mathcal{T}$, decide whether $\mathcal{T}$ can tile the $2^n \times 2^n$ grid in such a way that t_0 is placed onto $(0, 0)$. More precisely, let H and V be the binary relations on $\mathcal{T} \times \mathcal{T}$ such that $(t, t') \in H$ ($(t, t') \in V$) iff the colours of the right (upper) edge of t and the left (bottom) edge of t' coincide. Then $\mathcal{T}$ is said to *tile* the $2^n \times 2^n$ grid if there is a function $\tau : 2^n \times 2^n \to \mathcal{T}$ such that

- if $\tau(i, j) = t$ and $\tau(i + 1, j) = t'$ then $(t, t') \in H$, for all $i < 2^n - 1$, $j < 2^n$,
- if $\tau(i, j) = t$ and $\tau(i, j + 1) = t'$ then $(t, t') \in V$, for all $i < 2^n$, $j < 2^n - 1$,
- $\tau(0, 0) = t_0$.

Given a set $\mathcal{T} = \{t_1, \dots, t_m\}$ of tile types, we are going to construct two formulas φ_1 and φ_2 such that (i) $|\varphi_1|$ and $|\varphi_2|$ are polynomial in m and n, and (ii) $\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in L iff $\mathcal{T}$ *cannot* tile the $2^n \times 2^n$ grid in such a way that t_0 is placed onto $(0, 0)$.

To construct φ_1 and φ_2, we will use the following propositional variables:

- $\mathbf{p} = \{p_1, \dots, p_n\}$ and $\mathbf{q} = \{q_1, \dots, q_n\}$ to represent the points (i, j) of the $2^n \times 2^n$ grid in models by means of the standard binary encoding; for example, $(1, 2)$ is represented by a point w of some model iff

$$w \models p_1 \wedge \neg p_2 \wedge \cdots \wedge \neg p_n \quad \text{and} \quad w \models \neg q_1 \wedge q_2 \wedge \neg q_3 \wedge \cdots \wedge \neg q_n$$

(we will call a **p**-*literal* any conjunction $\neg_1 p_1 \wedge \cdots \wedge \neg_n p_n$ where each $\neg_i$ is either $\neg$ or blank),

- $\mathbf{t} = \{t_1, \ldots, t_m\}$: $w \models t_i$ will mean that the grid point represented by w is covered by a tile of type t_i,
- the set $\mathbf{A}$ of auxiliary variables $P_1, \ldots, P_n$, $Q_1, \ldots, Q_n$, $T_1, \ldots, T_m$; these variables will occur in φ_2, but not in φ_1.

The formula

$$\varphi_1 \;=\; (\neg p_1 \wedge \cdots \wedge \neg p_n) \wedge (\neg q_1 \wedge \cdots \wedge \neg q_n) \wedge t_0 \wedge \Box \bigvee_{i=1}^{m} t_i.$$

is supposed to say that if $(\mathfrak{M}, w) \models \varphi_1$ then w represents $(0, 0)$, which is covered by t_0, and each point of the grid represented by some R-successor of w is covered by at least one tile of a type from $\mathcal{T}$.

We say that a pointed model $(\mathfrak{M}, w)$ *represents a proper tiling of the* $2^n \times 2^n$ *grid by* $\mathcal{T}$ if $(\mathfrak{M}, w) \models \rho_{\mathcal{T},n}$, where $\rho_{\mathcal{T},n}$ is the conjunction of the following formulas:

$$\Diamond^+(l_1 \wedge l_2),$$

$$\Diamond^+(l_1 \wedge l_2 \wedge t) \to \Box^+\big(l_1 \wedge l_2 \to (t \wedge \neg \bigvee_{t' \neq t} t')\big),$$

$$\Diamond^+(l_1 \wedge l_2 \wedge t) \to \Diamond^+ \bigvee_{(t,t') \in H} \big((l_1 + 1) \wedge l_2 \wedge t'\big), \text{ for } l_1 < 2^n - 1,$$

$$\Diamond^+(l_1 \wedge l_2 \wedge t) \to \Diamond^+ \bigvee_{(t,t') \in V} \big(l_1 \wedge (l_2 + 1) \wedge t'\big), \text{ for } l_2 < 2^n - 1,$$

for all possible **p**-literals l_1, **q**-literals l_2, and $t, t' \in \mathbf{t}$. Here we use the abbreviations $\Box^+\psi = \psi \wedge \Box\psi$, $\Diamond^+\psi = \psi \vee \Diamond\psi$, and if l_i represents a number $k < 2^n - 1$ then $l_i + 1$ represents $k + 1$.

The formula φ_2 to be constructed below will have the property that, for every model $\mathfrak{M}$ with $(\mathfrak{M}, w) \models \varphi_1$, the following conditions are equivalent:

1. there is an $\mathbf{A}$-variant $\mathfrak{M}'$ of $\mathfrak{M}$ such that $(\mathfrak{M}', w) \models \varphi_2$,
2. $(\mathfrak{M}, w)$ *does not* represent a proper tiling of the $2^n \times 2^n$ grid by $\mathcal{T}$.

Suppose for the moment that we have managed to construct such a formula φ_2 of length polynomial in m and n. We claim then that $\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in L iff $\mathcal{T}$ cannot tile the $2^n \times 2^n$ grid in such a way that t_0 is placed onto $(0, 0)$. Indeed, assume first that $\mathcal{T}$ cannot tile the grid in this way, and consider any model $\mathfrak{M}$ based on a frame for L and satisfying φ_1 at its root w. Then $(\mathfrak{M}, w)$ cannot represent a proper tiling of the grid by means of $\mathcal{T}$, and so we can find an $\mathbf{A}$-variant $\mathfrak{M}'$ of $\mathfrak{M}$ such that $(\mathfrak{M}', w) \models \varphi_2$. Clearly, this means that $\varphi_1 \wedge \varphi_2$ is a conservative extension of φ_1 in L.

Now suppose that $\mathcal{T}$ can tile the grid. Clearly, we can satisfy $\varphi_1 \wedge \rho_{\mathcal{T},n}$ in a model based on a frame for L. But then $\rho_{\mathcal{T},n}$ is a witness for (φ_1, φ_2) in L because a model $(\mathfrak{M}, w)$ with $(\mathfrak{M}, w) \models \varphi_1 \wedge \varphi_2 \wedge \rho_{\mathcal{T},n}$ would trivially satisfy the former condition above but not the latter.

Now we construct the required formula φ_2. How to ensure that a point and its successors do not represent a tiling properly? There can be three different types of *defects*:

1. two different tiles cover the same point or two different tiles cover two points representing the same pair (i, j) on the grid,
2. there is a point representing (i, j) but there is no point representing $(i+1, j)$, for $i < 2^n-1$, or no point representing $(i, j+1)$, for $j < 2^n-1$,
3. colour mismatch.

To encode the *existence* of at least one of these defects we require our auxiliary variables P_i, Q_i, and T_i which will be used to carry, everywhere in the relevant part of the model, the information that there exists a point representing some grid point v covered by some tile t. The formula

$$\bigwedge_{i=1}^{n}(\Diamond^+ P_i \leftrightarrow \Box^+ P_i) \wedge \bigwedge_{i=1}^{n}(\Diamond^+ Q_i \leftrightarrow \Box^+ Q_i) \wedge \bigwedge_{i=1}^{m}(\Diamond^+ T_i \leftrightarrow \Box^+ T_i) \qquad (1)$$

says that each of the P_i, Q_i, and T_i has the same truth-value everywhere in the relevant part of the model. Suppose that (1) is true at the root w of our hypothetical model. Then, by making the formula

$$\exists p, q, t \;=\; \Diamond^+\Big(\bigwedge_{i=1}^{n}\big((p_i \leftrightarrow P_i) \wedge (q_i \leftrightarrow Q_i)\big) \wedge \bigwedge_{i=1}^{m}(t_i \leftrightarrow T_i)\Big)$$

true at w, we send—via the P_i, Q_i and T_i—to all worlds accessible from w (and w itself) the information that there is a point v in the model representing some grid point and covered by some tiles. In particular, the formula

$$at P, Q \;=\; \bigwedge_{i=1}^{n}\big((p_i \leftrightarrow P_i) \wedge (q_i \leftrightarrow Q_i)\big)$$

is true at a point u accessible from the root or the root itself iff u and v represent the same grid point, and

$$at P, Q, T \;=\; \bigwedge_{i=1}^{n}\big((p_i \leftrightarrow P_i) \wedge (q_i \leftrightarrow Q_i)\big) \wedge \bigwedge_{i=1}^{m}(t_i \leftrightarrow T_i)$$

is true at u iff u and v represent the same grid point and are covered by the same tiles. Using these formulas we can now express that the model under consideration contains a defect of type 1:

$$\exists p, q, t \wedge \bigvee_{i \neq j}\big(\Diamond^+(at P, Q \wedge t_i) \wedge \Diamond^+(at P, Q \wedge t_j)\big). \qquad (2)$$

To describe defects of type 2, we require the formulas

$$\exists p^+, q, t \;=\; \Diamond^+\Big(\neg \bigwedge_{k=1}^{n} p_k \wedge \bigwedge_{k\le n} \Big(\big(\bigwedge_{i<k} p_i \wedge \neg p_k\big) \to \bigwedge_{i<k} \neg P_i \wedge P_k \wedge \bigwedge_{j=k+1}^{n} (p_j \leftrightarrow P_j)\Big) \wedge \bigwedge_{i=1}^{n} (q_i \leftrightarrow Q_i) \wedge \bigwedge_{i=1}^{m} (t_i \leftrightarrow T_i)\Big)$$

and

$$\exists p, q^+, t \;=\; \Diamond^+\Big(\neg \bigwedge_{k=1}^{n} q_k \wedge \bigwedge_{k\le n} \Big(\big(\bigwedge_{i<k} q_i \wedge \neg q_k\big) \to \bigwedge_{i<k} \neg Q_i \wedge Q_k \wedge \bigwedge_{j=k+1}^{n} (q_j \leftrightarrow Q_j)\Big) \wedge \bigwedge_{i=1}^{n} (p_i \leftrightarrow P_i) \wedge \bigwedge_{i=1}^{m} (t_i \leftrightarrow T_i)\Big).$$

The latter, for instance, says that, for some point in the relevant part of the model representing some grid point (k,l) and covered by some tiles t, the variables P_i represent k, the Q_i represent $l+1$, and the T_i represent the same tiles t. (For describing defects of type 2 we do not need the last conjuncts for t_i in these formulas. They are required for defects of type 3.)

The existence of a defect of type 2 can be guaranteed then by the formula

$$(\exists p^+, q, t \wedge \neg\Diamond^+ atP, Q) \vee (\exists p, q^+, t \wedge \neg\Diamond^+ atP, Q), \tag{3}$$

and the existence of a defect of type 3 can be ensured by the formula

$$\Big(\exists p^+, q, t \wedge \Diamond^+\big(atP, Q \wedge \neg \bigvee_{(t_i,t_j)\in H} (T_i \wedge t_j)\big)\Big) \vee \Big(\exists p, q^+, t \wedge \Diamond^+\big(atP, Q \wedge \neg \bigvee_{(t_i,t_j)\in V} (T_i \wedge t_j)\big)\Big). \tag{4}$$

Finally, we define φ_2 by taking

$$\varphi_2 \;=\; (1) \wedge \big((2) \vee (3) \vee (4)\big).$$

It is easy to see that φ_2 is as required. We leave details to the reader. ■

4 The upper bound for S5

It is well known that the modal logic **S5** has uniform interpolation. One can easily construct a uniform interpolant $\exists_{\mathbf{S5}}\mathbf{q}.\varphi$ for a formula φ of double exponential size using the local tabularity of **S5**, according to which there are only $2^{2^{2^{O(n)}}}$ pairwise non-equivalent formulas in **S5** built from n variables. It follows from Lemma 1 and the decidability of **S5** in coNP that the non-conservativeness problem for **S5** is decidable in non-deterministic 2ExpTime.

In this section, we improve this bound by showing that **S5** has uniform interpolants of exponential size (which can be constructed in exponential time). Thus, we obtain, by Lemma 1 and Theorem 4, that the conservativeness problem for **S5** is coNEXPTIME-complete.

Suppose $\varphi(\mathbf{p},\mathbf{q})$ with disjoint $\mathbf{p}$ and $\mathbf{q}$ is given. The $\mathbf{p}$-literal given by a model $\mathfrak{M}$ and a point x in it will be denoted by $l_{\mathfrak{M}}(x)$ or simply $l(x)$ if $\mathfrak{M}$ is understood. Thus,

$$l(x) \;=\; \bigwedge\{p_i \mid (\mathfrak{M},x)\models p_i\}\cup\{\neg p_i \mid (\mathfrak{M},x)\not\models p_i\}.$$

Denote by $M(\varphi)$ the number of occurrences of the modal operator $\Box$ in φ plus one.

A uniform interpolant $\exists_{\mathbf{S5}}\mathbf{q}.\varphi$ of size at most exponential in $|\varphi|$ can be constructed in the following way. First we take the set of all pairwise non-isomorphic rooted **S5**-models over $\mathbf{p}$ and $\mathbf{q}$[1] with at most $M(\varphi)$ worlds and such that no two distinct worlds in the model validate precisely the same variables from $\mathbf{p}$ and $\mathbf{q}$. The total number of such models is not exceeding $2^{2n\cdot M(\varphi)}$. Then we partition this set of models into (disjoint) subsets, say $\mathcal{K}_1,\dots,\mathcal{K}_m$, such that all models from the same $\mathcal{K}_i$ validate the same subformulas of φ starting with $\Box$ (recall that we use $\Diamond$ as an abbreviation).

For each $\mathfrak{M}\in\mathcal{K}_i$ based on a frame (W,R) and each point x in $\mathfrak{M}$ with $x\models\varphi$, let

$$\chi_{\mathfrak{M}}(x) \;=\; l(x)\wedge\bigwedge_{w\in W\setminus\{x\}}\Diamond l(w),$$

and let ψ_i be the disjunction of all such $\chi_{\mathfrak{M}}(x)$, for all $\mathfrak{M}\in\mathcal{K}_i$ and x in $\mathfrak{M}$ with $x\models\varphi$. Denote by T_i the set of all $\mathbf{p}$-literals that are not satisfied in any model from $\mathcal{K}_i$, and set

$$\chi_i \;=\; \psi_i\wedge\bigwedge_{\mathsf{line}\in T_i}\neg\Diamond l, \qquad \exists_{\mathbf{S5}}\mathbf{q}.\varphi \;=\; \bigvee_{i=1}^{m}\chi_i.$$

Clearly, the size of $\exists_{\mathbf{S5}}\mathbf{q}.\varphi$ is at most exponential in the size of φ, and it can be constructed in exponential time. So it remains to prove the following:

THEOREM 5. *$\exists_{\mathbf{S5}}\mathbf{q}.\varphi$ is a uniform interpolant for φ in* **S5**.

Proof. To show that $\varphi\models_{\mathbf{S5}}\exists_{\mathbf{S5}}\mathbf{q}.\varphi$, consider an arbitrary rooted **S5**-model $\mathfrak{M}$ based on a frame $(W,W\times W)$ and such that $x\models\varphi$ for some $x\in W$. Let $i\in\{1,\dots,m\}$ be such that $\mathfrak{M}$ validates precisely the same subformulas of φ starting with $\Box$ or $\Diamond$ as the models from $\mathcal{K}_i$. Now observe that, for every point $y\in W$, we can always find a subset $Y\subseteq W$ containing y such that the restriction of $\mathfrak{M}$ to Y is (isomorphic to) some model from $\mathcal{K}_i$ (just pick up one 'witness' satisfying $\neg\psi$ for every subformula $\Box\psi$ of φ such that $y\models\neg\Box\psi$). It follows that $(\mathfrak{M},x)\models\chi_i$, and so $(\mathfrak{M},x)\models\exists_{\mathbf{S5}}\mathbf{q}.\varphi$.

[1] This means that we restrict valuations in models to the variables in $\mathbf{p}$ and $\mathbf{q}$.

Suppose now that we have a formula ψ with $var(\psi) \cap \mathbf{q} = \emptyset$. We need to prove that if $\exists_{\mathbf{S5}}\mathbf{q}.\varphi \not\models_{\mathbf{S5}} \psi$ then $\varphi \not\models_{\mathbf{S5}} \psi$. Let $\exists_{\mathbf{S5}}\mathbf{q}.\varphi \not\models_{\mathbf{S5}} \psi$. Take a model $\mathfrak{M} = (\mathfrak{F}, \mathfrak{V})$ with $\mathfrak{F} = (W, W \times W)$ such that $(\mathfrak{M}, w) \models \exists_{\mathbf{S5}}\mathbf{q}.\varphi$ and $(\mathfrak{M}, w) \not\models \psi$ for some $w \in W$. By the definition of $\exists_{\mathbf{S5}}\mathbf{q}.\varphi$, we can find $i \in \{1, \dots, m\}$, $\mathfrak{N} \in \mathcal{K}_i$, and x in $\mathfrak{N}$ such that $(\mathfrak{M}, w) \models \chi_i$ and $(\mathfrak{M}, w) \models \chi_{\mathfrak{N}}(x)$. We know that $(\mathfrak{N}, x) \models \varphi$. Now, with the help of $\mathfrak{M}$, we extend $\mathfrak{N}$ to a model $\mathfrak{K}$ such that $(\mathfrak{K}, x) \models \varphi$ and $(\mathfrak{K}, x) \not\models \psi$.

Let $\mathfrak{N} = (\mathfrak{G}, \mathfrak{U})$ and $\mathfrak{G} = (U, U \times U)$. By the definition of $\chi_{\mathfrak{N}}(x)$, for every $u \in U$ there is $w_u \in W$ such that $l_{\mathfrak{N}}(u) = l_{\mathfrak{M}}(w_u)$. Clearly, we can take $w_x = w$. We can also assume that W is disjoint from U. Now define a new model $\mathfrak{K} = (\mathfrak{H}, \mathfrak{W})$ based on the frame $\mathfrak{H} = (V, V \times V)$, where

$$V = U \cup (W \setminus \{w_u \mid u \in U\})$$

and, for each $u \in U$,

$$\mathfrak{W}(p, u) = \begin{cases} \mathfrak{U}(p, u), & \text{for } p \in \mathbf{p} \cup \mathbf{q} \\ \mathfrak{V}(p, w_u), & \text{otherwise.} \end{cases}$$

For the remaining points of V the valuation $\mathfrak{W}$ is defined as follows. For each $v \in V \setminus U$ we can find, by the second conjunct of χ_i, a model $\mathfrak{N}_v \in \mathcal{K}_i$ (with a valuation $\mathfrak{U}_v$) and a point z_v in it such that $\mathsf{line}_{\mathfrak{N}_v}(z_v) = \mathsf{line}_{\mathfrak{M}}(v)$. Then we set, for all such v,

$$\mathfrak{W}(p, v) = \begin{cases} \mathfrak{U}_v(p, z_v), & \text{for } p \in \mathbf{p} \cup \mathbf{q} \\ \mathfrak{V}(p, v), & \text{otherwise.} \end{cases}$$

We have $(\mathfrak{K}, x) \models \varphi$ because, restricted to the variables from $\mathbf{p} \cup \mathbf{q}$, the model $\mathfrak{K}$ consists of $\mathfrak{N} \in \mathcal{K}_i$ together with some points from other models in $\mathcal{K}_i$ validating, by definition, precisely the same subformulas of φ starting with $\Box$. Finally, we have $(\mathfrak{K}, x) \not\models \psi$ because $var(\psi) \cap \mathbf{q} = \emptyset$, and, restricted to the variables in $var(\psi)$, the model $\mathfrak{K}$ is isomorphic to $\mathfrak{M}$ with x corresponding to w. ■

5 The upper bound for K

As in the case of **S5**, it is known [17, 5] that the modal logic **K** has uniform interpolation, with a uniform interpolant for a given formula and set of variables being constructed in an effective way. Moreover, it has been recently shown in [14] that one can construct a uniform interpolant $\exists_{\mathbf{K}}\mathbf{q}.\varphi$ for φ in exponential time in such a way that the size of $\exists_{\mathbf{K}}\mathbf{q}.\varphi$ is at most exponential in the size $|\varphi|$ of φ — $2^{p(|\varphi|)}$ for a certain polynomial p which does not depend on φ, to be more exact, — and its modal depth does not exceed the modal depth of φ.

Using this result, the fact that the decision problem for **K** is PSpace-complete, and the algorithm from Section 2, one can obtain an algorithm deciding the conservativeness problem for **K** using exponential *space* in the size of the input formulas. In this section we improve this bound by providing a coNExpTime algorithm. Thus, we obtain

THEOREM 6. *The conservativeness problem for* $\mathbf{K}$ *is co*NExpTime-*complete.*

The coNExpTime lower bound follows from Theorem 4. Here we present a nondeterministic exponential time algorithm for deciding the complement of the conservativeness problem for $\mathbf{K}$. Suppose that we are given formulas $\varphi_1(\mathbf{p})$ and $\varphi_2(\mathbf{p},\mathbf{q})$ with disjoint $\mathbf{p} = \{p_1,\dots,p_n\}$ and $\mathbf{q} = \{q_1,\dots,q_n\}$. Denote by $\mathsf{sub}(\varphi_i)$, $i = 1, 2$, the closure under single negation of the set of all subformulas of φ_i. As usual, by a φ_i-*type* t we mean a Boolean-closed subset of $\mathsf{sub}(\varphi_i)$, i.e.,

- $\psi \in t$ iff $\neg\psi \notin t$, for every $\neg\psi \in \mathsf{sub}(\varphi_i)$,
- $\psi \wedge \chi \in t$ iff $\psi \in t$ and $\chi \in t$, for every $\psi \wedge \chi \in \mathsf{sub}(\varphi_i)$.

Denote by tp_i the set of all types for φ_i; clearly, $|\mathsf{tp}_i| \leq 2^{|\varphi_i|}$. Let d be the maximum of the modal depths of φ_1 and φ_2. The algorithm is presented in Fig. 1.

LEMMA 7. *The algorithm in Fig. 1 accepts input* φ_1, φ_2 *iff* $\varphi_1 \wedge \varphi_2$ *is not a conservative extension of* φ_1.

Proof. ($\Leftarrow$) Suppose that $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1. By Lemma 1, we have $\varphi_1 \not\models_{\mathbf{K}} \exists_{\mathbf{K}}\mathbf{q}.(\varphi_1 \wedge \varphi_2)$. According to [14], there exists a uniform interpolant $\exists_{\mathbf{K}}\mathbf{q}.(\varphi_1 \wedge \varphi_2)$ whose size does not exceed $N' = 2^{p(|\varphi_1 \wedge \varphi_2|)}$ and whose modal depth is $\leq d$. So there is a model $\mathfrak{M}$ based on an irreflexive and intransitive tree $\mathfrak{F} = (W, R)$ of depth $\leq d$ and branching factor $\leq N'$ such that both φ_1 and $\neg\exists_{\mathbf{K}}\mathbf{q}.(\varphi_1 \wedge \varphi_2)$ are true at the root r of $\mathfrak{F}$. Without loss of generality we may assume that, for every $x \in W$, if $x \models \neg\Box\psi_1 \wedge \cdots \wedge \neg\Box\psi_k$ for pairwise distinct $\neg\Box\psi_i \in \mathsf{sub}(\varphi_2)$ then there are k distinct R-successors $y_1,\dots,y_k$ of x such that $y_i \models \neg\psi_i$; if this is not the case, we can duplicate some relevant subtrees, thereby increasing the branching factor to $\leq N$. Thus, we may assume that, restricted to the variables in $\mathbf{p}$, the algorithm above guesses the model $\mathfrak{M}$.

Clearly, in Step 2 of the algorithm, we are in the 'else-case.' Suppose now that there is a $t_0 \in \ell(r)$ with $\varphi_2 \in t_0$, that is, the algorithm rejects the input φ_1, φ_2. Define a function f that maps each $x \in W$ to an element of $\ell(x)$ by induction starting from the root:

- Set $f(r) = t_0$.
- If $f(x)$ is already defined, but $f(\cdot)$ is not defined for the R-successors of x, then do the following. Let the negated box formulas in $f(x)$ be $\neg\Box\psi_1,\dots,\neg\Box\psi_k$ and the box formulas $\Box\vartheta_1,\dots,\Box\vartheta_h$. As $f(x) \in \ell(x)$, by the definition of ℓ there are distinct R-successors $y_1,\dots,y_k$ of x and types $t_1 \in \ell(y_1),\dots,t_k \in \ell(y_k)$ such that $\{\neg\psi_i,\vartheta_1,\dots,\vartheta_h\} \subseteq t_i$ for $1 \leq i \leq k$. Set $f(y_i) = t_i$ for $1 \leq i \leq k$. For each R-successor y of

Input: formulas $\varphi_1(\mathbf{p})$ and $\varphi_2(\mathbf{p},\mathbf{q})$.

1. Guess a Kripke model $\mathfrak{M}$ over the variables in $\mathbf{p}$ such that
   - $\mathfrak{M}$ is based on an irreflexive intransitive tree $\mathfrak{F} = (W, R)$,
   - the depth of $\mathfrak{F}$ is $\leq d$,
   - the branching factor of $\mathfrak{F}$ is bounded by $N = |\varphi_2| \cdot 2^{p(|\varphi_1 \wedge \varphi_2|)}$.
2. Check whether $(\mathfrak{M}, r) \models \varphi_1$, where r is the root of $\mathfrak{M}$. If this is not the case, reject the input. Else
3. label the points x of $\mathfrak{F}$ with subsets $\ell(x)$ of tp_2 by induction starting from the leafs as follows:
   - ff $x \in W$ is a leaf of $\mathfrak{F}$, then $\ell(x)$ consists of all types $t \in \mathsf{tp}_2$ such that $p_i \in t$ iff $x \models p_i$, for all $p_i \in \mathbf{p}$, and t contains all formulas of the form $\Box\psi \in \mathsf{sub}(\varphi_2)$.
   - Suppose now that $x \in W$ is not a leaf and all R-successors of x have already been labelled. Consider a type $t \in \mathsf{tp}_2$ and let $\Box\vartheta_1, \dots, \Box\vartheta_m$ be all the box formulas in t and $\neg\Box\psi_1, \dots, \neg\Box\psi_k$ be all the negated box formulas in t. Then $t \in \ell(x)$ iff the following conditions hold:
     (a) $p_i \in t$ iff $x \models p_i$, for all $p_i \in \mathbf{p}$,
     (b) for each R-successor y of x, there exists a $t' \in \ell(y)$ with $\{\vartheta_1, \dots, \vartheta_m\} \subseteq t'$,
     (c) there exist pairwise distinct R-successors $y_1, \dots, y_k$ of x and types $t_1 \in \ell(y_1), \dots, t_k \in \ell(y_k)$ such that $\{\neg\psi_i, \vartheta_1, \dots, \vartheta_m\} \subseteq t_i$, $1 \leq i \leq k$.
4. Check whether there is a $t \in \ell(r)$ such that $\varphi_2 \in t$. If this is the case, then reject the input; otherwise accept it.

Figure 1. Deciding non-conservativeness for **K**.

x with $y \notin \{y_1, \dots, y_k\}$, there is a $t \in \ell(y)$ such that $\{\vartheta_1, \dots, \vartheta_h\} \subseteq t$. Set $f(y) = t$.

Now define a $\mathbf{q}$-variant $\mathfrak{M}'$ of $\mathfrak{M}$ by taking, for all $x \in W$ and all $p \in \mathbf{q}$,

$$x \models p \quad \text{iff} \quad p \in f(x).$$

We still have $(\mathfrak{M}', r) \models \varphi_1$. Moreover, it can be easily shown by induction that $(\mathfrak{M}', x) \models \bigwedge_{\psi \in f(x)} \psi$ for all $x \in W$. Since $\varphi_2 \in f(r)$, we must have $(\mathfrak{M}', r) \models \varphi_2$. But then $(\mathfrak{M}, r) \models \exists_{\mathbf{K}}\mathbf{q}.(\varphi_1 \wedge \varphi_2)$, which is a contradiction.

($\Rightarrow$) Suppose now that our algorithm accepts φ_1, φ_2. Let $\mathfrak{M}$ be the model guessed by the algorithm and based on a tree $\mathfrak{F} = (W, R)$ with root r. Without loss of generality we may assume that, whenever xRy in $\mathfrak{F}$ then there are $|\varphi_2|$-many distinct R-successors z of x with $t^1_{\mathfrak{M}}(z) = t^1_{\mathfrak{M}}(y)$, where $t^1_{\mathfrak{M}}(x)$ the φ_1-*type* of x in $\mathfrak{M}$, that is,

$$t^1_{\mathfrak{M}}(x) \;=\; \{\psi \in \mathsf{sub}(\varphi_1) \mid (\mathfrak{M}, x) \models \psi\}.$$

With each $x \in W$ we associate inductively a formula $\psi(x)$ over $\mathbf{p}$ starting from the leaves of $\mathfrak{F}$:

- if $x \in W$ is a leaf, then

$$\psi(x) \;=\; \Box\bot \wedge \bigwedge_{x \models p_i} p_i \wedge \bigwedge_{x \not\models p_i} \neg p_i,$$

- if $x \in W$ is a non-leaf and $y_1, \dots, y_k$ are its R-successors, then

$$\psi(x) \;=\; \bigwedge_{x \models p_i} p_i \wedge \bigwedge_{x \not\models p_i} \neg p_i \wedge \bigwedge_{1 \le i \le k} \Diamond\psi(y_i) \wedge \Box \bigvee_{1 \le i \le k} \psi(y_i).$$

Now one can show that $\varphi_1 \wedge \varphi_2 \models_{\mathbf{K}} \neg\psi$, but $\varphi_1 \not\models_{\mathbf{K}} \neg\psi$, which means that $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1. ■

To complete the proof of Theorem 6, we note that the algorithm above runs in exponential time: the guessed model $\mathfrak{M}$ is at most exponentially large, and checking whether φ_1 is true in r can be done in exponential time. It remains to consider the $\ell(\cdot)$ labelling procedure. We have to label $|W|$ points—i.e., at most exponentially many. For each point x, we have to check for $|\mathsf{tp}_2|$ (exponentially) many types whether or not they should be included in $\ell(x)$. Condition (a) can be checked in polynomial time. Condition (b) can be checked in exponential time, since there are at most exponentially many successors in $\mathfrak{M}$. For condition (c) we have to consider all k-tuples of pairs (y, t) with y a successor of x and $t \in \ell(y)$. It is clear that there are at most exponentially many such tuples.

6 The upper bound for S4

In this section we present an algorithm deciding the conservativeness problem for **S4** in ExpSpace in the size of the input formulas.

Before proceeding to the technical details, we remind the reader that Kripke models for **S4** are based on quasi-orders $\mathfrak{F} = (W, R)$, that is, R is transitive and reflexive. A set $C \subseteq W$ is called a *cluster* in $\mathfrak{F}$ if $C = \{y \in W \mid xRy \;\&\; yRx\}$ for some $x \in W$; in this case we also say that C is the cluster *generated by* x and denote it by $C(x)$. Recall that every rooted model for **S4** is a p-morphic image of a model based on a *tree of clusters*, that is, a rooted quasi-order (W, R) such that, for all $x, y, z \in W$, if xRz and yRz, then either xRy or yRx or $C(x) = C(y)$. Without loss of generality we assume in this section that all our models are based on *finite trees of clusters*. Given a quasi-order $\mathfrak{F} = (W, R)$, we say that a cluster $C(x)$, for $x \in W$, is an *immediate strict predecessor* of a cluster $C(y)$ if xRy, $C(x) \neq C(y)$ and whenever $xRzRy$ then either $C(z) = C(x)$ or $C(z) = C(y)$. By the *depth* of $\mathfrak{F}$ we understand the length n of the longest sequence $C(x_1), \dots, C(x_n)$ of clusters in $\mathfrak{F}$ such that $C(x_i)$ is an immediate strict predecessor of $C(x_{i+1})$. A point y is a *strict successor* of a point x iff xRy and $C(x) \neq C(y)$. The *branching factor* of $\mathfrak{F}$ is the maximal number of immediate strict successor clusters of a cluster in $\mathfrak{F}$.

Suppose that we are given two formulas φ_1 and φ_2. The central role in our algorithm will be played by the following notion of a realisable triple for φ_1, φ_2. Consider a triple $\mathsf{t} = (t, \Gamma, \Delta)$ where t is a φ_1-type and Γ, Δ are sets of $\varphi_1 \wedge \varphi_2$-types. We call t *realisable* if there exists a pointed model $(\mathfrak{M}, x)$ based on a tree of clusters with root x such that

- $t = t^1_{\mathfrak{M}}(x)$ (where, as before, $t^1_{\mathfrak{M}}(x) = \{\psi \in \mathsf{sub}(\varphi_1) \mid (\mathfrak{M}, x) \models \psi\}$),
- Γ is the set of all $\varphi_1 \wedge \varphi_2$-types s such that $\bigwedge_{\sigma \in s} \sigma \wedge \chi_{var(\varphi_1)}(\mathfrak{M}, x)$ is satisfiable;
- Δ is the set of all $\varphi_1 \wedge \varphi_2$-types s for which there exists a point y in $\mathfrak{M}$ such that $\bigwedge_{\sigma \in s} \sigma \wedge \chi_{var(\varphi_1)}(\mathfrak{M}, y)$ is satisfiable.

In this case we say that $\mathsf{t} = (t, \Gamma, \Delta)$ is *realised* by $(\mathfrak{M}, x)$. Observe that if (t, Γ, Δ) is realisable, then $\Gamma \subseteq \Delta$ and, as follows from the main property of characteristic formulas and bisimulations, $t \subseteq s$ for every $s \in \Gamma$. The meaning of realisable triples will become clear from the following easily proved lemma.

LEMMA 8. *The following two conditions are equivalent for any formulas φ_1 and φ_2:*

(1) *$\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1,*

(2) *there exists a realisable triple $\mathsf{t} = (t, \Gamma, \Delta)$ (for φ_1, φ_1) such that $\varphi_1 \in t$ but $\varphi_1 \wedge \varphi_2 \notin s$ for any $s \in \Gamma$.*

Moreover, for every finite pointed model $(\mathfrak{M}, x)$ based on a tree of clusters, $\chi_{var(\varphi_1)}(\mathfrak{M}, x)$ is a witness for (φ_1, φ_2) iff the triple t realised by $(\mathfrak{M}, x)$ satisfies condition (2).

Observe that Δ does not play any role in the lemma above. Its role in realisable triples is indicated in the proof of Lemma 9 below.

We use the notion of a realisable triple to show that whenever $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1, then there exists a witness $\chi_{var(\varphi_1)}(\mathfrak{M}, x)$ of a certain bounded size. First, associate with every realisable triple $\mathsf{t} = (t, \Gamma, \Delta)$ the set

$$\Phi_{\mathsf{t}} \;=\; \{\{\Box\psi_1, \dots, \Box\psi_k\} \subseteq \mathsf{sub}(\varphi_1 \wedge \varphi_2) \mid \forall s \in \Gamma\ \{\Box\psi_1, \dots, \Box\psi_k\} \not\subseteq s\}.$$

LEMMA 9. *For every realisable triple $\mathsf{t} = (t, \Gamma, \Delta)$ for φ_1, φ_2, there is a realisable triple $\mathsf{t}' = (t, \Gamma', \Delta')$ such that $\Gamma \supseteq \Gamma'$, $\Delta \supseteq \Delta'$, and t' can be realised in a model $\mathfrak{M}'$ based on a tree of clusters $\mathfrak{F}'$ such that*

- *each cluster in $\mathfrak{F}'$ contains at most $2^{|\varphi_1|}$ points,*
- *the branching factor of $\mathfrak{F}'$ is bounded by $2^{|\varphi_1 \wedge \varphi_2|}$,*

- *the depth of* $\mathfrak{F}'$ *is bounded by* $1 + 2^{|\varphi_1 \wedge \varphi_2|}$.

Moreover, for any two points x, y *such that* y *is a strict successor of* x *in* $\mathfrak{F}'$ *and* x *is not in the root cluster of* $\mathfrak{F}'$, $\Phi_{\mathsf{t}_y} \subsetneq \Phi_{\mathsf{t}_x}$, *for the triples* t_x *and* t_y *realised by* $(\mathfrak{M}', x)$ *and* $(\mathfrak{M}', y)$, *respectively.*

Proof. (This is a sketch, a full proof can be found in [7].) Suppose that a triple $\mathsf{t} = (t, \Gamma, \Delta)$ for φ_1, φ_2 is realised in a model $(\mathfrak{M}, x)$ based on a finite tree of clusters $\mathfrak{F} = (W, R)$ with root x. The upper bound on the cardinality of clusters follows from the simple fact that if two points y and y' validate the same variables from φ_1 then we can omit one of these points and the resulting models will be $var(\varphi_1)$-bisimilar to the original one.

Consider now some $y \in W$ and denote by $\mathsf{t}_y = (t_y, \Gamma_y, \Delta_y)$ the triple for φ_1, φ_2 realised by $(\mathfrak{M}, y)$. For every $\varphi_1 \wedge \varphi_2$-type $s \notin \Gamma_y$ we take the set $\{\Box\psi_1, \dots, \Box\psi_k\}$ of all box formulas in s and choose a maximal (with respect R) strict successor z of y such that there does not exist a type $s' \in \Gamma_z$ with $\{\Box\psi_1, \dots, \Box\psi_k\} \subseteq s'$, if such a strict R-successor exists. Observe that (since $t_y \subseteq s$ for every $s \in \Gamma_y$) for each $\neg\Box\psi \in t_y$ we have also chosen a maximal strict R-successor z of y with $(\mathfrak{M}, z) \models \neg\Box\psi$, if such a successor exists.

Now remove from the submodel $\mathfrak{M}_y$ of $\mathfrak{M}$ generated by y all clusters $C(u)$ such that $C(u) \neq C(y)$, yRu and for no chosen point z do we have zRu. Denote the resulting model by $\mathfrak{N}$. Denote by $\mathsf{t}' = (t', \Gamma', \Delta')$ the triple realised by $(\mathfrak{N}, y)$. Our aim is to show

- $t_y = t'$;
- $\Gamma' \subseteq \Gamma_y$;
- $\Delta' \subseteq \Delta_y$.

Having shown this, we can replace $\mathfrak{M}_y$ in $\mathfrak{M}$ with $\mathfrak{N}$ where the number of immediate strict successors of $C(y)$ does not exceed $2^{|\varphi_1 \wedge \varphi_2|}$. It can be proved that $t = t''$, $\Gamma'' \subseteq \Gamma$, and $\Delta'' \subseteq \Delta$ for the triple $(t'', \Gamma'', \Delta'')$ realised in the resulting model $(\mathfrak{M}', x)$. For a proof we refer the reader to [7]. Note that the proof requires the third component of realizable triples because the second component alone only encodes the behaviour of the root of a model but not of non-root points.

Now, since every $\neg\Box\psi \in t_y$ has a witness in $\mathfrak{N}$, we clearly have $t_y = t'$. To prove that $\Gamma' \subseteq \Gamma_y$, suppose otherwise. Then we have some $s \in \Gamma'$ such that $s \notin \Gamma_y$. Two cases are possible now. *Case 1*: there is a strict successor z of y in $\mathfrak{M}_y$ such that there does not exist a type $s' \in \Gamma_z$ with $\{\Box\psi_1, \dots, \Box\psi_k\} \subseteq s'$ (where $\{\Box\psi_1, \dots, \Box\psi_k\}$ are all box formulas in s). Then a (possibly different) point with this property was chosen for $\mathfrak{N}$. From this one can easily derive a contradiction with s being in Γ'. *Case 2*: there is no such z in $\mathfrak{M}_y$. Let $C(y_1), \dots, C(y_m)$ be all immediate strict successors of $C(y)$ which do not belong to $\mathfrak{N}$. Denote by $\mathfrak{M}_i$ the submodel of $\mathfrak{M}_y$ generated by y_i. For each y_i there exists a pointed model $(\mathfrak{M}'_i, y'_i)$ which is $var(\varphi_1)$-bisimilar to $(\mathfrak{M}_i, y_i)$ and $(\mathfrak{M}'_i, y'_i) \models \Box\psi_i$ for every $\Box\psi \in s$. Let $\mathfrak{K}$

Input: formulas φ_1, φ_2.

1. Choose, non-deterministically, a triple (t, Γ, Δ), where t is a φ_1-type containing φ_1 and Γ, Δ are sets of $\varphi_1 \wedge \varphi_2$-types such that $\varphi_1 \wedge \varphi_2 \notin s$ for any $s \in \Gamma$.
2. Accept input φ_1, φ_2 iff $\mathsf{realise}_0(t, \Gamma, \Delta)$ returns 'true.'

Figure 2. Deciding non-conservativeness for **S4**.

be a model based on a tree of clusters with root v which is $var(\varphi_1)$-bisimilar to $(\mathfrak{N}, y)$ and satisfies s at v. Add the models $\mathfrak{M}_i'$ to $\mathfrak{K}$ as immediate strict successors of the points which are $var(\varphi_1)$-bisimilar to y and denote the resulting model by $\mathfrak{K}'$. Clearly, $(\mathfrak{K}', v)$ is $var(\varphi_1)$-bisimilar to $(\mathfrak{M}_y, y)$. On the other hand, s is still satisfied by $(\mathfrak{K}', v)$, contrary to $s \notin \Gamma_y$.

Finally, the inclusion $\Delta' \subseteq \Delta_y$ follows from $\Gamma' \subseteq \Gamma_y$ and the fact that if a point from a non-root cluster is chosen for $\mathfrak{N}$ then all of its successors belong to $\mathfrak{N}$ as well.

By recursively performing this operation whenever possible for some y, we obtain a model with the properties required. ■

We are now in a position to present a NExpSpace algorithm deciding the non-conservativeness problem for **S4**. The ExpSpace upper bound is then obtained from the fact that NExpSpace = ExpSpace. To formulate the algorithm, we require the following definition. Let $\mathcal{R}$ be some set of realisable triples. We say that a triple t is *obtained in one step from* $\mathcal{R}$ if there exists a pointed model $(\mathfrak{M}, w)$ based on a tree of clusters with root w such that

(a) $(\mathfrak{M}, w)$ realises t,

(b) for every strict immediate successor $C(y)$ of $C(w)$ in $\mathfrak{M}$, there is some $x \in C(y)$ such that the triple realised by $(\mathfrak{M}, x)$ belongs to $\mathcal{R}$,

(c) every triple from $\mathcal{R}$ is realised by $(\mathfrak{M}, y)$, for some strict immediate successor $C(y)$ of $C(w)$.

The algorithm is shown in Fig. 2; it uses the procedure realise described in Fig. 3.

The following lemma shows that this 'algorithm' can be refined in such a way that it indeed runs in ExpSpace (a proof can be found in [7]).

LEMMA 10. (i) *It can be decided in* ExpSpace (*in the size of* φ_1, φ_2) *whether a triple* (t, Γ, Δ) *is realisable in a finite pointed model* $(\mathfrak{M}, w)$ *based on a cluster.* (ii) *It is decidable in* ExpSpace (*in the size of* φ_1, φ_2) *whether a triple is obtained in one step from a set* $\mathcal{R}$ *of triples of cardinality not exceeding* $2^{|\varphi_1 \wedge \varphi_2|}$.

We can now analyse the algorithm in Fig. 2. By Lemma 10 and condition (1) of the procedure realise, the procedures realise and $\mathsf{realise}_0$ always terminate and require exponential space only.

Input: a triple $\mathsf{t} = (t, \Gamma, \Delta)$, where t is a φ_1-type and Γ, Δ are sets of $\varphi_1 \wedge \varphi_2$-types.

$\mathsf{realise}(\mathsf{t})$ returns 'true' iff t is realisable in a pointed model based on a cluster or there exists a set $\mathcal{R}$ of triples with $|\mathcal{R}| \leq 2^{|\varphi_1 \wedge \varphi_2|}$ such that

(1) for all $\mathsf{t}' \in \mathcal{R}$, $\Phi_{\mathsf{t}'} \subsetneq \Phi_{\mathsf{t}}$,

(2) for all $\mathsf{t}' \in \mathcal{R}$, $\mathsf{realise}(\mathsf{t}')$ returns 'true,'

(3) t is obtained in one step from $\mathcal{R}$.

The procedure $\mathsf{realise}_0$ is defined in the same way as $\mathsf{realise}$ with the exception that condition (1) is omitted. (In particular, it still calls $\mathsf{realise}$ in (2)).

Figure 3. The procedures $\mathsf{realise}(t, \Gamma, \Delta)$ and $\mathsf{realise}_0(t, \Gamma, \Delta)$.

Now suppose that $\varphi_1 \wedge \varphi_2$ is not a conservative extension of φ_1. Then there is a realisable triple (t, Γ, Δ) such that $\varphi_1 \in t$ but $\varphi_1 \wedge \varphi_2 \notin s$, for any $s \in \Gamma$. Take a pointed model $(\mathfrak{M}, w)$ with the properties of Lemma 9 which realises a $\mathsf{t}' = (t, \Gamma', \Delta')$ with $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$. Now, let the algorithm in Fig. 2 guess the triple t'. Then it obviously accepts input φ_1, φ_2. Observe that we start with the procedure $\mathsf{realise}_0$ instead of $\mathsf{realise}$ because we have not proved that $\Phi_{\mathsf{t}'} \supsetneq \Phi_{\mathsf{t}''}$ for every triple t'' realised in a strict successor of w.

In conclusion, we obtain:

THEOREM 11. *The conservativeness problem for* **S4** *is decidable in* ExpSpace.

7 Conservative extensions in GL.3

Recall that the set of finite rooted frames for **GL.3** coincides with the set of finite strict linear orders (W, R). We are going to show the following

THEOREM 12. *The conservativeness problem for* **GL.3** *is* ExpSpace-*complete.*

In what follows we will use the observation that for models based on strict linear orders, every $\mathbf{p}$-bisimulation between $(\mathfrak{M}, w)$ and $(\mathfrak{M}', w')$ is an isomorphism between the submodels of these models generated by w and w', respectively, and restricted to the variables from $\mathbf{p}$. To formulate the decision procedure we require the notion of a realisable pair: a pair (t, Γ), where t is a φ_1-type t and Γ is a set of $\varphi_1 \wedge \varphi_2$-types, is said to be *realised in a pointed model* $(\mathfrak{M}, w)$ if

- $(\mathfrak{M}, w) \models t$ and
- Γ is the set of $\varphi_1 \wedge \varphi_2$-types s such that $s \wedge \chi_{var(\varphi_1)}(\mathfrak{M}, w)$ is satisfiable (in a finite model for **GL.3**).

LEMMA 13. $\varphi_1 \wedge \varphi_2$ *is not a conservative extension of* φ_1 *in* **GL.3** *iff there exists a pointed model* $(\mathfrak{M}, w)$ *based on a finite strict linear order such that*

Input: formulas φ_1, φ_2.

1. Choose, non-deterministically, a pair (t, Γ), where t is a φ_1-type containing φ_1 and Γ is a set of $\varphi_1 \wedge \varphi_2$-types such that $t \subseteq s$ for all $s \in \Gamma$ and $\varphi_1 \wedge \varphi_2 \notin s$ for any $s \in \Gamma$.
2. Call $\mathsf{real}(t, \Gamma, 2^{2^{|\varphi_1 \wedge \varphi_2|}})$.

Definition of procedure $\mathsf{real}(t, \Gamma, n)$ (where n is a natural number coded in binary):

1. If no s in Γ contains a formula of the form $\neg\Box\psi$, then accept input φ_1, φ_2 and stop. Else,
2. if $n = 0$, then reject input φ_1, φ_2 and stop. Else,
3. choose, non-deterministically, a pair (t', Γ'), where t' is a φ_1-type and Γ' is a set of $\varphi_1 \wedge \varphi_2$-types. Check whether (t', Γ') has the following properties:
   - $t' \subseteq s'$, for all $s' \in \Gamma$,
   - (t, t') is a suitable pair,
   - for every $s \in \Gamma$, there exists $s' \in \Gamma'$ such that (s, s') is a suitable pair,
   - for each $\varphi_1 \wedge \varphi_2$-type $s \notin \Gamma$, $t \not\subseteq s$ or there is no $s' \in \Gamma'$ such that (s, s') is a suitable pair.
4. If (t', Γ') does not have these properties, reject the input φ_1, φ_2 and stop. Else,
5. set $n := n - 1$.
6. Call $\mathsf{real}(t', \Gamma', n)$.

Figure 4. Deciding non-conservativeness for **GL.3**.

- *for the pair (t, Γ) realised by $(\mathfrak{M}, w)$, $\varphi_1 \in t$ and $\varphi_1 \wedge \varphi_2 \notin s$, for any $s \in \Gamma$,*
- *for any two points $x \neq y$ in $\mathfrak{M}$, the pair realised by $(\mathfrak{M}, x)$ is different from the pair realised by $(\mathfrak{M}, y)$.*

In particular, the length of the finite strict linear order underlying $\mathfrak{M}$ does not exceed $2^{2^{|\varphi_1 \wedge \varphi_2|}}$.

We say that a pair of types (t, t') is *suitable* if

- $\Box\psi \in t$ implies $\psi, \Box\psi \in t'$, and
- $\neg\Box\psi \in t$ implies $\neg\psi \in t'$ or $\neg\Box\psi \in t'$.

The non-deterministic algorithm deciding non-conservativeness in **GL.3** is shown in Fig. 4. Clearly, this algorithm requires exponential space only. Using the fact that NExpSpace = ExpSpace, we obtain an ExpSpace algorithm. The correctness of this algorithm follows from Lemma 13.

ExpSpace-hardness of the conservativeness problem for **GL.3** is proved by simulating computations of Turing machines $\mathcal{M} = (Q, \Sigma, \Gamma, q_0, \Delta)$ that solve an ExpSpace-hard problem and consume at most 2^n tape cells if started on an input of length n.

Let $\mathsf{w} = a_0 \cdots a_{n-1} \in \Sigma^*$ be an input to $\mathcal{M}$. In [7], we construct formulas φ_1 and φ_2 (depending on $\mathcal{M}$ and w) such that $\varphi_1 \wedge \varphi_2$ is *not* a conservative extension of φ_1 if, and only if, $\mathcal{M}$ does accept w. More precisely, we construct φ_1 and φ_2 in such a way that, if ψ is a witness for (φ_1, φ_2), then rooted models of $\varphi_1 \wedge \psi$ describe an accepting computation of $\mathcal{M}$ on w. In these models, each point represents a tape cell of a configuration of $\mathcal{M}$, and moving to the immediate successor of a point means moving to the next tape cell in the same configuration, or, if we are already at the end of the configuration, moving to the first tape cell of a successor configuration. Such models will have depth $2^n \cdot 2^{2^n}$ since the length of computations is bounded by 2^{2^n}.

8 Discussion

We have investigated the complexity of the conservativeness problem for the local consequence relation of a number of basic modal logics. One interesting conclusion is that the complexity of deciding conservativeness is not monotonically related to the complexity of the logic in question: for example, the satisfiability problem is NP-complete for **GL.3** and PSpace-complete for **K**, while the conservativeness problem is (probably) more complex for **GL.3** than for **K**. This resembles the situation with products of modal logics where **GL.3** $\times$ **GL.3** is Π^1_1-complete [13], while **K** $\times$ **K** is decidable [4].

In this paper, we have considered modal languages with one modal operator only. It is not difficult, however, to modify the proofs above to show that conservativeness (for the local consequence relation) is still NExpTime-complete for multimodal **S5** and multimodal **K**. Similarly, for multimodal **S4** and **K4** it is still decidable in ExpSpace.

For the global consequence relation the results are different: recall that φ *follows globally* from ψ in a modal logic L if φ is true everywhere in a model based on a frame for L whenever ψ is true everywhere in this model. Conservativeness with respect to the global consequence relation is now defined in the obvious way. Of course, for m-transitive modal logics the complexity upper bound for deciding conservativeness with respect to the local consequence is an upper bound for deciding conservativeness relative to the global consequence relation as well. This applies to **S5**, **S4** and **GL.3**. For **K**, however, deciding conservativeness with respect to the global consequence becomes 2ExpTime-complete, as follows from the investigation of conservative extensions in description logics in [6]. We expect deciding conservativeness with respect to the global consequence in multimodal **S5** and **S4** to be 2ExpTime-complete as well.

Finally, as indicated by Lemma 2, there is a close connection between studying modal logics extended by bisimulation quantifiers (cf. [3]) and conservative extensions. For many modal logics (it remains to explore which ones exactly) checking conservativeness is equivalent to deciding validity of the Σ_1-fragment of its extension by bisimulation quantifiers. This raises, for example, the question whether the analysis presented here can be extended to higher prenex fragments.

Acknowledgements. The work on this paper was partially supported by U.K. EPSRC grants no. GR/S63175/02, GR/S61973/02, GR/S63182/01, GR/S61966/01.

BIBLIOGRAPHY

[1] G. Antoniou and K. Kehagias. A note on the refinement of ontologies. *International Journal of Intelligent Systems*, 15:623–632, 2000.

[2] A. Chagrov and M. Zakharyaschev. *Modal Logic*. Oxford University Press, Oxford, 1997.

[3] T. French. *Bisimulation Quantifiers for Modal Logics*. PhD thesis, University of Western Australia, 2006.

[4] D. Gabbay and V. Shehtman. Products of modal logics. Part I. *Logic Journal of the IGPL*, 6:73–146, 1998.

[5] S. Ghilardi. An algebraic theory of normal forms. *Annals of Pure and Applied Logic*, 71(3):189–245, 1995.

[6] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proceedings of the Tenth International Conference of Principles of Knowledge Representation and Reasoning*, 187–197, 2006.

[7] S. Ghilardi, C. Lutz, F. Wolter, and M. Zakharyaschev. Conservative extensions in modal logic. http://www.csc.liv.ac.uk/~frank/publ/publ.html, 2006.

[8] S. Ghilardi and M. Zawadowski. Undefinability of propositional quantifiers in the modal system S4. *Studia Logica*, 55(2):259–271, 1995.

[9] V. Goranko and M. Otto. Modal model theory. In J. van Benthem, P. Blackburn, and F. Wolter, editors, *Handbook of Modal Logic*. Elsevier, 2006.

[10] M. Kracht. Modal consequence relations. In J. van Benthem, P. Blackburn, and F. Wolter, editors, *Handbook of Modal Logic*. Elsevier, 2006.

[11] C. Papadimitriou. *Computational Complexity*. Addison Wesley, 1995.

[12] A. Pitts. On an interpretation of second-order quantification in first-order intuitionistic propositional logic. *Journal of Symbolic Logic*, 57(1):33–52, 1992.

[13] M. Reynolds and M. Zakharyaschev. On the products of linear modal logics. *Journal of Logic and Computation*, 11:909–931, 2001.

[14] B. ten Cate, W. Conradie, M. Marx, and Y. Venema. Definitorially complete description logics. In *Proceedings of the Tenth International Conference of Principles of Knowledge Representation and Reasoning*, 79–89, 2006.

[15] W.M. Turski and T. Maibaum. *The Specification of Computer Programs*. Addison-Wesley, 1987.

[16] P. van Emde Boas. The convenience of tilings. In A. Sorbi, editor, *Complexity, Logic and Recursion Theory*, volume 187 of *Lecture Notes in Pure and Applied Mathematics*, pages 331–363. Marcel Dekker Inc., 1997.

[17] A. Visser. Uniform interpolation and layered bisimulation. In *Gödel'96 (Brno, 1996)*, volume 6 of *Lecture Notes Logic*, pages 139–164. Springer, Berlin, 1996.

[18] F. Wolter and M. Zakharyaschev. Modal decision problems. In J. van Benthem, P. Blackburn, and F. Wolter, editors, *Handbook of Modal Logic*. Elsevier, 2006.

Silvio Ghilardi
Department of Computer Science
University of Milan, Italy
ghilardi@dsi.unimi.it

Carsten Lutz
Institut für Theoretische Informatik
TU Dresden, Germany
lutz@tcs.inf.tu-dresden.de

Frank Wolter
Department of Computer Science
The University of Liverpool
Liverpool L69 3BX, U.K.
frank@csc.liv.ac.uk

Michael Zakharyaschev
School of Computer Science and Information Systems
Birkbeck College
Malet Street, London WC1E 7HX, U.K.
michael@dcs.bbk.ac.uk