The Interpolant Existence Problem for Weak K4 and Difference Logic

Agi Kurucz

Department of Informatics
King's College London, U.K.

Frank Wolter

Department of Computer Science
University of Liverpool, U.K.

Michael Zakharyaschev

School of Computing and Mathematical Sciences
Birkbeck, University of London, U.K.

Abstract

As well known, weak K4 and the difference logic DL do not enjoy the Craig interpolation property. Our concern here is the problem of deciding whether any given implication does have an interpolant in these logics. We show that the nonexistence of an interpolant can always be witnessed by a pair of bisimilar models of polynomial size for DL and of triple-exponential size for weak K4, and so the interpolant existence problems for these logics are decidable in coNP and coN3ExpTime, respectively. We also establish coNExpTime-hardness of this problem for weak K4, which is higher than the PSpace-completeness of its decision problem.

Keywords: Craig interpolant, propositional modal logic, computational complexity.

1 Introduction

Weak K4 is the modal logic one obtains when the $\Diamond$-operator of the propositional classical (uni)modal language is interpreted by the derivative operation [1] in topological spaces [7]—rather than the more conventional topological closure, which results in classical S4 [17]. In terms of Kripke semantics, weak K4 is characterised [7] by the class of weakly transitive frames, i.e., those $\mathfrak{F} = (W, R)$ that satisfy the condition

$$\forall x, y, z \in W \big(xRyRz \to (x = z) \vee xRz\big), \tag{1}$$

[1] The *derived set* of a subset X of a topological space comprises all limit points of X, i.e., those x all of whose neighbourhoods contain a point in X different from x.

which explains the moniker 'weak $\mathsf{K4}$' or $\mathsf{wK4}$ for this logic. Syntactically, $\mathsf{wK4}$ is obtained by adding the axiom $\Diamond\Diamond p \to (p \vee \Diamond p)$ to the basic normal modal logic K. A notable extension of $\mathsf{wK4}$ is the difference logic DL, which goes back to the 'logic of elsewhere' [23,19] and can be axiomatised by adding the Brouwersche axiom $p \to \Box\Diamond p$ to $\mathsf{wK4}$. While an arbitrary frame for DL is a symmetric and weakly transitive relation, DL is also known to be characterised by the class of *difference frames* (that is, Kripke frames of the form $(W, \neq)$) [19].

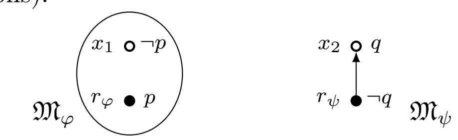
Despite their apparent similarity to $\mathsf{K4}$, $\mathsf{S4}$ and $\mathsf{S5}$, the logics $\mathsf{wK4}$ and DL have—or rather lack—one important feature: they do not enjoy the Craig interpolation property (CIP) [12], according to which each valid implication $\varphi \to \psi$ in a logic L has an interpolant in L, viz. a formula ι built from common variables of φ and ψ such that $(\varphi \to \iota) \in L$ and $(\iota \to \psi) \in L$; if φ and ψ have no variables in common, variable-free interpolant ι is built from the logical constants $\top$ and $\bot$.

Example 1.1 In the pictures below, $\bullet$ always denotes an irreflexive point, $\circ$ a reflexive one, and an ellipse represents a cluster (a set of points, in which any two distinct ones 'see' each other).

(i) Consider the following formulas without common variables:

$$\varphi = \Diamond\Diamond p \wedge \neg\Diamond p, \qquad \psi = \Diamond\Diamond\neg q \vee q \ (\equiv \Box\Box q \to q).$$

It is easy to see that $\varphi \to \psi$ is true in all models based on weakly transitive frames, and so $(\varphi \to \psi) \in \mathsf{wK4}$. On the other hand, the picture below shows models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ based on weakly transitive frames such that $\mathfrak{M}_\varphi, r_\varphi \models \varphi$, $\mathfrak{M}_\psi, r_\psi \models \neg\psi$, and the universal relation $\boldsymbol{\beta}$ between the points of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ is a ϱ-bisimulation for the shared signature $\varrho = \emptyset$ of φ and ψ, with $r_\varphi \boldsymbol{\beta} r_\psi$ (see Sec. 2 for definitions).

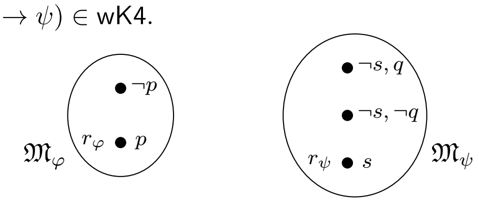


It follows that there is no variable-free formula ι with $(\varphi \to \iota) \in \mathsf{wK4}$ and $(\iota \to \psi) \in \mathsf{wK4}$ because ϱ-bisimulations preserve the truth-values of ϱ-formulas.

(ii) Consider the formulas

$$\varphi = \Diamond\Diamond p \wedge \neg\Diamond p \wedge \neg\Diamond\Diamond\Diamond p, \qquad \psi = \Diamond\Diamond s \wedge \neg\Diamond s \to \neg\big(\Diamond(q \wedge \Diamond s) \wedge \Diamond(\neg q \wedge \Diamond s)\big).$$

It is easy to see that $\varphi \to \psi$ is true in all models based on a weakly transitive frame, and so $(\varphi \to \psi) \in \mathsf{wK4}$.



On the other hand, the models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ above are based on difference frames with $\mathfrak{M}_\varphi, r_\varphi \models \varphi$, $\mathfrak{M}_\psi, r_\psi \models \neg\psi$, and the universal relation $\boldsymbol{\beta}$ between the points of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ is a ϱ-bisimulation with $r_\varphi \boldsymbol{\beta} r_\psi$, for the shared signature $\varrho = \emptyset$ of φ, ψ. Therefore, φ and ψ do not have an interpolant in any logic between wK4 and DL. $\dashv$

Our concern in this paper is the following

***interpolant existence problem (IEP)* for** $L \in \{\mathsf{wK4}, \mathsf{DL}\}$**:** given formulas φ and ψ, decide whether $\varphi \to \psi$ has an interpolant in L.

We show that the IEP for wK4 is decidable in coN3ExpTime, being coNExpTime-hard (harder than its decision problem), while the IEP for DL is coNP-complete (as is its decision problem).

In Section 2, we introduce the necessary technical tools and demonstrate them in the case of DL. Then we focus on the much more involved IEP for wK4, establishing the upper bound in Section 3 and the lower one in Section 4. Finally, we discuss related and open problems in Section 5.

2 Preliminaries

All (Kripke) frames $\mathfrak{F} = (W, R)$ we deal with in this paper are assumed to be *weakly transitive* (1). By a *cluster* in $\mathfrak{F}$ we mean any set of the form

$$C(x) = \{x\} \cup \{y \in W \mid xRy \wedge yRx\}, \quad x \in W.$$

A cluster $\mathfrak{F}$ can contain *reflexive* points y, for which yRy, as well as *irreflexive* points z, for which $\neg(zRz)$. A cluster with a single point, which is irreflexive, is said to be *degenerate*. Given $x, y \in W$, we write $xR^s y$ iff xRy and $C(x) \neq C(y)$.

Suppose C and C' are clusters in $\mathfrak{F}$ and $x \in W$. We write CRx if there exists $y \in C$ such that yRx, and CRC' if there are $x \in C$ and $y \in C'$ with xRy. Observe that CRC iff C is non-degenerate. We write $CR^s x$ if CRx and $x \notin C$, and $CR^s C'$ if CRC' and $C \neq C'$. Thus, R^s is a strict partial order on the set W_c of clusters in $\mathfrak{F}$. The reflexive closure of R^s is denoted by R^r. The frame $\mathfrak{F}$ is called *rooted* if there is $r \in W$, a *root* of $\mathfrak{F}$, with $W_c = \{C(x) \mid C(r) R^r C(x)\}$.

By a *signature* we mean any *finite* set of propositional variables, p_i. Given a signature σ, a σ*-formula* is built from variables in σ and logical constants $\bot$, $\top$ using the Boolean connectives $\wedge$, $\neg$ and the modal possibility operator $\Diamond$. The other Boolean connectives and the necessity operator $\Box$ are regarded as standard abbreviations. We denote by $sig(\varphi)$ the set of variables in a formula φ and by $sub(\varphi)$ the set of subformulas of φ together with their negations, setting $|\varphi| = |sub(\varphi)|$. We also use abbreviations $\Diamond^+\varphi = \varphi \vee \Diamond\varphi$, $\Box^+\varphi = \varphi \wedge \Box\varphi$ and $\Diamond\Gamma = \{\Diamond\varphi \mid \varphi \in \Gamma\}$, for a set Γ of formulas.

A σ*-model* based on a frame $\mathfrak{F} = (W, R)$ is a pair $\mathfrak{M} = (\mathfrak{F}, \mathfrak{v})$ with a *valuation* $\mathfrak{v}\colon \sigma \to 2^W$. The *truth-relation* $\mathfrak{M}, x \models \varphi$, for $x \in W$ and a σ-formula φ, is defined by induction as usual in Kripke semantics (e.g., $\mathfrak{M}, x \models \Diamond\varphi$ iff $\mathfrak{M}, y \models \varphi$, for some $y \in W$ with xRy). For any $\varrho \subseteq \sigma$, the *ϱ-type of $x \in W$ in $\mathfrak{M}$* is the set $t^\varrho_{\mathfrak{M}}(x)$ of all ϱ-formulas that are true at x in $\mathfrak{M}$, and the *atomic ϱ-type of $x \in W$ in $\mathfrak{M}$* is $at^\varrho_{\mathfrak{M}}(x) = t^\varrho_{\mathfrak{M}}(x) \cap \varrho$. For a set X of points in $\mathfrak{M}$, we let $t^\varrho_{\mathfrak{M}}(X) = \{t^\varrho_{\mathfrak{M}}(x) \mid x \in X\}$ and $at^\varrho_{\mathfrak{M}}(X) = \{at^\varrho_{\mathfrak{M}}(x) \mid x \in X\}$. A set Γ of

σ-formulas is *finitely satisfiable* in $\mathfrak{M}$ if, for every finite $\Gamma' \subseteq \Gamma$, there is $x' \in W$ such that $\Gamma' \subseteq t^\sigma_{\mathfrak{M}}(x')$; Γ is *satisfiable in* $\mathfrak{M}$ if $\Gamma \subseteq t^\sigma_{\mathfrak{M}}(x)$, for some $x \in W$.

A σ-model $\mathfrak{M}$ is *descriptive* if, for any $x, y \in W$ and any set Γ of σ-formulas,

(dif) $x = y$ iff $t^\sigma_{\mathfrak{M}}(x) = t^\sigma_{\mathfrak{M}}(y)$,

(ref) xRy iff $\Diamond t^\sigma_{\mathfrak{M}}(y) \subseteq t^\sigma_{\mathfrak{M}}(x)$ iff $\{\varphi \mid \Box\varphi \in t^\sigma_{\mathfrak{M}}(x)\} \subseteq t^\sigma_{\mathfrak{M}}(y)$,

(com) if Γ is finitely satisfiable in $\mathfrak{M}$, then Γ is satisfiable in $\mathfrak{M}$.

(In other words, descriptive models are based on finitely generated descriptive frames for $\mathsf{wK4}$ [5].) We remind the reader that, for any σ-formula φ, we have $\varphi \in \mathsf{wK4}$ iff $\neg\varphi$ is not satisfiable in a (finite) σ-model iff $\neg\varphi$ is not satisfiable in a (finite) descriptive σ-model. The finite model property of $\mathsf{wK4}$ was established in [3,4,13]. The decision problem for $\mathsf{wK4}$ is PSPACE-complete [20]. For the *difference logic* DL, we have $\varphi \in \mathsf{DL}$ iff $\neg\varphi$ is not satisfiable in a polynomial-size σ-model based on a frame for DL; the decision problem for DL is coNP-complete [6].

In the remainder of this section, we present the technical tools and results we need for deciding the interpolant existence problem for DL and $\mathsf{wK4}$.

Given $\varrho \subseteq \sigma$, we call a cluster C *ϱ-maximal* in $\mathfrak{M}$ if, for any $x \in C$ and $y \in W$, whenever CRy and $t^\varrho_{\mathfrak{M}}(x) = t^\varrho_{\mathfrak{M}}(y)$ then $y \in C$. The following fundamental properties of descriptive σ-models for $\mathsf{wK4}$ (that are similar to the corresponding properties of finitely generated descriptive frames for $\mathsf{K4}$ [8,5]) are proved in Appendix A:

Lemma 2.1 *Suppose $\mathfrak{M}$ is a descriptive σ-model based on a $\mathsf{wK4}$-frame, $\varrho \subseteq \sigma$, C is a cluster in $\mathfrak{M}$, and Γ a set of σ-formulas. Then the following hold*:

(a) $|C| \leq 2^{|\sigma|}$;

(b) if $\mathfrak{M}, x \models \Diamond \bigwedge \Gamma'$ for every finite $\Gamma' \subseteq \Gamma$, then there is y with xRy and $\mathfrak{M}, y \models \Gamma$;

(c) there exists a ϱ-maximal cluster C' such that $CR^r C'$ and $t^\varrho_{\mathfrak{M}}(C) \subseteq t^\varrho_{\mathfrak{M}}(C')$.

Let $\mathfrak{M}_i$, $i = 1, 2$, be σ-models based on frames $\mathfrak{F}_i = (W_i, R_i)$ for $\mathsf{wK4}$ and let $\varrho \subseteq \sigma$. A relation $\boldsymbol{\beta} \subseteq W_1 \times W_2$ is called a *ϱ-bisimulation* between $\mathfrak{M}_1$ and $\mathfrak{M}_2$ in case the following conditions hold: whenever $x_1 \boldsymbol{\beta} x_2$,

(atom) $at^\varrho_{\mathfrak{M}_1}(x_1) = at^\varrho_{\mathfrak{M}_2}(x_2)$;

(move) if $x_1 R_1 y_1$, then there is y_2 such that $x_2 R_2 y_2$ and $y_1 \boldsymbol{\beta} y_2$; and, conversely, if $x_2 R_2 y_2$, then there is y_1 with $x_1 R_1 y_1$ and $y_1 \boldsymbol{\beta} y_2$.

If there is such $\boldsymbol{\beta}$ with $z_1 \boldsymbol{\beta} z_2$, we write $\mathfrak{M}_1, z_1 \sim^\varrho \mathfrak{M}_2, z_2$. The following characterisation of bisimulations between descriptive models in terms of types is well-known; see [10] and references therein:

Lemma 2.2 *For any $\varrho \subseteq \sigma$, descriptive σ-models $\mathfrak{M}_i$, $i = 1, 2$, and $x_i \in W_i$,*

$$t^\varrho_{\mathfrak{M}_1}(x_1) = t^\varrho_{\mathfrak{M}_2}(x_2) \quad \textit{iff} \quad \mathfrak{M}_1, x_1 \sim^\varrho \mathfrak{M}_2, x_2.$$

The implication ($\Leftarrow$) holds for arbitrary (not necessarily descriptive) models.

Variations of the next criterion of interpolant (non-)existence are implicit in various (dis-)proofs of the CIP in modal logics [16,10]:

Lemma 2.3 *Let $\sigma = sig(\varphi) \cup sig(\psi)$ and $L \in \{\mathsf{wK4}, \mathsf{DL}\}$. Then $\varphi \to \psi$ has no interpolant in L iff there are descriptive σ-models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ based on frames for L with roots r_φ and r_ψ, respectively, such that*

(*i*) $\mathfrak{M}_\varphi, r_\varphi \models \varphi$ *and* $\mathfrak{M}_\psi, r_\psi \models \neg\psi$;

(*ii*) $\mathfrak{M}_\varphi, r_\varphi \sim^\varrho \mathfrak{M}_\psi, r_\psi$, *where* $\varrho = sig(\varphi) \cap sig(\psi)$.

As both wK4 *and* DL *are canonical, the requirement that models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ be descriptive can be omitted.*

We first apply this criterion to decide the IEP for the difference logic DL. A key observation is that, from any two σ-models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ witnessing the nonexistence of an interpolant for a given $\varphi \to \psi$ in DL in the sense of Lemma 2.3, we can extract sub-models of polynomial size in $|\varphi|$ and $|\psi|$ that also satisfy the above criterion. We call this phenomenon the *polysize bisimilar model property* of DL, which clearly implies that the IEP for DL is decidable in coNP. Indeed, to check that $\varphi \to \psi$ has no interpolant in DL, we can guess polynomial-size $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ together with a relation $\boldsymbol{\beta}$ between them and then check whether they satisfy the criterion of Lemma 2.3.

We remind the reader that rooted frames for DL (and so the frames $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ are based on) are clusters, containing possibly both reflexive and irreflexive points. To show the polysize bisimilar model property of DL, we proceed in two steps. First, for every $\alpha \in sub(\varphi)$ ($\alpha \in sub(\psi)$) satisfiable in $\mathfrak{M}_\varphi$ (respectively, $\mathfrak{M}_\psi$), we pick two points x_α, x'_α satisfying α in $\mathfrak{M}_\varphi$ (in $\mathfrak{M}_\psi$) if they exist, otherwise a single such point x_α. Denote the set of the points selected this way by M_φ (M_ψ), assuming that $r_\varphi \in M_\varphi$ and $r_\psi \in M_\psi$. Let

$$T = \{t^\varrho_{\mathfrak{M}_\varphi}(x) \mid x \in M_\varphi\} \cup \{t^\varrho_{\mathfrak{M}_\psi}(x) \mid x \in M_\psi\}.$$

As $\mathfrak{M}_\varphi, r_\varphi \sim^\varrho \mathfrak{M}_\psi, r_\psi$, every ϱ-type $t \in T$ is satisfied in both $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$. Now, for each $t \in T$, we pick two distinct points satisfying t in $\mathfrak{M}_\varphi$, if they exist, and otherwise a single such point, and add them to M_φ if they were not already there. We do the same for $\mathfrak{M}_\psi$ and M_ψ. Let $\mathfrak{M}^\dagger_\varphi$ and $\mathfrak{M}^\dagger_\psi$ be the restrictions of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ to the resulting M_φ and M_ψ; and let $\boldsymbol{\beta}^\dagger$ be the restriction of $\boldsymbol{\beta}$ to $M_\varphi \times M_\psi$. It is readily seen that $\mathfrak{M}^\dagger_\varphi, r_\varphi \models \varphi$, $\mathfrak{M}^\dagger_\psi, r_\psi \models \neg\psi$ and $\boldsymbol{\beta}^\dagger$ is a ϱ-bisimulation between $\mathfrak{M}^\dagger_\varphi$ and $\mathfrak{M}^\dagger_\psi$ with $r_\varphi \boldsymbol{\beta}^\dagger r_\psi$. By the construction, $|M_\varphi|$ and $|M_\psi|$ are polynomial in $|\varphi|$ and $|\psi|$. Thus, we obtain:

Theorem 2.4 (*i*) DL *enjoys the polysize bisimilar model property.*

(*ii*) *The interpolant existence property for* DL *is* coNP*-complete.*

3 Deciding Interpolant Existence for wK4

In this section we show that wK4 has the 3-exponential-size bisimilar model property, which means that the IEP is decidable in coN3ExpTime.

Given formulas φ and ψ, let $sub(\varphi, \psi) = sub(\varphi) \cup sub(\psi)$, $\sigma = sig(\varphi) \cup sig(\psi)$ and $\varrho = sig(\varphi) \cap sig(\psi)$. If $\varphi \to \psi$ does not have an interpolant in $\mathsf{wK4}$, then Lemma 2.3 provides two pointed descriptive σ-models $\mathfrak{M}_\varphi, r_\varphi$ and $\mathfrak{M}_\psi, r_\psi$ based on weakly transitive frames. To simplify notation, we will operate with a single descriptive σ-model $\mathfrak{M}$ based on a weakly transitive frame $\mathfrak{F} = (W, R)$—the disjoint union of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$—containing two points r_φ, r_ψ such that $\mathfrak{M}, r_\varphi \models \varphi$, $\mathfrak{M}, r_\psi \models \neg\psi$ and $\mathfrak{M}, r_\varphi \sim^\varrho \mathfrak{M}, r_\psi$. Our aim is to convert $\mathfrak{M}$ into a model $\mathfrak{M}^\dagger$ based on a weakly transitive frame $\mathfrak{F}^\dagger = (W^\dagger, R^\dagger)$ that still witnesses the lack of an interpolant for $\varphi \to \psi$ in the above sense, and has $W^\dagger$ of triple-exponential size in $|sub(\varphi, \psi)|$.

Given a point x in $\mathfrak{M}$, we define the *φ, ψ-type* $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) = t^\sigma_{\mathfrak{M}}(x) \cap sub(\varphi, \psi)$. For a set X of points in $\mathfrak{M}$, we let $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(X) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \mid x \in X\}$. Our construction is an elaborate $sub(\varphi, \psi)$-filtration. (For the well-known *filtration* techniques in modal logic see, e.g., [5].) As usual, it keeps track of the φ, ψ-types of points in $\mathfrak{M}$ to satisfy condition (i) of Lemma 2.3. In addition, some connections of these φ, ψ-types with the ϱ-types of points in $\mathfrak{M}$ are also noted in order to satisfy condition (ii) of Lemma 2.3. When applied to satisfiability checking, our construction reduces to the filtration of [13]; see Remark 3.12.

To begin with, we define an equivalence relation $\approx$ between clusters in $\mathfrak{M}$ by taking $C \approx C'$ iff there exists a sequence $C = C_0, \dots, C_n = C'$ of clusters in $\mathfrak{M}$ such that, for each $i < n$ there are $x_i \in C_i$ and $y_{i+1} \in C_{i+1}$ with $t^\varrho_{\mathfrak{M}}(x_i) = t^\varrho_{\mathfrak{M}}(y_{i+1})$. Let $[C] = \{C' \mid C' \approx C\}$ and $T^\varrho[C] = \bigcup_{C' \in [C]} t^\varrho_{\mathfrak{M}}(C')$. It follows from the definition that $[C] = \{C' \mid t^\varrho_{\mathfrak{M}}(C') \cap T^\varrho[C] \neq \emptyset\}$. By Lemma 2.1, for every $C' \in [C]$, there is a ϱ-maximal $D \in [C]$ with $C' R^r D$ and $t^\varrho_{\mathfrak{M}}(C') \subseteq t^\varrho_{\mathfrak{M}}(D)$.

Lemma 3.1 *If $D \in [C]$ is a ϱ-maximal cluster, then $T^\varrho[C] = t^\varrho_{\mathfrak{M}}(D)$, and so $|T^\varrho[C]| \le 2^{|\varrho|}$.*

Proof. Let $D \in [C]$ be a ϱ-maximal cluster. Clearly, it is enough to prove that $T^\varrho[C] \subseteq t^\varrho_{\mathfrak{M}}(D)$. Let $x \in C$. We need to show that there exists $y \in D$ with $t^\varrho_{\mathfrak{M}}(y) = t^\varrho_{\mathfrak{M}}(x)$. By the definition of $\approx$, we have $z \in D$ with such that $\Diamond^n t^\varrho_{\mathfrak{M}}(z) \subseteq t^\varrho_{\mathfrak{M}}(x)$ and $\Diamond^n t^\varrho_{\mathfrak{M}}(x) \subseteq t^\varrho_{\mathfrak{M}}(z)$, for some $n \ge 0$. If $n = 0$, then we take $y = z$. If $n > 0$, then by Lemma 2.1, there are $u, v \in W$ with $z R^n u R^n v$, $t^\varrho_{\mathfrak{M}}(u) = t^\varrho_{\mathfrak{M}}(x)$ and $t^\varrho_{\mathfrak{M}}(v) = t^\varrho_{\mathfrak{M}}(z)$. By weak transitivity, $z = v$ or zRv. In either case, by the ϱ-maximality of D, we must have $v \in D$, and so $u \in D$, yielding $y = u$. $\square$

For every equivalence class $[C]$ in $\mathfrak{M}$, we let $AT^\varrho[C] = \{t \cap \varrho \mid t \in T^\varrho[C]\}$. Given a cluster C, we define the *cluster-type* of C in $\mathfrak{M}$ as the function $\tau_C \colon AT^\varrho[C] \to 2^{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(C)}$ where, for any $a \in AT^\varrho[C]$,

$$\tau_C(a) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \mid x \in C,\ \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \cap \varrho = at^\varrho_{\mathfrak{M}}(x) = a\}.$$

Observe that, by Lemma 3.1, $AT^\varrho[C'] = AT^\varrho[C]$ for every $C' \in [C]$, and so τ_C and $\tau_{C'}$ have the same domain $\mathsf{dom}\,\tau_C = \mathsf{dom}\,\tau_{C'} = AT^\varrho[C]$. As $\bigcup_{a \in \mathsf{dom}\,\tau_C} \tau_C(a) = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(C)$, τ_C keeps a record of both φ, ψ-types and atomic

ϱ-types of points in C in the context of the whole equivalence class $[C]$. Also, $\tau_C(a)$ might be empty for some $a \in \mathsf{dom}\,\tau_C$, but if C is ϱ-maximal then $\tau_C(a) \neq \emptyset$ for all $a \in \mathsf{dom}\,\tau_C$. Note that the number of pairwise distinct cluster-types in $\mathfrak{M}$ does not exceed $2^{|\varrho|} \cdot \big(2^{2^{|sub(\varphi,\psi)|}}\big)^{2^{|\varrho|}} = 2^{|\varrho|+2^{|sub(\varphi,\psi)|+|\varrho|}}$.

By the *mosaic of* $[C]$ *in* $\mathfrak{M}$ we mean the set $M_{[C]} = \{\tau_{C'} \mid C' \in [C]\}$ of cluster-types of the same domain (also called as the *domain of* $M_{[C]}$ and denoted by $\mathsf{dom}\,M_{[C]}$). M is a *mosaic in* $\mathfrak{M}$ if $M = M_{[C]}$ for some C. Clearly, the number of pairwise distinct mosaics in $\mathfrak{M}$ is $\mathcal{O}\big(2^{2^{2^{|sub(\varphi,\psi)|}}}\big)$.

We are now in a position to define the model $\mathfrak{M}^\dagger = (\mathfrak{F}^\dagger, \mathfrak{v}^\dagger)$ and its underlying frame $\mathfrak{F}^\dagger = (W^\dagger, R^\dagger)$. Suppose $x \in W$. Then we set

$$\boldsymbol{w}(x) = \big(\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x), at^{\varrho}_{\mathfrak{M}}(x), \tau_{C(x)}, M_{[C(x)]}\big) \quad \text{and} \quad W^\dagger = \{\boldsymbol{w}(x) \mid x \in W\}.$$

Observe that if $\boldsymbol{w} \in W^\dagger$ and $\boldsymbol{w} = (\boldsymbol{t}, a, \tau, M)$, then $a = \boldsymbol{t} \cap \varrho$, $\tau \in M$ and $\boldsymbol{t} \in \tau(a)$ always hold. Moreover,

$$(\boldsymbol{t}, a, \tau, M) \in W^\dagger, \text{ for all mosaics } M,\ \tau \in M,\ a \in \mathsf{dom}\,M, \text{ and } \boldsymbol{t} \in \tau(a). \tag{2}$$

We call M and a the *mosaic* and the *ϱ-index* of $\boldsymbol{w}$, respectively. Later on, we shall see that $\boldsymbol{w}$ and $\boldsymbol{w}'$ are ϱ-bisimilar if they share the same mosaic and ϱ-index. By the above calculations, $|W^\dagger| = \mathcal{O}\big(2^{2^{2^{|sub(\varphi,\psi)|}}}\big)$. The valuation $\mathfrak{v}^\dagger$ on $W^\dagger$ is inherited from $\mathfrak{v}$ in $\mathfrak{M}$: for every $p \in \sigma$,

$$\mathfrak{v}^\dagger(p) = \{\boldsymbol{w}(x) \mid x \in \mathfrak{v}(p)\} = \{\boldsymbol{w}(x) \mid p \in t^{\varphi,\psi}_{\mathfrak{M}}(x),\ x \in W\}.$$

To define the accessibility relation $R^\dagger$ in $\mathfrak{F}^\dagger$, we require some new notions. Let $\boldsymbol{t}$ and $\boldsymbol{t}'$ be φ,ψ-types, τ_C and $\tau_{C'}$ cluster-types, M and M' mosaics. Define a relation $\twoheadrightarrow$ between such pairs by taking:

- $\boldsymbol{t} \twoheadrightarrow \boldsymbol{t}'$ iff, for every $\Diamond\psi \in sub(\varphi,\psi)$, whenever χ or $\Diamond\chi$ is in $\boldsymbol{t}'$, then $\Diamond\chi \in \boldsymbol{t}$;
- $\tau_C \twoheadrightarrow \tau_{C'}$ iff $\boldsymbol{t} \twoheadrightarrow \boldsymbol{t}'$ for all $\boldsymbol{t} \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(C)$ and $\boldsymbol{t}' \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(C')$;
- $M \twoheadrightarrow M'$ iff for every $\tau \in M$ there is $\tau' \in M'$ with $\tau \twoheadrightarrow \tau'$.

It is readily checked that the defined relation $\twoheadrightarrow$ has the following properties:

$$\twoheadrightarrow \text{ is transitive in all three settings;} \tag{3}$$
$$\text{if } xRy, \text{ then } \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y); \tag{4}$$
$$\text{if } CRC' \text{ and } C \neq C', \text{ then } \tau_C \twoheadrightarrow \tau_{C'}. \tag{5}$$

Note that CRC does not necessarily imply $\tau_C \twoheadrightarrow \tau_C$; see Example 3.2. By (5), Lemmas 2.1 (c) and 3.1, we also have that

$$\text{for all } C \text{ there is a } \varrho\text{-maximal } D \in [C] \text{ such that } C = D \text{ or } \tau_C \twoheadrightarrow \tau_D. \tag{6}$$

Next, for any mosaic M, we define a subset I_M of $\mathsf{dom}\,M$ by taking

$$I_M = \{a \in \mathsf{dom}\,M \mid \Diamond t \not\subseteq t, \text{ for all clusters } C \text{ with } M = M_{[C]} \text{ and all } t \in T^{\varrho}[C] \text{ with } t \cap \varrho = a\}.$$

As $a \in I_M$ implies that for every cluster C with $M = M_{[C]}$ there is at most one $x \in C$ such that $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \in \tau_C(a)$ and such an x is irreflexive, it follows that

$$\text{if } \tau_C \in M, \text{ then } |\tau_C(a)| \leq 1 \text{ for every } a \in I_M. \tag{7}$$

We also claim that

$$\text{if } C \text{ is not } \varrho\text{-maximal and } \tau_C \in M, \text{ then } \tau_C(a) = \emptyset \text{ for every } a \in I_M. \tag{8}$$

Indeed, by Lemmas 2.1 (c) and 3.1, there is a ϱ-maximal $D \in [C(x)]$ with $C \neq D$ and CRD. Let $a \in \mathsf{dom}\, M_{[C]}$ be such that $\tau_C(a) \neq \emptyset$. There there is $x \in C$ with $t^{\varrho}_{\mathfrak{M}}(x) \cap \varrho = a$. By Lemma 3.1, there is $y \in D$ with $t^{\varrho}_{\mathfrak{M}}(y) = t^{\varrho}_{\mathfrak{M}}(x)$. Thus, $\Diamond t^{\varrho}_{\mathfrak{M}}(x) \subseteq t^{\varrho}_{\mathfrak{M}}(x)$, and so $a \notin I_{M_{[C]}}$.

Now, to define $R^\dagger$ on $W^\dagger$, suppose $\boldsymbol{w}, \boldsymbol{w}' \in W^\dagger$, where

$$\boldsymbol{w} = (\boldsymbol{t}, a, \tau, M), \qquad \boldsymbol{w}' = (\boldsymbol{t}', a', \tau', M').$$

The definition of $R^\dagger$ for $\boldsymbol{w}$, $\boldsymbol{w}'$ depends on whether $M = M'$ and $M \twoheadrightarrow M$:

Case $M \neq M'$: then $\boldsymbol{w} R^\dagger \boldsymbol{w}'$ iff $M \twoheadrightarrow M'$ and $\tau \twoheadrightarrow \tau'$.

Case $M = M'$, $M \twoheadrightarrow M$: then $\boldsymbol{w} R^\dagger \boldsymbol{w}'$ iff
- either $\boldsymbol{w} \neq \boldsymbol{w}'$ and ($\tau = \tau'$ or $\tau \twoheadrightarrow \tau'$),
- or $\boldsymbol{w} = \boldsymbol{w}'$ and $\boldsymbol{t} \twoheadrightarrow \boldsymbol{t}$.

Case $M = M'$, $M \not\twoheadrightarrow M$: then $\boldsymbol{w} R^\dagger \boldsymbol{w}'$ iff
- either $\boldsymbol{w} \neq \boldsymbol{w}'$ and $\big(\tau = \tau'$ or $(\tau \twoheadrightarrow \tau'$ and $\tau(b) = \emptyset$ for all $b \in I_M)\big)$,
- or $\boldsymbol{w} = \boldsymbol{w}'$, $\boldsymbol{t} \twoheadrightarrow \boldsymbol{t}$, and $a \notin I_M$.

Example 3.2 Consider φ, ψ, $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ with $\sigma = \{p, q\}$ and $\varrho = \emptyset$ from Example 1.1 (i). Then $sub(\varphi, \psi)$ consists of the formulas p, q, $\Diamond p$, $\Diamond\neg q$, $\Diamond\Diamond p$, $\Diamond\Diamond\neg q$, φ, ψ and their negations. Let $\mathfrak{M}$ be the disjoint union of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$. Then $C(r_\varphi) = \{r_\varphi, x_1\} = C(x_1)$, $C(r_\psi) = \{r_\psi\}$, $C(x_2) = \{x_2\}$, with all of these clusters being $\approx$-equivalent, $T^\varrho[C(r_\varphi)] = \{t\}$ for $t = t^{\varrho}_{\mathfrak{M}}(r_\varphi) = \{\Diamond^n \top \mid n < \omega\}$, and $\mathsf{dom}\, \tau_{C(r_\varphi)} = \mathsf{dom}\, \tau_{C(r_\psi)} = \mathsf{dom}\, \tau_{C(x_2)} = AT^\varrho[C(r_\varphi)] = \{\emptyset\}$. This gives $\tau_{C(r_\varphi)}(\emptyset) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi), \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_1)\}$, where

$$\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi) = \{p, \neg q, \neg\Diamond p, \Diamond\neg q, \Diamond\Diamond p, \Diamond\Diamond\neg q, \varphi, \psi\},$$
$$\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_1) = \{\neg p, \neg q, \Diamond p, \Diamond\neg q, \Diamond\Diamond p, \Diamond\Diamond\neg q, \neg\varphi, \psi\}.$$

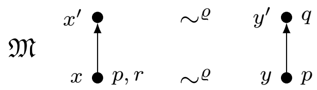
On the other hand, $\tau_{C(r_\psi)}(\emptyset) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi)\}$ and $\tau_{C(x_2)}(\emptyset) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_2)\}$, where

$$\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi) = \{\neg p, \neg q, \neg\Diamond p, \neg\Diamond\neg q, \neg\Diamond\Diamond p, \neg\Diamond\Diamond\neg q, \neg\varphi, \neg\psi\},$$
$$\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_2) = \{\neg p, q, \neg\Diamond p, \neg\Diamond\neg q, \neg\Diamond\Diamond p, \neg\Diamond\Diamond\neg q, \neg\varphi, \psi\}.$$

Then $\boldsymbol{w}(r_\varphi)$, $\boldsymbol{w}(x_1)$, $\boldsymbol{w}(r_\psi)$, and $\boldsymbol{w}(x_2)$ are all different, but they share the same ϱ-index $\emptyset$ and the same mosaic $M_{[C(r_\varphi)]} = \{\tau_{C(r_\varphi)},\ \tau_{C(r_\psi)}, \tau_{C(x_2)}\}$ with $\mathsf{dom}\, M_{[C(r_\varphi)]} = \{\emptyset\}$. As $\Diamond t \subseteq t$, we have $I_{M_{[C(r_\varphi)]}} = \emptyset$.

We have $C(r_\varphi)RC(r_\varphi)$ but $\tau_{C(r_\varphi)} \not\twoheadrightarrow \tau_{C(r_\varphi)}$ as $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi) \not\twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi)$ because $p \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi)$ but $\Diamond p \notin \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\varphi)$. So $\boldsymbol{w}(r_\varphi)R^\dagger\boldsymbol{w}(r_\varphi)$ does not hold. In fact, it is not hard to check that $\mathfrak{M}^\dagger$ is isomorphic to $\mathfrak{M}$. For example, we do not have $\boldsymbol{w}(r_\psi)R^\dagger\boldsymbol{w}(r_\psi)$ as $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi) \not\twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi)$ because $\neg q \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi)$ but $\Diamond\neg q \notin \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(r_\psi)$. But we obtain $\boldsymbol{w}(r_\varphi)R^\dagger\boldsymbol{w}(x_1)$ because they share the same cluster-type $\tau_{C(r_\varphi)}$, and $\boldsymbol{w}(x_1)R^\dagger\boldsymbol{w}(x_1)$ because $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_1) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x_1)$. $\dashv$

Example 3.3 We now illustrate the role of the set I_M in the definition of $R^\dagger$. Consider the model $\mathfrak{M}$ below and $\varphi = p \wedge \neg\Diamond q$ and $\psi = \neg p \wedge \Diamond r$ with $\varrho = \{p\}$ and $\sigma = \{p, q, r\}$. (The form of φ, ψ is not important here, but $sub(\varphi, \psi)$ is.)



Then $t^\varrho_{\mathfrak{M}}(x) = t^\varrho_{\mathfrak{M}}(y) = t$ with $\Diamond t \not\subseteq t$. Let $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) = \boldsymbol{t}_x$ and $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y) = \boldsymbol{t}_y$. Then $\boldsymbol{t}_x \not\twoheadrightarrow \boldsymbol{t}_x$ because $r \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$ and $\Diamond r \notin \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$, while $\boldsymbol{t}_y \twoheadrightarrow \boldsymbol{t}_y$ because $\Diamond p \notin sub(\varphi, \psi)$; we also have $\boldsymbol{t}_x \not\twoheadrightarrow \boldsymbol{t}_y$ and $\boldsymbol{t}_y \not\twoheadrightarrow \boldsymbol{t}_x$. Now, consider

$$\boldsymbol{w}_x = (\boldsymbol{t}_x, \{p\}, \tau_{C(x)}, M), \quad \boldsymbol{w}_y = (\boldsymbol{t}_y, \{p\}, \tau_{C(y)}, M),$$

where $T^\varrho[C(x)] = \{t\}$, $[C(x)] = \{C(x), C(y)\}$, and $M = \{\tau_{C(x)}, \tau_{C(y)}\}$ with $\mathsf{dom}\, M = AT^\varrho[C(x)] = \big\{\{p\}\big\}$, $\tau_{C(x)}(\{p\}) = \{\boldsymbol{t}_x\}$, and $\tau_{C(y)}(\{p\}) = \{\boldsymbol{t}_y\}$. Then $M \not\twoheadrightarrow M$ and $I_M = \big\{\{p\}\big\}$. By the last item in the definition of $R^\dagger$, neither $\boldsymbol{w}_xR^\dagger\boldsymbol{w}_x$ nor $\boldsymbol{w}_yR^\dagger\boldsymbol{w}_y$ holds because $\{p\} \in I_M$. However, without the condition $a \notin I_M$ in the definition, we would have $\boldsymbol{w}_yR^\dagger\boldsymbol{w}_y$ but still not $\boldsymbol{w}_xR^\dagger\boldsymbol{w}_x$, which would destroy the ϱ-bisimilarity of $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$. $\dashv$

The following lemma is proved in Appendix A by a straightforward checking of all the cases in the definition of $R^\dagger$:

Lemma 3.4 *The relation $R^\dagger$ on $W^\dagger$ is weakly transitive.*

The next lemma says that $R^\dagger$ contains the smallest $sub(\varphi, \psi)$-filtration:

Lemma 3.5 *For all $x, y \in W$, if xRy, then $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$.*

Proof. Suppose we have xRy. To begin with, we claim that

$$\text{if } [C(x)] \neq [C(y)] \text{ then } M_{[C(x)]} \twoheadrightarrow M_{[C(y)]}. \tag{9}$$

Indeed, take any $\tau \in M_{[C(x)]}$ and let $C \in [C(x)]$ be such that $\tau_C = \tau$. By (6), there is a ϱ-maximal $D \in [C(x)]$ with $C = D$ or $\tau_C \twoheadrightarrow \tau_D$. So by Lemma 3.1, there is $x' \in D$ with $t^\varrho_{\mathfrak{M}}(x) = t^\varrho_{\mathfrak{M}}(x')$. As xRy and $\mathfrak{M}$ is descriptive, it follows from Lemma 2.2 that there is y' with $x'Ry'$ and $t^\varrho_{\mathfrak{M}}(y') = t^\varrho_{\mathfrak{M}}(y)$. Then $C(y') \in [C(y)] \neq [C(x)] = [C(x')] = [D]$. Therefore, $\tau_D \twoheadrightarrow \tau_{C(y')}$ by (5), and so $\tau_C \twoheadrightarrow \tau_{C(y')}$ by (3), as required.

Now we show that $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$ follows from (9) in all cases of the definition of $R^\dagger$. Assume first that $M_{[C(x)]} \neq M_{[C(y)]}$. Then $[C(x)] \neq [C(y)]$, and so $M_{[C(x)]} \twoheadrightarrow M_{[C(y)]}$ by (9), and $\tau_{C(x)} \twoheadrightarrow \tau_{C(y)}$ by (5), as required.

Assume next that $M_{[C(x)]} = M_{[C(y)]}$ and $M_{[C(x)]} \twoheadrightarrow M_{[C(y)]}$. By (5), $\tau_{C(x)} \not\twoheadrightarrow \tau_{C(y)}$ implies $C(x) = C(y)$, and so $\tau_{C(x)} = \tau_{C(y)}$. So we have $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$ when $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$. If $\boldsymbol{w}(x) = \boldsymbol{w}(y)$ then $t^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow t^{\varphi,\psi}_{\mathfrak{M}}(y)$ follows by (4), and so we also have $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$.

Finally, assume that $M_{[C(x)]} = M_{[C(y)]}$, $M_{[C(x)]} \not\twoheadrightarrow M_{[C(y)]}$. Then (9) implies that $[C(x)] = [C(y)]$. There are two cases:

Case $C(x)$ is ϱ-maximal: Then $C(x) = C(y)$, and so $\tau_{C(x)} = \tau_{C(y)}$. Thus, we have $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$ if $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$. If $\boldsymbol{w}(x) = \boldsymbol{w}(y)$ then we have $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$ by (4). As $at^{\varrho}_{\mathfrak{M}}(x) = at^{\varrho}_{\mathfrak{M}}(y)$ and $C(x) = C(y)$, we have $t^{\varrho}_{\mathfrak{M}}(x) = t^{\varrho}_{\mathfrak{M}}(y)$, and so $\Diamond t^{\varrho}_{\mathfrak{M}}(x) \subseteq t^{\varrho}_{\mathfrak{M}}(x)$ by xRy. Therefore, $at^{\varrho}_{\mathfrak{M}}(x) \notin I_{M_{[C(x)]}}$, and so $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$.

Case $C(x)$ is not ϱ-maximal: Then, by (8), $\tau_{C(x)}(a) = \emptyset$ for every $a \in I_{M_{[C(x)]}}$. In particular, as $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \in \tau_{C(x)}(at^{\varrho}_{\mathfrak{M}}(x))$, it follows that $at^{\varrho}_{\mathfrak{M}}(x) \notin I_{M_{[C(x)]}}$. If $\boldsymbol{w}(x) = \boldsymbol{w}(y)$ then we have $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$ by (4), and so $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$. If $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$ and $C(x) = C(y)$, then $\tau_{C(x)} = \tau_{C(y)}$, and so $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$. And if $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$ and $C(x) \neq C(y)$, then $\tau_{C(x)} \twoheadrightarrow \tau_{C(y)}$ by (5), and so $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$ again.

This completes the proof of the lemma. $\square$

The next lemma says that $R^\dagger$ is contained in the largest $sub(\varphi,\psi)$-filtration:

Lemma 3.6 *If* $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$, *then* $\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ *implies* $\Diamond\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$, *for every* $\Diamond\chi \in sub(\varphi,\psi)$.

Proof. As $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ implies that, for every $\Diamond\chi \in sub(\varphi,\psi)$, whenever $\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ then $\Diamond\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$, and $\tau_{C(x)} \twoheadrightarrow \tau_{C(y)}$ implies $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$, we only need to check those cases when $\boldsymbol{w}(x)R^\dagger\boldsymbol{w}(y)$ but neither $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \twoheadrightarrow \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ nor $\tau_{C(x)} \twoheadrightarrow \tau_{C(y)}$ hold.

An inspection of the definition of $R^\dagger$ shows that this can only happen when $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$, $M_{[C(x)]} = M_{[C(y)]}$ and $\tau_{C(x)} = \tau_{C(y)}$. In this case, $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) \in \tau_{C(x)}\big(at^{\varrho}_{\mathfrak{M}}(x)\big) = \tau_{C(y)}\big(at^{\varrho}_{\mathfrak{M}}(x)\big)$, and so there is $y' \in C(y)$ with $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y') = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$. Then $y' \neq y$ follows, as otherwise we would have $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$, and so $at^{\varrho}_{\mathfrak{M}}(x) = at^{\varrho}_{\mathfrak{M}}(y)$ as well, contradicting $\boldsymbol{w}(x) \neq \boldsymbol{w}(y)$. Now it follows that for every $\Diamond\chi \in sub(\varphi,\psi)$, whenever $\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ then $\Diamond\chi \in \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y') = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)$, as required. $\square$

As a consequence of Lemmas 3.5 and 3.6, we obtain the usual 'filtration lemma' for $\mathfrak{M}^\dagger$ that can be proved by induction on χ:

Lemma 3.7 *For any* $\chi \in sub(\varphi,\psi)$ *and any* $\boldsymbol{w} = (\boldsymbol{t}, a, \tau, M) \in W^\dagger$, *we have* $\mathfrak{M}^\dagger, \boldsymbol{w} \models \chi$ *iff* $\chi \in \boldsymbol{t}$.

As a consequence of Lemma 3.7 we obtain:

Corollary 3.8 $\mathfrak{M}^\dagger, \boldsymbol{w}(r_\varphi) \models \varphi$ *and* $\mathfrak{M}^\dagger, \boldsymbol{w}(r_\psi) \models \neg\psi$.

Define a binary relation $\boldsymbol{\beta}^\dagger$ on $W^\dagger$ by taking

$$(\boldsymbol{t}, a, \tau, M)\, \boldsymbol{\beta}^\dagger\, (\boldsymbol{t}', a', \tau', M') \quad \text{iff} \quad a = a',\ M = M'.$$

Lemma 3.9 *The relation $\boldsymbol{\beta}^\dagger$ is a ϱ-bisimulation on $\mathfrak{M}^\dagger$ with $\boldsymbol{w}(r_\varphi)\boldsymbol{\beta}^\dagger\boldsymbol{w}(r_\psi)$.*

Proof. Since $\mathfrak{M}, r_\varphi \sim^\varrho \mathfrak{M}, r_\psi$, we have $t^\varrho_{\mathfrak{M}}(r_\varphi) = t^\varrho_{\mathfrak{M}}(r_\psi)$. Thus, $at^\varrho_{\mathfrak{M}}(r_\varphi) = at^\varrho_{\mathfrak{M}}(r_\psi)$ and $[C(r_\varphi)] = [C(r_\psi)]$, and so $\boldsymbol{w}(r_\varphi)\boldsymbol{\beta}^\dagger\boldsymbol{w}(r_\psi)$.

Condition **(atom)** follows from Lemma 3.7. To prove **(move)**, suppose $\boldsymbol{w}_1\boldsymbol{\beta}^\dagger\boldsymbol{w}'_1$ and $\boldsymbol{w}_1 R^\dagger \boldsymbol{w}_2$, for $\boldsymbol{w}_1 = (\boldsymbol{t}_1, a_1, \tau_1, M_1)$, $\boldsymbol{w}'_1 = (\boldsymbol{t}'_1, a_1, \tau'_1, M_1)$, and $\boldsymbol{w}_2 = (\boldsymbol{t}_2, a_2, \tau_2, M_2)$. We show that there is $\boldsymbol{w}'_2$ with $\boldsymbol{w}_2\boldsymbol{\beta}^\dagger\boldsymbol{w}'_2$ and $\boldsymbol{w}'_1 R^\dagger \boldsymbol{w}'_2$, that is, there exist $\boldsymbol{t}'_2$ and τ'_2 such that $\boldsymbol{w}'_2 = (\boldsymbol{t}'_2, a_2, \tau'_2, M_2) \in W^\dagger$ and $\boldsymbol{w}'_1 R^\dagger \boldsymbol{w}'_2$. We proceed by case distinction.

Case $M_1 \neq M_2$: As $\boldsymbol{w}_1 R^\dagger \boldsymbol{w}_2$, we have $M_1 \twoheadrightarrow M_2$. As $\tau'_1 \in M_1$, there is some $\tau'_2 \in M_2$ with $\tau'_1 \twoheadrightarrow \tau'_2$. By (3) and (6), we may assume that $\tau'_2 = \tau_D$ for some ϱ-maximal D, and so $\tau'_2(a_2) \neq \emptyset$. Take any $\boldsymbol{t}'_2 \in \tau'_2(a_2)$. Then τ'_2 and $\boldsymbol{t}'_2$ are as required, by (2).

Case $M_1 = M_2$, $M_1 \twoheadrightarrow M_1$: As $\tau'_1 \in M_1$, there is $\tau'_2 \in M_1$ with $\tau'_1 \twoheadrightarrow \tau'_2$. By (3) and (6), we may assume that $\tau'_2 = \tau_D$ for some ϱ-maximal D, and so $\tau'_2(a_2) \neq \emptyset$. Take any $\boldsymbol{t}'_2 \in \tau'_2(a_2)$. Then $\boldsymbol{w}'_2 = (\boldsymbol{t}'_2, a_2, \tau'_2, M_1) \in W^\dagger$, by (2). As $\tau'_1 \twoheadrightarrow \tau'_2$ implies $\boldsymbol{t}'_1 \twoheadrightarrow \boldsymbol{t}'_2$, we have $\boldsymbol{w}'_1 R^\dagger \boldsymbol{w}'_2$ if $\boldsymbol{w}'_1 = \boldsymbol{w}'_2$ or $\boldsymbol{w}'_1 \neq \boldsymbol{w}'_2$.

Case $M_1 = M_2$, $M_1 \not\twoheadrightarrow M_2$, $a_1 \neq a_2$: If $\tau'_1(a_2) \neq \emptyset$, then take any $\boldsymbol{t}'_2 \in \tau'_2(a_2)$. Then $\boldsymbol{w}'_2 = (\boldsymbol{t}'_2, a_2, \tau'_1, M_1)$ is as required. If $\tau'_1(a_2) = \emptyset$, then $\tau'_1 \neq \tau_D$ for any ϱ-maximal D. Thus, by (8), $\tau'_1(a) = \emptyset$ for every $a \in I_{M_1}$. Also, by (6), there is some ϱ-maximal D with $\tau_D \in M_1$ and $\tau'_1 \twoheadrightarrow \tau_D$. Take any $\boldsymbol{t}'_2 \in \tau_D(a_2)$. By (2), $\boldsymbol{w}'_2 = (\boldsymbol{t}'_2, a_2, \tau_D, M_1)$ is as required.

Case $M_1 = M_2$, $M_1 \not\twoheadrightarrow M_2$, $a_1 = a_2$: We claim that $a_1 \notin I_{M_1}$. Indeed, suppose $a_1 \in I_{M_1}$. As $\boldsymbol{w}_1 R^\dagger \boldsymbol{w}_2$, $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$ follows. If $\tau_1 = \tau_2$ held, then $\{\boldsymbol{t}_1\} = \tau_1(a_1) = \tau_2(a_1) = \{\boldsymbol{t}_2\}$ by (7), and so $\boldsymbol{t}_1 = \boldsymbol{t}_2$ would follow, contradicting $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$. So $\tau_1 \neq \tau_2$, and thus $\boldsymbol{w}_1 R^\dagger \boldsymbol{w}_2$ implies that $\tau_1(a) = \emptyset$ for all $a \in I_{M_1}$. As $\boldsymbol{t}_1 \in \tau_1(a_1)$, $a_1 \notin I_{M_1}$ follows, as required.

As $a_1 \notin I_{M_1}$, there exist C with $M_1 = M_{[C]}$ and $t \in T^\varrho[C]$ such that $t \cap \varrho = a_1$ and $\Diamond t \not\subseteq t$. By Lemma 3.1, it follows that, for every ϱ-maximal $D \in [C]$, either (i) there are at least two $x \in D$ with $t^\varrho_{\mathfrak{M}}(x) = t$ or (ii) xRx for the single $x \in D$ with $t^\varrho_{\mathfrak{M}}(x) = t$.

If $\tau'_1 = \tau_D$ for a ϱ-maximal $D \in [C]$, then in case (ii) we have $\tau'_1(a_1) = \{\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x)\} = \{\boldsymbol{t}'_1\}$ and $\boldsymbol{t}'_1 \twoheadrightarrow \boldsymbol{t}'_1$. Therefore, $\boldsymbol{w}'_2 = (\boldsymbol{t}'_1, a_1, \tau'_1, M_1)$ is as required. In case (i), if $\boldsymbol{t}'_1 \not\twoheadrightarrow \boldsymbol{t}'_1$, then there is $\boldsymbol{t} \in \tau'_1(a_1)$ with $\boldsymbol{t} \neq \boldsymbol{t}'_1$, and so $\boldsymbol{w}'_2 = (\boldsymbol{t}, a_1, \tau'_1, M_1)$ is as required.

If $\tau'_1 \neq \tau_D$ for any ϱ-maximal $D \in [C]$, then $\tau'_1(a) = \emptyset$ for every $a \in I_{M_1}$, by (8). By (6), there is a ϱ-maximal D such that $\tau_D \in M_1$ and $\tau'_1 \twoheadrightarrow \tau_D$. Take any $\boldsymbol{t}'_2 \in \tau_D(a_1)$. By (2), $\boldsymbol{w}'_2 = (\boldsymbol{t}'_2, a_1, \tau_D, M_1)$ is as required,

completing the proof of the lemma. □

The results obtained above yield the following:

Theorem 3.10 *Any given implication $\varphi \to \psi$ does not have an interpolant in* wK4 *iff there are models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ satisfying the criterion of Lemma 2.3 and having size triple-exponential in $|\varphi|$ and $|\psi|$.*

Thus, to decide whether $\varphi \to \psi$ does not have an interpolant in wK4, we can guess models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ of triple-exponential size in $|\varphi|$ and $|\psi|$ together with a binary relation $\sim^\varrho$ between their points and check in polynomial time in the size of $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ whether the conditions of Lemma 2.3 are met.

Theorem 3.11 *The IEP for* wK4 *is decidable in* coN3ExpTime.

Remark 3.12 We can use the above construction to check whether a formula ψ is in wK4 as follows. We clearly have $\psi \notin$ wK4 iff $\varphi \to \psi$ has no interpolant in wK4, for $\varphi = \psi \vee \neg\psi$. In this case, $\sigma = \varrho = sig(\psi)$, and if $\mathfrak{M}$ is a descriptive σ-model, then $\boldsymbol{w}(x) = \boldsymbol{w}(y)$ iff $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(x) = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}(y)$ and $\boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}\big(C(x)\big) = \boldsymbol{t}^{\varphi,\psi}_{\mathfrak{M}}\big(C(y)\big)$. Also, our filtration becomes the $sub(\psi)$-filtration given in [13]. Note that this filtration gives a double-exponential bound on the size of the model satisfying $\neg\psi$, which is not optimal as the decision problem for wK4 is PSpace-complete [20].

4 Lower Bound

Theorem 4.1 *The IEP for* wK4 *is* coNExpTime*-hard.*

Proof. We show NExpTime-hardness of interpolant non-existence by a reducion of the *exponential torus tiling problem*. A *tiling system* is a triple $P = (T, H, V)$, where T is a finite set of *tile types* and $H, V \subseteq T \times T$ are the *horizontal* and *vertical* matching conditions, respectively. An *initial condition* for P takes the form $\bar{t} = (t_0, \dots, t_{n-1}) \in T^n$. A map $\tau\colon \{0, \dots, 2^n - 1\} \times \{0, \dots, 2^n - 1\} \to T$ is a *solution for P and $\bar{t}$* if $\tau(i, 0) = t_i$ for all $i < n$, and for all $i, j < 2^n$, the following conditions hold (where $\oplus$ denotes addition modulo 2^n):

- if $\tau(i, j) = t$ and $\tau(i \oplus 1, j) = t'$, then $(t, t') \in H$;
- if $\tau(i, j) = t$ and $\tau(i, j \oplus 1) = t'$, then $(t, t') \in V$.

It is well-known that the problem of deciding whether there is a solution for given P and $\bar{t}$ is NExpTime-hard [2, Section 5.2.2].

Given a tiling system P and an initial condition $\bar{t}$ of length n, we define formulas φ, ψ of size polynomial in $|P|$ and n, such that, for $\sigma = sig(\varphi) \cup sig(\psi)$ and $\varrho = sig(\varphi) \cap sig(\psi)$,

$$\begin{aligned} &\text{there is a solution for } P \text{ and } \bar{t} \quad \text{iff} \\ &\qquad \text{there exist } \sigma\text{-models } \mathfrak{M}_\varphi \text{ and } \mathfrak{M}_\psi \text{ based on frames for wK4} \\ &\qquad \text{with } \mathfrak{M}_\varphi, r_\varphi \sim^\varrho \mathfrak{M}_\psi, r_\psi,\ \mathfrak{M}_\varphi, r_\varphi \models \varphi \text{ and } \mathfrak{M}_\psi, r_\psi \models \neg\psi. \end{aligned} \tag{10}$$

The shared signature ϱ consists of

- a variable e that will be used to force connections between the two ϱ-bisimilar models;
- a variable t, for each tile type $t \in T$; and

– variables $\mathsf{b}_0, \dots, \mathsf{b}_{2n-1}$, that serve as bits in the binary representation of *grid positions* (i, j) with $i, j < 2^n$. We will use $[\mathsf{b} = (i, j)]$ as a shorthand for the formula where $\mathsf{b}_0, \dots, \mathsf{b}_{n-1}$ represent the horizontal coordinate i and $\mathsf{b}_n, \dots, \mathsf{b}_{2n-1}$ the vertical coordinate j; for instance, $[\mathsf{b} = (2, 3)]$ stands for $\neg\mathsf{b}_0 \wedge \mathsf{b}_1 \wedge \mathsf{b}_n \wedge \mathsf{b}_{n+1} \wedge \bigwedge_{n+1<k<2n} \neg\mathsf{b}_k$.

The formula φ is defined as

$$\varphi = \mathsf{e} \wedge (\Diamond\Diamond p \wedge \neg\Diamond p) \wedge \Box(\mathsf{e} \to \Diamond p),$$

to which we add the other symbols in ϱ using tautologies. Observe that φ has the first formula of Example 1.1 (i) as a conjunct. Thus, if $\mathfrak{M}, x \models \varphi$, then x is irreflexive, $C(x)$ contains a point different from x, and e is true everywhere in $C(x)$ and nowhere else.

Our formula ψ takes form $\chi \to (\Box\Box q \to q)$ (cf. the second formula in Example 1.1 (i)). We next define χ. To begin with, χ has conjuncts that use variables $\mathsf{a}_0, \dots, \mathsf{a}_{2n-1}$ and variables $\mathsf{level}_0, \dots, \mathsf{level}_{2n}$ to generate a binary tree on nodes satisfying e such that a counter implemented using $\mathsf{a}_0, \dots, \mathsf{a}_{2n-1}$ is realised at its leaves:

$$\mathsf{level}_0 \wedge \Box^+ \bigwedge_{i<j\leq 2n} \neg(\mathsf{level}_i \wedge \mathsf{level}_j), \tag{11}$$

$$\Box^+\big(\mathsf{level}_i \to \Diamond(\mathsf{level}_{i+1} \wedge \mathsf{a}_i) \wedge \Diamond(\mathsf{level}_{i+1} \wedge \neg\mathsf{a}_i)\big), \quad \text{for } i < 2n, \tag{12}$$

$$\Box\big(\mathsf{level}_{i+1} \wedge \mathsf{a}_i \to \Box(\mathsf{level}_j \to \mathsf{a}_i)\big) \wedge \Box\big(\mathsf{level}_{i+1} \wedge \neg\mathsf{a}_i \to \Box(\mathsf{level}_j \to \neg\mathsf{a}_i)\big), \quad \text{for } i < j \leq 2n, \tag{13}$$

$$\Box^+(\mathsf{level}_i \to \mathsf{e}), \quad \text{for } i \leq 2n. \tag{14}$$

Next, we express that any leaf making $[\mathsf{a} = (i, j)]$ true has an R-successor making $\neg\mathsf{e} \wedge [\mathsf{b} = (i, j)]$ and a unique tile-variable t true, by defining

$$\mathsf{grid}: \quad \bigwedge_{k<2n} (\mathsf{a}_k \leftrightarrow \mathsf{b}_k),$$

and then adding the following conjuncts to χ:

$$\Box\Big(\mathsf{level}_{2n} \to \Diamond\big(\neg\mathsf{e} \wedge \mathsf{grid} \wedge \bigvee_{t\in T} \mathsf{t}\big)\Big), \tag{15}$$

$$\Box\big(\mathsf{level}_{2n} \wedge \mathsf{a}_i \to \Box(\neg\mathsf{e} \to \mathsf{a}_i)\big) \wedge \Box\big(\mathsf{level}_{2n} \wedge \neg\mathsf{a}_i \to \Box(\neg\mathsf{e} \to \neg\mathsf{a}_i)\big), \quad \text{for } i < 2n, \tag{16}$$

$$\Box \bigwedge_{t\neq t'\in T} \neg(\mathsf{t} \wedge \mathsf{t}'), \tag{17}$$

$$\Box\Big(\mathsf{level}_{2n} \to \bigwedge_{t\in T} \big(\Diamond(\neg\mathsf{e} \wedge \mathsf{grid} \wedge \mathsf{t}) \to \Box(\neg\mathsf{e} \wedge \mathsf{grid} \to \mathsf{t})\big)\Big). \tag{18}$$

The following formulas, respectively, express that '$[\mathsf{a} = (i, j)]$ and $[\mathsf{b} = (i \oplus 1, j)]$'

and '$[\mathsf{a} = (i,j)]$ and $[\mathsf{b} = (i,j\oplus 1)]$':

$$\mathsf{succ}_x: \quad \Big(\bigvee_{m<n} \big(\mathsf{b}_m \wedge \neg\mathsf{a}_m \wedge \bigwedge_{k<m} (\neg\mathsf{b}_k \wedge \mathsf{a}_k) \wedge \bigwedge_{m<k<n} (\mathsf{b}_k \leftrightarrow \mathsf{a}_k)\big) \vee \bigwedge_{m<n} (\neg\mathsf{b}_m \wedge \mathsf{a}_m)\Big) \wedge \bigwedge_{n\le k<2n} (\mathsf{b}_k \leftrightarrow \mathsf{a}_k),$$

$$\mathsf{succ}_y: \quad \Big(\bigvee_{n\le m<2n} \big(\mathsf{b}_m \wedge \neg\mathsf{a}_m \wedge \bigwedge_{n\le k<m} (\neg\mathsf{b}_k \wedge \mathsf{a}_k) \wedge \bigwedge_{m<k<2n} (\mathsf{b}_k \leftrightarrow \mathsf{a}_k)\big) \vee \bigwedge_{n\le m<2n} (\neg\mathsf{b}_m \wedge \mathsf{a}_m)\Big) \wedge \bigwedge_{k<n} (\mathsf{b}_k \leftrightarrow \mathsf{a}_k).$$

We add the following conjuncts to χ to ensure the tiling matching conditions:

$$\Box\Big(\mathsf{level}_{2n} \to \bigwedge_{t\in T} \big(\Diamond(\neg\mathsf{e} \wedge \mathsf{grid} \wedge \mathsf{t}) \to \Box(\neg\mathsf{e} \wedge \mathsf{succ}_x \to \bigvee_{(t,t)'\in H} \mathsf{t}')\big)\Big), \tag{19}$$

$$\Box\Big(\mathsf{level}_{2n} \to \bigwedge_{t\in T} \big(\Diamond(\neg\mathsf{e} \wedge \mathsf{grid} \wedge \mathsf{t}) \to \Box(\neg\mathsf{e} \wedge \mathsf{succ}_y \to \bigvee_{(t,t)'\in V} \mathsf{t}')\big)\Big). \tag{20}$$

Finally, we ensure that the initial condition $\bar{t}$ holds, that is $\tau(i,0) = t_i$ for $i < n$. To this end, we add to χ the conjuncts

$$\Box\big(\mathsf{level}_{2n} \wedge [\mathsf{a} = (i,0)] \to \Box(\neg\mathsf{e} \wedge \mathsf{grid} \to \mathsf{t}_i)\big), \quad \text{for } i < n. \tag{21}$$

It follows from the argument in Example 1.1 (i) that $(\varphi \to \psi) \in \mathsf{wK4}$. Below we show that (10) holds.

($\Rightarrow$) See Appendix A.

($\Leftarrow$) Suppose $\mathfrak{M}_\varphi, r_\varphi \sim^\varrho \mathfrak{M}_\psi, r_\psi$ with $\mathfrak{M}_\varphi, r_\varphi \models \varphi$ and $\mathfrak{M}_\psi, r_\psi \models \neg\psi$, for some σ-models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ based on frames for $\mathsf{wK4}$. Thus, r_φ is irreflexive, the cluster C of r_φ contains a point different from r_φ, and e is true everywhere in C and nowhere else in $\mathfrak{M}_\varphi$. As $\mathfrak{M}_\psi, r_\psi \models \neg\psi$ and by (11)–(13), r_ψ is the root of a full binary tree of depth $2n$ having its leaves $e_{i,j}$, $i,j < 2^n$, marked by level_{2n} and the corresponding formula $[\mathsf{a} = (i,j)]$. (Note that, as $\mathfrak{M}_\psi, r_\psi \models \Box\Box q \wedge \neg q$, none of $e_{i,j}$ is in $C(r_\psi)$.) By (14), for all $i,j < 2^n$, $\mathfrak{M}_\psi, e_{i,j} \models \mathsf{e}$, and so we must have $\mathfrak{M}_\varphi, x_{i,j} \sim_\varrho \mathfrak{M}_\psi, e_{i,j}$ for some $x_{i,j} \in C$. Thus, by (13) and (15),

(t1) for all $i,j < 2^n$, C has an R-successor w such that $\mathfrak{M}_\varphi, w \models [\mathsf{b} = (i,j)] \wedge \mathsf{t}$ for some $t \in T$.

We claim that

(t2) for all $i,j < 2^n$, if w, w' are R-successors of C with $\mathfrak{M}_\varphi, w \models [\mathsf{b} = (i,j)] \wedge \mathsf{t}$ and $\mathfrak{M}_\varphi, w' \models [\mathsf{b} = (i,j)] \wedge \mathsf{t}'$, then $t = t'$.

Indeed, by ϱ-bisimilarity, there exist R-successors u, u' of $e_{i,j}$ such that $\mathfrak{M}_\psi, u \models \neg\mathsf{e} \wedge [\mathsf{b} = (i,j)] \wedge \mathsf{t}$ and $\mathfrak{M}_\psi, u' \models \neg\mathsf{e} \wedge [\mathsf{b} = (i,j)] \wedge \mathsf{t}'$. By (16), $[\mathsf{a} = (i,j)]$ is true at both u and u', and so grid is true at both u and u' as well. Thus, $t = t'$ follows from (17) and (18).

Now we define a map τ by taking, for all $i, j < 2^n$, $\tau(i, j) = t$ iff C has an R-successor w with $\mathfrak{M}_\varphi, w \models [\mathsf{b} = (i, j)] \wedge \mathsf{t}$. By **(t1)** and **(t2)**, τ is well-defined. We claim that

(t3) for all $i, j < 2^n$, if $\tau(i, j) = t$ and $\tau(i \oplus 1, j) = t'$ then $(t, t') \in H$, and

(t4) for all $i, j < 2^n$, if $\tau(i, j) = t$ and $\tau(i, j \oplus 1) = t'$ then $(t, t') \in V$.

Indeed, for **(t3)**, by the definition of τ and ϱ-bisimilarity, there are R-successors u, u' of $e_{i,j}$ such that $\mathfrak{M}_\psi, u \models \neg\mathsf{e} \wedge [\mathsf{b} = (i, j)] \wedge \mathsf{t}$ and $\mathfrak{M}_\psi, u' \models \neg\mathsf{e} \wedge [\mathsf{b} = (i \oplus 1, j)] \wedge \mathsf{t}'$. By (16), $[\mathsf{a} = (i, j)]$ is true at both u and u'. So grid is true at u and succ_x is true at u', and so $(t, t') \in H$ follows from (19). Condition **(t4)** can be shown similarly, using (20) in place of (19).

It follows from **(t3)**, **(t4)** and (21) that τ is a solution for P and $\bar{t}$, completing the proof of (10). Now the theorem follows by Lemma 2.3. □

5 Discussion

Our investigation of the interpolant existence problem for weak K4 and the difference logic DL is part of a research programme that aims to understand Craig interpolants for logics not enjoying the CIP. It turns out that wK4 shares with standard modal logics with nominals [1], decidable fragments of first-order modal logics [14], and the guarded and two-variable fragment of first-order logic [11] that interpolant existence is still decidable but computationally harder than validity. In contrast, the difference logic DL shares with normal extensions of K4.3 [15] that interpolant existence has the same complexity as validity. Linear temporal logic LTL is another example of a logic without the CIP, for which interpolant existence is decidable [18]. For both wK4 and LTL, establishing tight complexity bounds remains an interesting open problem. Further open problems include the following. As our decision procedures are non-constructive, it would be of interest to develop algorithms that compute interpolants whenever they exist. Also, is it possible to establish general decidability results for interpolant existence for families of extensions of wK4? Another related question is to find out which additional logical connectives would repair the CIP for wK4; see [21,22] for elegant answers to such questions for modal logics with nominals and other fragments of first-order logic.

The results of this paper are also relevant to the *explicit definition existence problem* (EDEP) for $L \in \{\mathsf{DL}, \mathsf{wK4}\}$: given formulas φ, ψ and a signature ϱ, decide whether there is a ϱ-formula χ with $\varphi \to (\psi \leftrightarrow \chi) \in L$, called an *explicit ϱ-definition of ψ modulo φ in L*. The EDEP reduces trivially to validity for logics enjoying the projective Beth definability property [9], which is not the case for DL and wK4. In fact, one can prove in exactly the same way as for fragments of first-order modal logics [14] that, for $L \in \{\mathsf{DL}, \mathsf{wK4}\}$, the IEP and EDEP are polynomial-time reducible to each other. Thus, our results above also provide complexity bounds for the EDEP in DL and wK4.

Appendix

A Omitted proofs

Proof of Lemma 2.1. (a) It is easy to see that if $x, y \in C$ and $at^{\sigma}_{\mathfrak{M}}(x) = at^{\sigma}_{\mathfrak{M}}(y)$, then $t^{\sigma}_{\mathfrak{M}}(x) = t^{\sigma}_{\mathfrak{M}}(y)$. It follows by **(dif)** that $|C| \leq 2^{|\sigma|}$.

(b) By the assumption, the set $\Xi = \Gamma \cup \{\varphi \mid \Box\varphi \in t^{\sigma}_{\mathfrak{M}}(x)\}$ is finitely satisfiable, and so, by **(com)**, $\mathfrak{M}, y \models \Xi$, for some y, with xRy by **(ref)**.

(c) Take any $x \in C$ and any $\subseteq$-maximal R^s chain X' of R^s-successors of x in $\mathfrak{F}$ such that $t^{\varrho}_{\mathfrak{M}}(x') = t^{\varrho}_{\mathfrak{M}}(x)$ for all $x' \in X'$, which exists by Zorn's lemma. Let $X = \{x\} \cup X'$. If X has an R^s-maximal point y, then $C(y)$ is ϱ-maximal. Indeed, otherwise there is $z \in C(y)$ with an R^s-successor z' such that $t^{\varrho}_{\mathfrak{M}}(z) = t^{\varrho}_{\mathfrak{M}}(z')$. But then $\Diamond t^{\varrho}_{\mathfrak{M}}(x) \subseteq t^{\varrho}_{\mathfrak{M}}(z)$, and so, by (b), there is an R-successor z'' of z' with $t^{\varrho}_{\mathfrak{M}}(z'') = t^{\varrho}_{\mathfrak{M}}(x)$, contrary to the maximality of chain X. A similar argument shows that $t^{\varrho}_{\mathfrak{M}}(C) \subseteq t^{\varrho}_{\mathfrak{M}}\big(C(y)\big)$.

Now suppose X does not have an R^s-maximal element, that is, there is no $z \in X$ such that yR^sz for all $y \in X$. The set

$$\Xi = t^{\varrho}_{\mathfrak{M}}(x) \cup \{\varphi \mid \Box\varphi \in t^{\sigma}_{\mathfrak{M}}(y) \text{ for some } y \in X\}$$

is finitely satisfiable, and so, by **(com)**, there is z with $t^{\varrho}_{\mathfrak{M}}(x) = t^{\varrho}_{\mathfrak{M}}(z)$. By **(ref)**, we also have yRz for all $y \in X$, so $z \in X$, which is a contradiction. $\dashv$

Proof of Lemma 3.4. Suppose $\boldsymbol{w}R^{\dagger}\boldsymbol{w}'R^{\dagger}\boldsymbol{w}''$ and $\boldsymbol{w} \neq \boldsymbol{w}''$, where

$$\boldsymbol{w} = (\boldsymbol{t}, a, \tau, M), \quad \boldsymbol{w}' = (\boldsymbol{t}', a', \tau', M'), \quad \boldsymbol{w}'' = (\boldsymbol{t}'', a'', \tau'', M'').$$

We need to show that $\boldsymbol{w}R^{\dagger}\boldsymbol{w}''$. This is trivial if $\boldsymbol{w} = \boldsymbol{w}'$ or $\boldsymbol{w}' = \boldsymbol{w}''$, so we assume that $\boldsymbol{w} \neq \boldsymbol{w}'$ and $\boldsymbol{w}' \neq \boldsymbol{w}''$.

Suppose first that $M \neq M''$. We need to show that $M \twoheadrightarrow M''$ and $\tau \twoheadrightarrow \tau''$.

Case $M \neq M'$, $M' \neq M''$: By the definition of $R^{\dagger}$, we have $M \twoheadrightarrow M' \twoheadrightarrow M''$ and $\tau \twoheadrightarrow \tau' \twoheadrightarrow \tau''$, which gives $M \twoheadrightarrow M''$ and $\tau \twoheadrightarrow \tau''$ by (3).

Case $M \neq M'$, $M' = M''$: Then $M \twoheadrightarrow M'$ and $\tau \twoheadrightarrow \tau'$ because $\boldsymbol{w}R^{\dagger}\boldsymbol{w}'$, and so $M \twoheadrightarrow M''$. Then no matter whether $M' \twoheadrightarrow M''$ or $M' \not\twoheadrightarrow M''$, we have either $\tau' = \tau''$ or $\tau' \twoheadrightarrow \tau''$ as $\boldsymbol{w}'R^{\dagger}\boldsymbol{w}''$ and $\boldsymbol{w}' \neq \boldsymbol{w}''$. Thus, $\tau \twoheadrightarrow \tau''$ by (3).

Case $M = M'$, $M' \neq M''$: Then $M' \twoheadrightarrow M''$ and $\tau' \twoheadrightarrow \tau''$ because $\boldsymbol{w}'R^{\dagger}\boldsymbol{w}''$, and so $M \twoheadrightarrow M''$. Then no matter whether $M \twoheadrightarrow M'$ or $M \not\twoheadrightarrow M'$, either $\tau = \tau'$ or $\tau \twoheadrightarrow \tau'$, because $\boldsymbol{w}R^{\dagger}\boldsymbol{w}'$ and $\boldsymbol{w} \neq \boldsymbol{w}'$. Thus, $\tau \twoheadrightarrow \tau''$ by (3).

Now suppose $M = M''$. If $M \neq M'$, then $M' \neq M''$. By the definition of $R^{\dagger}$, we have $M \twoheadrightarrow M' \twoheadrightarrow M$ and $\tau \twoheadrightarrow \tau' \twoheadrightarrow \tau''$, which gives $M \twoheadrightarrow M$ and $\tau \twoheadrightarrow \tau''$ by (3). As $\boldsymbol{w} \neq \boldsymbol{w}''$, these imply $\boldsymbol{w}R^{\dagger}\boldsymbol{w}''$.

Finally, suppose $M = M' = M''$. Then two cases are possible.

Case $M \twoheadrightarrow M$: As $\boldsymbol{w} \neq \boldsymbol{w}''$, we need to show that $\tau = \tau''$ or $\tau \twoheadrightarrow \tau''$. As $\boldsymbol{w}R^{\dagger}\boldsymbol{w}'$ and $\boldsymbol{w} \neq \boldsymbol{w}'$, we have $\tau = \tau'$ or $\tau \twoheadrightarrow \tau'$. Similarly, as $\boldsymbol{w}'R^{\dagger}\boldsymbol{w}''$ and $\boldsymbol{w}' \neq \boldsymbol{w}''$, we have $\tau' = \tau''$ or $\tau' \twoheadrightarrow \tau''$, which yields the required.

Case $M \not\twoheadrightarrow M$: We need to show that $\tau = \tau''$ or ($\tau \twoheadrightarrow \tau''$ and $\tau(b) = \emptyset$ for all $b \in I_M$). As $\boldsymbol{w} R^\dagger \boldsymbol{w}'$ and $\boldsymbol{w} \neq \boldsymbol{w}'$, we have $\tau = \tau'$ or ($\tau \twoheadrightarrow \tau'$ and $\tau(b) = \emptyset$ for all $b \in I_M$). Similarly, as $\boldsymbol{w}' R^\dagger \boldsymbol{w}''$ and $\boldsymbol{w}' \neq \boldsymbol{w}''$, we have $\tau' = \tau''$ or ($\tau' \twoheadrightarrow \tau''$ and $\tau'(b) = \emptyset$ for all $b \in I_M$). By (3), it follows that $\tau = \tau''$ or ($\tau \twoheadrightarrow \tau''$ and $\tau(b) = \emptyset$ for all $b \in I_M$). As $\boldsymbol{w} \neq \boldsymbol{w}''$, these imply $\boldsymbol{w} R^\dagger \boldsymbol{w}''$.

This completes the proof of the lemma. $\dashv$

Proof of the ($\Rightarrow$) **direction of** (10)**.** Suppose τ is a solution for P and $\bar{t}$. We define σ-models $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ as follows. The underlying $\mathsf{wK4}$-frame of $\mathfrak{M}_\varphi$ consist of a two-element cluster C having an irreflexive point r_φ and a reflexive point x, and C has 2^{2n} irreflexive R-successors $w_{k,\ell}$, $k, \ell < 2^n$. The valuation in $\mathfrak{M}_\varphi$ is such that p holds at r_φ, e holds everywhere in C, and for $k, \ell < 2^n$, $\mathfrak{M}_\varphi, w_{k,\ell} \models [\mathsf{b} = (k,\ell)] \wedge \mathsf{t}$, where $t = \tau(k,\ell)$. The underlying $\mathsf{wK4}$-frame of $\mathfrak{M}_\psi$ consist of a full binary tree (T, R_T) of depth $2n$, with an irreflexive root r_ψ, all other nodes being reflexive, and having 2^{2n} leaves $e_{i,j}$, $i,j < 2^n$, and then each leaf $e_{i,j} \in T$ has 2^{2n} irreflexive R-successors $u^{k,\ell}_{i,j}$, $k, \ell < 2^n$. The valuation in $\mathfrak{M}_\psi$ is such that q holds everywhere apart from r_ψ, e holds everywhere in T, $\mathsf{level}_0, \dots, \mathsf{level}_{2n}$ and $\mathsf{a}_0, \dots, \mathsf{a}_{2n-1}$ 'mark' the nodes of the tree (T, R_T) in such a way that, for $i,j < 2^n$, $\mathfrak{M}_\psi, e_{i,j} \models \mathsf{level}_{2n} \wedge [\mathsf{a} = (i,j)]$, and for $i,j,k,\ell < 2^n$, $\mathfrak{M}_\psi, u^{k,\ell}_{i,j} \models [\mathsf{a} = (i,j)] \wedge [\mathsf{b} = (k,\ell)] \wedge \mathsf{t}$, where $t = \tau(k,\ell)$.

It is not hard to check that $\mathfrak{M}_\varphi, r_\varphi \models \varphi$ and $\mathfrak{M}_\psi, r_\psi \models \neg\psi$, and the relation

$$\boldsymbol{\beta} = (C \times T) \cup \left\{ \left(w_{k,\ell}, u^{k,\ell}_{i,j}\right) \mid i,j,k,\ell < 2^n \right\}$$

is a ϱ-bisimulation between $\mathfrak{M}_\varphi$ and $\mathfrak{M}_\psi$ with $r_\varphi \boldsymbol{\beta} r_\psi$. $\dashv$

References

[1] Artale, A., J. C. Jung, A. Mazzullo, A. Ozaki and F. Wolter, *Living without Beth and Craig: Definitions and interpolants in description and modal logics with nominals and role inclusions*, ACM Trans. Comput. Logic **24** (2023).

[2] Baader, F., I. Horrocks, C. Lutz and U. Sattler, "An Introduction to Description Logic," Cambridge University Press, 2017.

[3] Bezhanishvili, G., L. Esakia and D. Gabelaia, *Spectral and T_0-spaces in d-semantics*, in: N. Bezhanishvili, S. Löbner, K. Schwabe and L. Spada, editors, *Logic, Language, and Computation - 8th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2009, Bakuriani, Georgia, September 21-25, 2009. Revised Selected Papers*, Lecture Notes in Computer Science **6618** (2009), pp. 16–29.

[4] Bezhanishvili, G., S. Ghilardi and M. Jibladze, *An algebraic approach to subframe logics. modal case*, Notre Dame J. Formal Log. **52** (2011), pp. 187–202.

[5] Chagrov, A. V. and M. Zakharyaschev, "Modal Logic," Oxford logic guides **35**, Oxford University Press, 1997.

[6] de Rijke, M., *The modal logic of inequality*, J. Symb. Log. **57** (1992), pp. 566–584.

[7] Esakia, L., *Weak transitivity—a restitution*, Logical Investigations **8** (2001), pp. 244–255, (In Russian).

[8] Fine, K., *Logics containing* $\mathsf{K4}$*. part I*, J. Symb. Log. **39** (1974), pp. 31–42.

[9] Gabbay, D. M. and L. Maksimova, "Interpolation and Definability: Modal and Intuitionistic Logics," Oxford University Press, 2005.

[10] Goranko, V. and M. Otto, *Model theory of modal logic*, in: *Handbook of Modal Logic*, Elsevier, 2007 pp. 249–329.

[11] Jung, J. C. and F. Wolter, *Living without Beth and Craig: Definitions and interpolants in the guarded and two-variable fragments*, in: *Proceedings of the 36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021* (2021), pp. 1–14.

[12] Karpenko, A. and L. Maksimova, *Simple weakly transitive modal algebras*, Algebra and Logic **49** (2010), pp. 233–245.

[13] Kudinov, A. and I. Shapirovsky, *Filtrations for* wK4 *and its relatives*, CoRR **abs/2401.00457** (2024).
URL `https://doi.org/10.48550/arXiv.2401.00457`

[14] Kurucz, A., F. Wolter and M. Zakharyaschev, *Definitions and (uniform) interpolants in first-order modal logic*, in: P. Marquis, T. C. Son and G. Kern-Isberner, editors, *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR 2023, Rhodes, Greece, September 2-8, 2023*, 2023, pp. 417–428.

[15] Kurucz, A., F. Wolter and M. Zakharyaschev, *A non-uniform view of Craig interpolation in modal logics with linear frames*, CoRR **abs/2312.05929** (2023).
URL `https://doi.org/10.48550/arXiv.2312.05929`

[16] Marx, M., *Interpolation in modal logic*, in: *Proceedings of the 7th International Conference on Algebraic Methodology and Software Technology, AMAST 1998* (1998), pp. 154–163.

[17] McKinsey, J. and A. Tarski, *The algebra of topology*, Annals of Mathematics **45** (1944), pp. 141–191.

[18] Place, T. and M. Zeitoun, *Separating regular languages with first-order logic*, Log. Methods Comput. Sci. **12** (2016).

[19] Segerberg, K., *A note on the logic of elsewhere*, Theoria **46** (1980), pp. 183–187.

[20] Shapirovsky, I., *Satisfiability problems on sums of Kripke frames*, ACM Trans. Comput. Log. **23** (2022), pp. 15:1–15:25.

[21] ten Cate, B., “Model theory for extended modal languages,” Phd (2005).

[22] ten Cate, B. and J. Comer, *Craig interpolation for decidable first-order fragments*, CoRR **abs/2310.08689** (2023).
URL `https://doi.org/10.48550/arXiv.2310.08689`

[23] von Wright, G., *A modal logic of place*, in: E. Sosa, editor, *The philosophy of Nicolas Rescher*, Dordrecht, 1979 pp. 65–73.