TECHNISCHE UNIVERSITÄT DRESDEN

Technische Universitat Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Decidability of $\mathcal{ALC}^P(\mathcal{D})$ for concrete domains with the EHD-property

Claudia Carapelle   Anni-Yasmin Turhan

LTCS-Report 16-01

Postal Address:
Lehrstuhl fur Automatentheorie
Institut fur Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nothnitzer Str. 46
Dresden

Decidability of $\mathcal{ALC}^P(\mathcal{D})$ for concrete domains with the EHD-property

Claudia Carapelle   Anni-Yasmin Turhan *

April 2016

Abstract

Reasoning for Description logics with concrete domains and w.r.t. general TBoxes easily becomes undecidable. For particular, restricted concrete domains decidablity can be regained. We introduce a novel way to integrate a concrete domain $\mathcal{D}$ into the well-known description logic $\mathcal{ALC}$, we call the resulting logic $\mathcal{ALC}^P(\mathcal{D})$. We then identify sufficient conditions on $\mathcal{D}$ that guarantee decidability of the satisfiability problem, even in the presence of general TBoxes. In particular, we show decidability of $\mathcal{ALC}^P(\mathcal{D})$ for several domains over the integers, for which decidability was open. More generally, this result holds for all negation-closed concrete domains with the EHD-*property*, which stands for the existence of a homomorphism is definable. Such technique has recently been used to show decidability of CTL^* with local constraints over the integers.

Contents

*This work is supported by the German Research Foundation (DFG) within the Collaborative Research Center SFB 912 HAEC.

1 Introduction

Description Logics (DLs) are a collection of knowledge representation formalisms with well-founded semantics. They allow to characterize notions from an application domain by *concepts*. Each DL offers a set of concept constructors that allow to build complex concepts from atomic concepts and roles, which are binary relations. Concepts can be related to each other via sub-concept statements called *general concept inclusions* (GCIs). The (possibly cyclic) GCIs represent the terminological knowledge of the application domain and are collected in the *TBox*. DLs are employed nowadays in a range of application areas such as the biomedical field or the semantic web. DLs are the foundations of the web ontology language OWL 2 [18]. DLs are investigated for a range of reasoning problems. Several of the classical reasoning problems can be reduced to the satisfiability problem (if negation is present in the set of concept constructors), so we concentrate on this reasoning problem in this report.

DLs are an excellent tool to represent abstract knowledge, but practical applications often require concrete properties with values from a fixed domain, such as integers or strings, and to support built-in predicates. In [1], DLs were extended with *concrete domains*. These are relational structures $\mathcal{D} = (D, R_1, R_2, \dots)$, with a domain D and n-ary relations $R_1, R_2, \dots$. The resulting logic $\mathcal{ALC}(\mathcal{D})$ extends the standard DL $\mathcal{ALC}$ by the concept constructor *concrete domain restriction* and the DL can be parameterized with a concrete domain $\mathcal{D}$. Concrete domain restrictions can be used to build complex concepts based on concrete qualities of their instances such as the age, temperature or even measured values. For instance, $\mathcal{ALC}(\mathcal{D})$-concepts can characterize drivers of a motor vehicle by requiring that they are at least as old as the legal driving age of 18:

$$\mathsf{motor\text{-}vehicle\text{-}driver} \sqsubseteq \mathsf{Person} \sqcap \exists \mathsf{has} - \mathsf{age}. \geq_{18}$$

A concrete domain restriction can connect several abstract objects via *feature-paths*, i.e. paths of functional roles, and assert a predicate of arbitrary arity for concrete quantities of those objects. Concrete domains are incorporated in a weakend form in the ontology language OWL as data-types for which only unary predicates are admitted [18].

If definitorial and acyclic TBoxes are used, then reasoning for $\mathcal{ALC}$ with certain concrete domains, so called *admissible* ones, is decidable [1]. Reasoning becomes undecidable in the presence of general TBoxes [10, 14] for this kind of concrete domains. There have been several attempts to regain decidability for reasoning in $\mathcal{ALC}(\mathcal{D})$ with general TBoxes. Some approaches impose syntactic restrictions on the concept constructor – allowing only unary predicates [9] or admitting only feature-paths of length 1 [8] in concrete domain restrictions. These restrictions limit the modelling capabilities severely. Lutz and Miličić took a different approach and showed that if a concrete domain respects a criterion called *ω-admissibility*, then satisfiability for $\mathcal{ALC}(\mathcal{D})$ with general TBoxes is decidable [15]. The condition of ω-admissibility essentially allows to lift local satisfiability of finite (connected) concrete domain parts to global satisfiability by requiring compactness and that the concrete domain parts need to conform on the predicates asserted for the shared objects. This condition indicates decidability of DL reasoning for some concrete domains, for instance, the RCC8 relations and the Allen relations over the reals [15]. However, several interesting domains do not satisfy ω-admissibility, for instance domains based on non-dense domains, as the integers or the natural numbers. In [13], Lutz considers a concrete domain over the rational numbers, and proves that reasoning w.r.t. general TBoxes is decidable. Such domain can however not be used to reasonably represent some situations: certain concrete features, such as 'number of children', cannot possibly be fractions.

In this paper we devise a new criterion for concrete domains that guarantees decidability of the satisfiability problem in the presence of general TBoxes. This criterion holds also for some concrete domains that are known to be not ω-admissible, such as the integers. To this end we introduce the new DL $\mathcal{ALC}^P(\mathcal{D})$ that uses path constraints instead of concrete domain restrictions. Unlike the latter, which only allow feature-paths to connect an individual and a concrete value, path constraints can use the full expressiveness of role-paths. This enables to model for instance 'person who only has younger siblings', as an individual whose age is greater than that of all his siblings, where the sibling relation need not be functional.

We show decidability of the satisfiability problem of $\mathcal{ALC}^P(\mathcal{D})$ w.r.t. general TBoxes, if used with concrete domains that:

- are *negation-closed*, which requires that the complement of each (atomic) relation is effectively definable by a positive existential first-order formula, and
- have the *EHD-property*, which stands for 'the existence of a homomorphism is definable' and is a property of a relational structure $\mathcal{A}$, expressing the ability of a logic L to distinguish between those structures $\mathcal{B}$ which can be mapped to $\mathcal{A}$ by a homomorphism and those who cannot.

Our approach to show decidability of $\mathcal{ALC}^P(\mathcal{D})$ with concrete domains that fulfill the above conditions is an adaptation of the EHD-method, used in [6, 7] for CTL* and ECTL*. This, in turn, uses a recent decidability result by Bojańczyk and Toruńczyk for WMSO+B over infinite trees, an extension of weak monadic second order logic by the bounding quantifier B (see [3]). The idea for testing satisfiability of an $\mathcal{ALC}^P(\mathcal{D})$-concept C w.r.t. an $\mathcal{ALC}^P(\mathcal{D})$-TBox $\mathcal{T}$ is to proceed in two steps. First, an ordinary $\mathcal{ALC}$ interpretation is built that satisfies an abstracted version of C and $\mathcal{T}$, where each path constraint is replaced by a fresh concept name. Second, this interpretation is used to generate a so-called constraint graph, which is a structure in charge of remembering the contribution of the constraints that were abstracted away. We show that the fact that such a constraint graph allows a homomorphism to the concrete domain is enough to guarantee that the constraints are satisfied. In contrast to the mentioned CTL variants, $\mathcal{ALC}^P(\mathcal{D})$ is multi-modal and uses features, i.e. functional roles, which required some adaptation to apply the techniques from [6, 7].

By the newly established criterion for decidability of $\mathcal{ALC}^P(\mathcal{D})$ w.r.t. TBoxes, we confirm what the authors of Lutz and Miličić have conjectured: ω-admissibility is a sufficient, but not a necessary condition. We show, in fact, that reasoning with non ω-admissible concrete domains over the natural numbers and the integers w.r.t. general TBoxes is decidable. We also show that it is possible to add a feature to a concrete domain over the rational numbers presented in [11, 10], that allows to ask that a certain concrete value is an integer.

This report is structured as follows. In the next section we introduce basic notions on structures, the logic BMWB capturing Boolean combinations of WMSO+B and MSO. Section 3 gives preliminaries on DL and introduces the new DL $\mathcal{ALC}^P(\mathcal{D})$ with some of its properties and normal forms. In Section 4 we introduce the EHD-property and structures that enjoy such property, we explain the EHD method and show decidability of negation-closed concrete domains that have the EHD-property. Section 5 shows that we can add to a concrete domain a mechanism to mimic undefined concrete values.

2 Preliminary Notions

Structures

Definition 1. A (*relational*) *signature* $\sigma = \{R_1, R_2, \ldots\}$ is a countable (finite or infinite) set of relation symbols. Every relation symbol $R \in \sigma$ has an associated arity $\mathsf{ar}(R) \geq 1$.

A σ-*structure* is a tuple $\mathcal{A} = (A, R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \ldots)$, where A is a non-empty set (the universe of the structure) and for each $R \in \sigma$, $R^{\mathcal{A}} \subseteq A^{\mathsf{ar}(R)}$ is the interpretation of the relation symbol R in $\mathcal{A}$, that is an $\mathsf{ar}(R)$-ary relation over A.

Example 2. A simple example of a $\{=, <\}$-structure is $\mathcal{Z} = (\mathbb{Z}, =^{\mathcal{Z}}, <^{\mathcal{Z}})$, where $=^{\mathcal{Z}}$ and $<^{\mathcal{Z}}$ are defined as expected, namely as $\{(a, b) \in \mathbb{Z}^2 \mid a = b\}$ and $\{(a, b) \in \mathbb{Z}^2 \mid a < b\}$, respectively.

In case it does not create ambiguity, we often identify the relation $R^{\mathcal{A}}$ with the relation symbol R, and we specify a σ-structure as $(A, R_1, R_2, \ldots)$ where $\sigma = \{R_1, R_2, \ldots\}$. In the example above, then, we would simply write $(\mathbb{Z}, =, <)$.

Definition 3. For a σ-structure $\mathcal{A}$ and a τ-structure $\mathcal{B}$ such that $\tau \subseteq \sigma$, a *homomorphism* from $\mathcal{B}$ to $\mathcal{A}$ is a mapping $h : B \to A$ such that for all $R \in \tau$ and all tuples $(b_1, \ldots, b_{\mathsf{ar}(R)}) \in B^{\mathsf{ar}(R)}$ we have

$$(b_1, \ldots, b_{\mathsf{ar}(R)}) \in R^{\mathcal{B}} \Rightarrow (h(b_1), \ldots, h(b_{\mathsf{ar}(R)})) \in R^{\mathcal{A}} .$$

We write $\mathcal{B} \preceq \mathcal{A}$ if there is a homomorphism from $\mathcal{B}$ to $\mathcal{A}$. Note that we do not require this homomorphism to be injective.

MSO, WMSO+B and BMWB

We fix countably infinite sets $\mathsf{V_e}$ and $\mathsf{V_s}$ of element variables and set variables, respectively.

Monadic second-order logic (MSO) is the extension of first-order logic where also quantification over subsets of the underlying structure is allowed. Let us fix a signature σ.

Definition 4 (MSO Syntax)**.** MSO-formulas over the signature σ are defined by the following grammar, where $R \in \sigma$, $x, y, x_1, \ldots, x_{\mathsf{ar}(R)} \in \mathsf{V_e}$ and $X \in \mathsf{V_s}$:

$$\varphi ::= R(x_1, \ldots, x_{\mathsf{ar}(R)}) \mid x = y \mid x \in X \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \exists x\, \varphi \mid \exists X \varphi \ . \tag{1}$$

MSO-formulas are evaluated on σ-structures, where element and set variables range over elements and subsets of the domain, respectively.

Definition 5 (MSO Semantics)**.** If $\mathcal{A} = (A, R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \ldots)$ is a σ-structure, the semantics of MSO-formulas on $\mathcal{A}$ are defined inductively on the structure of the formula with the help of a valuation function $\nu : \mathsf{V_e} \cup \mathsf{V_s} \to A \cup 2^A$. We write $\nu[x \mapsto a]$ to denote the function which assigns a to x and is otherwise identical to ν.

- $(\mathcal{A}, \nu) \models R(x_1, \ldots, x_{\mathsf{ar}(R)})$ iff $(\nu(x_1), \ldots, \nu(x_{\mathsf{ar}(R)})) \in R^{\mathcal{A}}$;
- $(\mathcal{A}, \nu) \models x = y$ iff $\nu(x) = \nu(y)$;
- $(\mathcal{A}, \nu) \models x \in X$ iff $\nu(x) \in \nu(X)$;
- $(\mathcal{A}, \nu) \models \neg\varphi$ iff it is not the case that $(\mathcal{A}, \nu) \models \varphi$;
- $(\mathcal{A}, \nu) \models (\varphi_1 \wedge \varphi_2)$ iff $(\mathcal{A}, \nu) \models \varphi_1$ and $(\mathcal{A}, \nu) \models \varphi_2$;
- $(\mathcal{A}, \nu) \models \exists x\, \varphi$ iff there exists $b \in A$ such that $(\mathcal{A}, \nu[x \mapsto b]) \models \varphi$;
- $(\mathcal{A}, \nu) \models \exists X \varphi$ iff there exists $B \subseteq A$ such that $(\mathcal{A}, \nu[X \mapsto B]) \models \varphi$;

Remark 6. Introducing disjunction as

- $(\varphi_1 \vee \varphi_2) := \neg(\neg\varphi_1 \wedge \neg\varphi_2)$,

and universal quantification over element and set variables

- $\forall x \varphi := \neg\exists x \neg\varphi$,
- $\forall X \varphi := \neg\exists X \neg\varphi$,

we can associate to each formula φ its semantically equivalent negation normal form $\widehat{\varphi}$, where negation only appears in front of atomic formulas and relations.

Remark 7. Note that, if in a formula φ no variable occurs freely, i.e. all variables appear in the scope of a quantifier, the semantics of φ do not depend on the choice of ν. We can therefore simply write $\mathcal{A} \models \varphi$.

Weak monadic second-order logic (WMSO) has the same syntax as MSO (1), but second-order variables are interpreted as finite subsets of the underlying universe.

WMSO+B is the extension of WMSO by the *bounding quantifier* $\mathsf{B}X\varphi$ for $X \in \mathsf{V_s}$. The semantics of $\mathsf{B}X\varphi$ in the structure $\mathcal{A}$ with universe A are defined as follows: $(\mathcal{A}, \nu) \models \mathsf{B}X\varphi(X)$ if and only if there is a bound $b \in \mathbb{N}$ such that whenever $(\mathcal{A}, \nu) \models \varphi(B)$ for some finite subset $B \subseteq A$, then $|B| \leq b$. The dual quantifier is denoted by U. It is called the *unbounding quantifier* and $\mathsf{U}X\varphi = \neg\mathsf{B}X\varphi$ expresses that there are arbitrarily large finite sets that satisfy φ.

Finally, let BMWB denote the set of all Boolean combinations of MSO-formulas and (WMSO+B)-formulas.

MSO and WMSO+B can express many interesting properties of relational structures.

Example 8. Given a graph $\mathcal{G} = (V, E)$, WMSO can express reachability in $\mathcal{G}$. We define the WMSO-formula $\mathsf{reach}_Z(x_1, x_2)$ to be

$$x_1 \in Z \wedge \forall Y \subseteq Z\big[\big(x_1 \in Y \wedge \forall y \forall z(y \in Y \wedge z \in Z \wedge E(y, z)) \rightarrow z \in Y\big) \rightarrow x_2 \in Y\big] \ .$$

It is easy to see that for every finite subset $B \subseteq A$, we have $\mathcal{A} \models \mathsf{reach}_B(a, b)$ if and only if $(a, b) \in (E^* \cap B^2)$, i.e., b is reachable from a in the subgraph $\mathcal{G}_{|B}$. Note that reach_Z is the standard MSO-formula for reachability but restricted to the subgraph induced by Z. If we define $\mathsf{reach}(x, y) := \exists Z \mathsf{reach}_Z(x, y)$, the semantics of reach seen as an MSO-formula or a WMSO-formula are the same because b is reachable from a in the graph $\mathcal{G}$ if and only if it is in some finite subgraph of $\mathcal{G}$.

BMWB-satisfiability is decidable over n-trees

Consider the signature $S2S = \{S_0, S_1, p_1, p_2, \dots\}$ formed by two binary relation symbols and countably many unary predicates. A *(labeled) binary tree* with left and right successor is then a structure $\mathcal{T} = (\{0,1\}^*, S_0, S_1, p_1^{\mathcal{T}}, p_2^{\mathcal{T}}, \dots)$[1] where $S_i = \{(x, xi) \mid x \in \{0,1\}^*\}$ for $i = 0, 1$, and $p_j^{\mathcal{T}} \subseteq \{0,1\}^*$ are arbitrary interpretations for each p_j.

In [3] Bojańczyk and Toruńczyk proved that WMSO+B over the signature $S2S$ has a decidable satisfiability problem over binary trees.

Theorem 9 (cf. [3]). *One can decide whether for a given formula $\varphi \in$ WMSO+B over the signature $S2S$ there is a binary tree $\mathcal{T}$ such that $\mathcal{T} \models \varphi$.*

Let $[1, n]$ indicate the set $\{1, 2, \dots, n\}$, and $S1S$ be the signature $\{S, p_1, p_2, \dots\}$. We call *(labeled) n-tree* a structure $\mathcal{T}_n = ([1, n]^*, S, p_1^{\mathcal{T}}, p_2^{\mathcal{T}}, \dots)$ where $S = \{(x, xi) \mid x \in [1, n]^*$ and $i \in [1, n]\}$ is the successor relation. Using the above Theorem 9, together with some properties of MSO and WMSO+B, it is proven in [7] that BMWB over the signature $S1S$ has a decidable satisfiability problem over n-trees:

Theorem 10 (cf. [7]). *One can decide whether for a given formula $\varphi \in$ BMWB over $S1S$ there exists an n-tree $\mathcal{T}_n$ such that $\mathcal{T}_n \models \varphi$.*

[1] We omit the superscript $\mathcal{T}$ for the interpretation of the relations S_1 and S_2 because they are always interpreted in the same way on a binary tree.

3 The Description Logic $\mathcal{ALC}^P(\mathcal{D})$

Let us fix for the rest of this section a countably infinite set of register variables Reg, a relational signature σ, and an arbitrary σ-structure $\mathcal{D} = (D, R_1, R_2, \ldots)$, called the *concrete domain*.

Definition 11. We define a *constraint* $c(x_1, \ldots, x_k)$ of arity k over $\mathcal{D}$ as a Boolean combination of *atomic constraints* $R(x_{i_1}, \ldots, x_{i_{\mathsf{ar}(R)}})$, where $R \in \sigma$ and $i_j \in \{1, \ldots, k\}$. We write $\mathcal{D} \models c(a_1, \ldots, a_k)$ if the constraint is satisfied in $\mathcal{D}$ by the assignment $x_i \mapsto a_i$.

Example 12. Consider as concrete domain $\mathcal{Z} = (\mathbb{Z}, <, =)$, the relational structure introduced in Example 2. We use infix notation for the relations to improve readability. Then $c(x, y, z) = [(x < y \vee x = y) \wedge \neg\, y < z]$ is a constraint of arity 3 over $\mathcal{Z}$, and $\mathcal{Z} \models c(0, 1, 0)$.

Let us fix two countably infinite sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of *concept names* and *role names*, respectively. Let then $\mathsf{N_F} \subseteq \mathsf{N_R}$ be the set of *features*, a special kind of roles that are interpreted as partial functions. We call a finite sequence $P = r_1 \cdots r_n$ of role names a *role-path* of length n.

Definition 13. We recursively define $\mathcal{ALC}^P(\mathcal{D})$*-concepts* as follows

$$C := A \mid \neg C \mid (C \sqcap C) \mid \exists r.C \mid \exists P.c(S^{i_1} x_1, \ldots, S^{i_k} x_k)$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, P is a role-path of length $n \geq 0$, c is a constraint of arity k, $x_1, \ldots, x_k \in \mathsf{Reg}$ and $i_1, \ldots, i_k \leq n$. We call $\exists P.c(S^{i_1} x_1, \ldots, S^{i_k} x_k)$ a *path constraint*. The symbol S appearing in the path constraints stands for successor, as the term $S^i x$ points at the register variable x in the i-th position of the path P.

Definition 14. A *general concept inclusion* (GCI) is an expression of the form $C \sqsubseteq D$, where C and D are concepts. A *TBox* is a finite set of GCIs.

Definition 15. A $\mathcal{D}$*-interpretation* $\mathcal{I}$ is a tuple $(\Delta, \cdot^{\mathcal{I}}, \gamma)$, where Δ is a set called *the domain*, $\cdot^{\mathcal{I}}$ is the *interpretation function*, and $\gamma : \Delta \times \mathsf{Reg} \to D$ is the *valuation function*, assigning a value from the concrete domain to each register variable in each element of the interpretation domain. The interpretation function maps

- each concept name $A \in \mathsf{N_C}$ to some $A^{\mathcal{I}} \subseteq \Delta$,
- each role name $r \in \mathsf{N_R}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta \times \Delta$,
- if $f \in \mathsf{N_F}$ the binary relation $f^{\mathcal{I}}$ has to be functional, i.e. for all $a, b, c \in \Delta$, $(a, b), (a, c) \in f^{\mathcal{I}}$ implies $b = c$.

It is then extended to role-paths and arbitrary concepts as follows:

$$\begin{aligned}
(r_1 \cdots r_n)^{\mathcal{I}} &:= \{(v_0, \ldots, v_n) \in \Delta^{n+1} \mid (v_{i-1}, v_i) \in r_i^{\mathcal{I}} \text{ for } i = 1, \ldots, n\} \\
(\neg C)^{\mathcal{I}} &:= \Delta \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &:= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(\exists r.C)^{\mathcal{I}} &:= \{v \in \Delta \mid \exists w \in \Delta \text{ with } (v, w) \in r^{\mathcal{I}} \text{ and } w \in C^{\mathcal{I}}\}
\end{aligned}$$

and, if P has length n, we define $(\exists P.c(S^{i_1} x_1, \ldots, S^{i_k} x_k))^{\mathcal{I}}$ as

$$\{v \in \Delta \mid \exists (v_0, \ldots, v_n) \in P^{\mathcal{I}} \text{ s.t. } v_0 = v, \text{ and } \mathcal{D} \models c(\gamma(v_{i_1}, x_1), \ldots, \gamma(v_{i_k}, x_k))\}\,.$$

So the fact that an element $v \in \Delta$ belongs to the interpretation of $\exists P.c(S^{i_1} x_1, \ldots, S^{i_k} x_k)$ means that there exists an instance of the path $P^{\mathcal{I}}$ starting in v, namely some $(v_0, v_1, \ldots, v_n) \in P^{\mathcal{I}}$ with $v_0 = v$, such that the assignment $y_j \mapsto \gamma(v_{j_i}, x_j)$ satisfies the constraint $c(y_1, \ldots, y_k)$. A term S^i inside the constraint is used to point at the i-th element of the path $P^{\mathcal{I}}$. Note that the requirement that $i_1, \ldots, i_k \leq n$ ensures that such element is well-defined.

Note also that an atomic constraint $R(S^{i_1}x_1, \ldots, S^{i_k}x_k)$ is *local* in the sense that it involves only nodes in a fixed *neighborhood* of the position at which they are evaluated. We call $d := \max\{i_1, \ldots, i_k\}$ the *depth* of R. By extension, the depth of a constraint c is the maximum depth of all the atomic constraints which appear in c.

If we call $\mathsf{Reg}_{C,\mathcal{T}}$ the set of register variables that occur in C and $\mathcal{T}$, it is clear that the relevance of the valuation function γ is limited to the domain $(\Delta \times \mathsf{Reg}_{C,\mathcal{T}})$.

Definition 16. A $\mathcal{D}$-interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ ($\mathcal{I} \models \mathcal{T}$) if and only if every GCI $C \sqsubseteq D \in \mathcal{T}$ is satisfied, that is, if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Given a concept C and a TBox $\mathcal{T}$, we say C is *satisfiable with respect to* $\mathcal{T}$ if and only if there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}} \neq \emptyset$. We write $\mathcal{I} \models_{\mathcal{T}} C$.

Remark 17. We define some additional operators:

- $C \sqcup D := \neg(\neg C \sqcap \neg D)$,
- $\forall r.C := \neg\exists r.\neg C$,
- $\forall P.c := \neg\exists P.\neg c$,
- $\exists P.C := \exists r_1.\exists r_2.\cdots\exists r_n.C$, where $P = r_1 \cdots r_n$

and special concepts:

- $\top := A \sqcup \neg A$,
- $\bot := A \sqcap \neg A$,

Using this extended set of operators and DeMorgan's laws we can, given an $\mathcal{ALC}^P(\mathcal{D})$-concept C, obtain an equivalent concept in negation normal form $\mathsf{nnf}(C)$, where negation only appears before concept names or relations from the concrete domain $R(x_1, \ldots, x_k)$ with $R \in \sigma$.

A *TBox-concept* of the TBox $\mathcal{T}$ is defined as $C_{\mathcal{T}} = \bigsqcap_{C \sqsubseteq D \in \mathcal{T}} (\neg C \sqcup D)$.

Note that it is equivalent to ask that an interpretation $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ satisfies all GCIs $C \sqsubseteq D$ in $\mathcal{T}$, and to ask that the $C_{\mathcal{T}}$ is globally satisfied, i.e. $(C_{\mathcal{T}})^{\mathcal{I}} = \Delta$. Vice-versa, any globally satisfied concept C can be seen as the GCI $\top \sqsubseteq C$. For technical reasons, it is convenient for us to adopt this view, and from now on we will always assume that a TBox consists of a single concept $C_{\mathcal{T}}$ that needs to be globally satisfied. We say a TBox $\mathcal{T}$ is in negation normal form if so is $C_{\mathcal{T}}$.

Example 18. Take again $\mathcal{Z} = (\mathbb{Z}, <, =)$ as concrete domain and consider the following TBox: $\mathcal{T} = \{\exists \mathsf{neighbor}.(\textit{green grass} < S\textit{green grass}),\ \neg\mathsf{GreenThumb} \sqcup (\textit{alive plants} = \textit{plants})\}$[2]. Here we consider three register variables: *green grass* measures the degree of *greenness* of an individual's lawn, while *plants* and *alive plants* count the number of plants (total or alive) of an individual. In any model of $\mathcal{T}$, every individual has a neighbor whose grass is greener, and individuals with a green thumb keep all their plants alive.

Remark 19. In example 18, there cannot exist a model for $\mathcal{T}$ with a finite underlying domain, as the degree of greenness of neighboring lawns is strictly increasing. This is never the case for ordinary $\mathcal{ALC}$, which enjoys the finite model property.

[2]Here the absence of a path quantifier before (*alive plants* = *plants*) means that we are referring to a "path of length zero".

Remark 20. In the literature on description logics with concrete domains (for instance in [1, 15]) one finds constraints of the kind $\exists R(P_1 x_1, \dots, P_k x_k)$, where R is a relation from the concrete domain and each P_i is a path composed of features only. The constraint is satisfied by an element d if there exist k elements, $d_1 \dots d_k$, reachable from d via the feature-paths $P_1 \dots P_k$, such that the tuple $(\gamma(d_1, x_1), \dots, \gamma(d_k, x_k))$ belongs to the relation R in the concrete domain. Nonetheless, in many interesting cases this kind of constraint can be replaced with path constraints by introducing some additional register variables. For example, the constraint $\exists(P_1 x_1 < P_2 x_2)$ can be expressed as $\exists P_1.(S^{|P_2|} x_1 < z) \sqcap \exists P_2.(z \leq S^{|P_2|} x_2)$, where z is a fresh register variable. Likewise, the constraint $\forall(P_1 x_1 < P_2 x_2)$ can be replaced by the expression $\neg(\exists P_1.\top \sqcap \exists P_2.\top) \sqcup (\exists P_1.(S^{|P_1|} x_1 < z) \sqcap \exists P_2.(z \leq S^{|P_2|} x_2))$.[3] On the other hand, our constraints can use role-paths of arbitrary length, which—to the best of our knowledge—is not allowed in the previously existing literature, where they are limited in length or disallowed completely in favor of feature-paths. Therefore, although these two kinds of constraints are generally incomparable in expressiveness, they are strictly more expressive on interesting concrete domains.

3.1 $\mathcal{ALC}^P(\mathcal{D})$ has the tree-model property

Definition 21. Let $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}}, \gamma)$ be a $\mathcal{D}$-interpretation and define $\rightarrow := \bigcup_{r \in \mathsf{N_R}} r^{\mathcal{I}}$. We say $\mathcal{I}$ is a *tree-shaped $\mathcal{D}$-interpretation* if and only if $(\Delta, \rightarrow)$ is a tree, that is:

- $\Delta \subseteq \Sigma^*$ is (isomorphic to) a prefix-closed set of strings over some alphabet Σ, and
- for all $u, v \in \Delta$, $u \rightarrow v$ if and only if $v = ua$ for some $a \in \Sigma$.

We call $\mathcal{I}$ an *n-tree $\mathcal{D}$-interpretation* if $\Delta = [1, n]^*$ for some $n \in \mathbb{N}$, where $[1, n]$ denotes the closed interval $\{1, \dots, n\}$.

A logic has the *tree model property*, if for every concept C and every TBox $\mathcal{T}$, C is satisfiable w.r.t. $\mathcal{T}$ if and only if there exists a tree-shaped $\mathcal{D}$-interpretation $\mathcal{J}$ such that $\mathcal{J} \models_{\mathcal{T}} C$, in particular, the root ε of $\mathcal{J}$ is such that $\varepsilon \in C^{\mathcal{J}}$.

We denote by $\mathsf{Sub}(\mathcal{T}, C)$ the set of all concepts which appear in a TBox $\mathcal{T}$ and in a concept C.

Theorem 22. *$\mathcal{ALC}^P(\mathcal{D})$ has the tree-model property.*

Proof. We show that given a $\mathcal{D}$-interpretation $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}}, \gamma)$ such that $\mathcal{I} \models_{\mathcal{T}} C$, we can build a tree-shaped $\mathcal{D}$-interpretation $\mathcal{J} = (\Delta', \cdot^{\mathcal{J}}, \gamma')$ such that $\mathcal{J} \models_{\mathcal{T}} C$. This is done by the process commonly known as *unravelling*, made non-standard only by the fact that we have to deal with the concrete values assigned by the valuation function.

So let $v \in \Delta$ be such that $v^{\mathcal{I}} \in C^{\mathcal{I}}$, such element exists because C is satisfiable. First of all we define Δ', the domain of $\mathcal{J}$, as the set of all paths in Δ originating in v, namely as

$$\{v_0 r_1 v_1 r_2 v_2 \cdots r_n v_n \mid v_0 = v \text{ and } (v_{i-1}, v_i) \in r_i^{\mathcal{I}} \text{ for all } i = 1, \dots, n\}.$$

We successively define $\cdot^{\mathcal{J}}$ according to the *last node semantics*:

- $A^{\mathcal{J}} := \{v_0 \cdots v_n \in \Delta' \mid v_n \in A^{\mathcal{I}}\}$ for all concept names $A \in \mathsf{N_C}$,
- $r^{\mathcal{J}} := \{(v_0 \cdots v_n, v_0 \cdots v_n r v_{n+1}) \mid (v_n, v_{n+1}) \in r^{\mathcal{I}}\}$ for all role names $r \in \mathsf{N_R}$,

and extend it to arbitrary concepts in the usual way.

[3]Such translations must be applied after the concepts are converted to strong negation normal form (see Sec. 3.2) because they preserve satisfiability but are not necessarily closed under negation.

Finally the valuation function γ' is defined for all $v_0 \cdots v_n \in \Delta'$ and for all $x \in \mathsf{Reg}$ as $\gamma'(v_0 \cdots v_n, x) := \gamma(v_n, x)$.

Now, if we consider $\rightarrow = \bigcup_{r \in \mathsf{N_R}} r^{\mathcal{J}}$, it is easy to see that $(\Delta', \rightarrow)$ is a tree with root v. To prove that $\mathcal{J} \models_{\mathcal{T}} C$, we show the following (HP): For all $v_0 \cdots v_n \in \Delta'$ and for all concepts $D \in \mathsf{Sub}(\mathcal{T}, C)$, $v_0 \cdots v_n \in D^{\mathcal{J}} \Leftrightarrow v_n \in D^{\mathcal{I}}$.

Using (HP) and the fact that $\mathcal{I} \models \mathcal{T}$ we obtain that $v_0 \cdots v_n \in (C_{\mathcal{T}})^{\mathcal{J}}$ for all $v_0 \cdots v_n \in \Delta'$, where $C_{\mathcal{T}}$ is the TBox concept for $\mathcal{T}$, thus showing $\mathcal{J} \models \mathcal{T}$. Furthermore the one-node path v belongs to $C^{\mathcal{J}}$, which implies that $\mathcal{J} \models_{\mathcal{T}} C$, as wanted.

Let us now prove (HP) by structural induction on a concept D. Let $p = v_0 \cdots v_n$ be a path originating in v:

- If D is a concept name, then (HP) is satisfied by definition of $D^{\mathcal{J}}$.
- If $D = \neg E$ for some concept E, then $p \in D^{\mathcal{J}}$ if and only if $p \notin E^{\mathcal{J}}$. By induction hypothesis v_n does not belong to $E^{\mathcal{I}}$, that is, $v_n \in (\neg E)^{\mathcal{I}}$.
- Let $D = E \sqcap F$. If $p \in D^{\mathcal{J}}$, by definition p belongs to both $E^{\mathcal{J}}$ and $F^{\mathcal{J}}$. By induction hypothesis this holds if and only if $v_n \in E^{\mathcal{I}} \cap F^{\mathcal{I}}$, that is $v_n \in D^{\mathcal{I}}$.
- Let $D = \exists r.E$. Then $p \in D^{\mathcal{J}}$ if and only if there exists $p' \in \Delta'$ such that $(p, p') \in r^{\mathcal{J}}$ and such that $p' \in E^{\mathcal{J}}$. By definition of $r^{\mathcal{J}}$, this means that $p' = v_0 \cdots v_n r v'$ for some $v' \in \Delta$ such that $(v_n, v') \in r^{\mathcal{I}}$. Also, by induction hypothesis $p' \in E^{\mathcal{J}}$ holds if and only if $v' \in E^{\mathcal{I}}$. This yields $p \in D^{\mathcal{J}}$ if and only if $v_n \in D^{\mathcal{I}}$, as wanted.
- Let now $D = \exists P.c(S^{i_1}x_1, \ldots, S^{i_k}x_k)$, with $P = r_1 \cdots r_m$. We know that $v_0 \cdots v_n \in D^{\mathcal{J}}$ if and only if there exists $(p_0, p_1, \ldots, p_m) \in P^{\mathcal{J}}$ with $p_0 = v_0 \cdots v_n$ such that $\mathcal{D} \models c(\gamma(p_{i_1}, x_1), \ldots, \gamma(p_{i_k}, x_k))$. Now, notice that $(p_0, \ldots, p_m)$ belongs to $P^{\mathcal{J}}$ if and only if $(p_{i-1}, p_i) \in r_i^{\mathcal{J}}$ for $i = 1, \ldots, m$, which is true if and only if $p_i = p_{i-1} r_i w_i$ for some $w_1, \ldots, w_m \in \Delta$. But this is equivalent to $(w_0, w_1, \ldots, w_m) \in P^{\mathcal{I}}$ where $w_0 = v_n$. Now, by definition of γ', $\gamma(p_{i_j}, x_j) = \gamma(w_{i_j}, x_j)$ for all $j = 0, \ldots, k$, and therefore $v_n \in D^{\mathcal{I}}$.

□

We prove now, that $\mathcal{ALC}^P(\mathcal{D})$ actually has a stronger form of the tree-model property, where the branching degree is bounded. We denote by $\#_E(\mathcal{T}, C)$ the number of existentially quantified subconcepts that occur in $\mathsf{Sub}(\mathcal{T}, C)$. We can prove the following:

Lemma 23. *Consider a concept C and a TBox $\mathcal{T}$, both in negation normal form, let d be the maximum depth of an existential path constraint occurring in C and $\mathcal{T}$, and let $e = \#_E(\mathcal{T}, C)$. Given a tree-shaped $\mathcal{D}$-interpretation $\mathcal{I} = (\Delta_{\mathcal{I}}, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ such that $\mathcal{I} \models_{\mathcal{T}} C$, we can obtain an n-tree $\mathcal{D}$-interpretation $\mathcal{H} = (\Delta_{\mathcal{H}}, \cdot^{\mathcal{H}}, \gamma_{\mathcal{H}})$ such that $\mathcal{H} \models_{\mathcal{T}} C$, where $n = d \cdot e$.*

Proof. The idea is the following: first we prove that, given the tree-model $\mathcal{I}$, we can build a tree-shaped interpretation $\mathcal{J}$ for C w.r.t. $\mathcal{T}$ which is an infinitely branching infinite tree. Successively we prune such model to obtain a new one where each node has exactly $n = d \cdot e$ many successors.

Let us look in detail at the first step. Suppose the role name $s \in \mathsf{N_R} \setminus \mathsf{N_F}$ does not appear in C nor in $\mathcal{T}$. The idea is to introduce new s-successors to the elements of $\Delta_{\mathcal{I}}$, and then attach copies of $\mathcal{I}$ to it, until we obtain an infinitely branching infinite tree: $\mathcal{J} = (\Delta_{\mathcal{J}}, \cdot^{\mathcal{J}}, \gamma_{\mathcal{J}})$. We define $\Delta_{\mathcal{J}} \subseteq (\Delta_{\mathcal{I}} \times \mathbb{N})^*$ recursively, and for each $v \in \Delta_{\mathcal{I}}$ we write v^i instead of (v, i) to increase readability.

First of all set $(\Delta_{\mathcal{I}} \times \{0\}) \subseteq \Delta_{\mathcal{J}}$. We also set, for all $v, w \in \Delta_{\mathcal{I}}$ and for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$, $v^0 \in A^{\mathcal{J}}$ if and only if $v \in A^{\mathcal{I}}$, and $(v^0, w^0) \in r^{\mathcal{J}}$ if and only if $(v, w) \in r^{\mathcal{I}}$. Furthermore we define $\gamma_{\mathcal{J}}(v^0, x) = \gamma_{\mathcal{I}}(v, x)$, for all $x \in \mathsf{Reg}$.

Now let $v \in \Delta_{\mathcal{J}}$ be some node with only finitely many successors. Then for each $w \in \Delta_{\mathcal{I}}$ we add infinitely many elements $vw^1, vw^2, vw^3 \dots$ to $\Delta_{\mathcal{J}}$. If ε is the minimal element of $\Delta_{\mathcal{I}}$, we add $(v, v\varepsilon^i)$ to $s^{\mathcal{J}}$ for all $i = 1, 2, 3 \dots$. Furthermore, for all $r \in \mathsf{N_R}$, if $(w, z) \in r^{\mathcal{I}}$, we add (vw^i, vz^i) to $r^{\mathcal{J}}$ for all $i = 1, 2, 3 \dots$. For all concept names $A \in \mathsf{N_C}$ and for all $w \in \Delta_{\mathcal{I}}$ such that $w \in A^{\mathcal{I}}$, we add vw^i to $A^{\mathcal{J}}$. Finally, for all $w \in \Delta_{\mathcal{I}}$ and $x \in \mathsf{Reg}$, we set $\gamma_{\mathcal{J}}(vw^i, x) = \gamma_{\mathcal{I}}(w, x)$.

We repeat this procedure until an infinitely branching infinite tree is obtained. Note that all elements v of Δ' are either of the form $v = w^0$ or $v = v'w^i$ for some $v' \in \Delta'$ and some $w \in \Delta$. We say then that v is *a copy of* w. Note also that this procedure respects the functional requirements for all $f \in \mathsf{N_F}$.

We now show by induction the following (HP): For all $v \in \Delta'$ and $w \in \Delta$ such that v is a copy of w and for all concepts $D \in \mathsf{Sub}(\mathcal{T}, C)$, $v \in D^{\mathcal{J}}$ if and only if $w \in D^{\mathcal{I}}$. This implies that $\mathcal{J} \models_{\mathcal{T}} C$, as wanted.

Let us now prove (HP) by structural induction on $D \in \mathsf{Sub}(\mathcal{T}, C)$. Let $v \in \Delta'$ be a copy of $w \in \Delta$:

- If D is a concept name, then (HP) is satisfied by definition of $D^{\mathcal{J}}$.
- If $D = \neg E$ for some concept E, then $v \in D^{\mathcal{J}}$ if and only if $v \notin E^{\mathcal{J}}$, if and only if (by induction hypothesis) w does not belong to $E^{\mathcal{I}}$, that is, $w \in (\neg E)^{\mathcal{I}}$.
- Let $D = E \sqcap F$. If $v \in D^{\mathcal{J}}$, by definition v belongs to both $E^{\mathcal{J}}$ and $D^{\mathcal{J}}$. By induction hypothesis this holds if and only if w belongs to $E^{\mathcal{I}}$ and $F^{\mathcal{I}}$, that is $w \in D^{\mathcal{I}}$.
- Let $D = \exists r.E$. Note that $r \neq s$, because $D \in \mathsf{Sub}(\mathcal{T}, C)$ and we chose s specifically among those role names which do neither appear in C nor in $\mathcal{T}$. Now, $v \in D^{\mathcal{J}}$ if and only if there exists v_1 such that $(v, v_1) \in r^{\mathcal{J}}$ and $v_1 \in E^{\mathcal{J}}$. By construction of $r^{\mathcal{J}}$, v_1 must be a copy of some $w_1 \in \Delta$ such that $(w, w_1) \in r^{\mathcal{I}}$. Furthermore, by induction hypothesis $v_1 \in E^{\mathcal{J}}$ if and only if $w_1 \in E^{\mathcal{I}}$ which yields $w \in D^{\mathcal{I}}$, as wanted. For the vice-versa, suppose now that $w \in D^{\mathcal{I}}$. Then there exists $w_1 \in \Delta$ such that $(w, w_1) \in r^{\mathcal{I}}$ and $w_1 \in E^{\mathcal{I}}$. Deducing that $v \in D^{\mathcal{J}}$ is as simple as proving that there exists an element $v_1 \in \Delta'$ such that v_1 is a copy of w_1 and such that $(v, v_1) \in r^{\mathcal{J}}$. But this is a straightforward consequence of the construction we used: If $v = w^0$, then w_1^0 also belongs to Δ' and it is such that $(w^0, w_1^0) \in r^{\mathcal{J}}$. If $v = v'w^i$ for some $v' \in \Delta'$, then $v'w_1^i$ belongs to Δ' and respects our requirements.
- Let now $D = \exists P.c(S^{i_1}x_1, \dots, S^{i_k}x_k)$, with $P = r_1 \cdots r_m$. If $v \in D^{\mathcal{J}}$, then there exist $v_0, v_1, \dots, v_m \in \Delta'$ such that $v_0 = v$ and $(v_{j-1}, v_j) \in r_j^{\mathcal{J}}$ for $j = 1, \dots, m$. Furthermore, $\mathcal{D} \models c(\gamma_{\mathcal{J}}(v_{i_1}, x_1), \dots, \gamma_{\mathcal{J}}(v_{i_k}, x_k))$. Again, notice that none of the $r_1, \dots, r_m$ which appear in the path P can be the role s. Suppose $v = w^0$, by construction we can find elements $w_0, \dots, w_m$ such that $v_j = w_j^0$ for $j = 0 \dots m$ (in particular $w_0 = w$). By construction we find that $(w_{j-1}, w_j) \in r^{\mathcal{I}}$ for $j = 1 \dots m$. According to how we defined $\gamma_{\mathcal{J}}$, we know that $\gamma_{\mathcal{I}}(w, x) = \gamma_{\mathcal{J}}(w^0, x)$ and therefore we deduce that $\mathcal{D} \models c(\gamma_{\mathcal{I}}(w_{i_1}, x_1), \dots, \gamma_{\mathcal{I}}(w_{i_k}, x_k))$. Then $w \in E^{\mathcal{I}}$, as wanted. The case where $v = v'w^i$ for some $v' \in \Delta'$ and some $i \geq 1$ is treated analogously. Also the direction $w \in D^{\mathcal{I}} \Rightarrow v \in D^{\mathcal{J}}$ is proved using the same tools.

The second step consists in pruning $\mathcal{J}$ until we are left with an n-tree interpretation. To this end, we select only those elements of $\Delta_{\mathcal{J}}$ which are in some sense *necessary*. Let us see how.

As a base step we take $v_1 = \varepsilon$, the root of $\Delta_{\mathcal{J}}$, and $\Delta^0_{\mathcal{H}} = \{\varepsilon\}$. We define $\Delta^1_{\mathcal{H}}$ by applying the procedure shown in Figure 1, instantiated for $i = 1$.

For the i-th step we first choose a node v_i from $\Delta^{i-1}_{\mathcal{H}}$, namely the successor of v_{i-1} according to the same-level traversal, and then define $\Delta^i_{\mathcal{H}} \supset \Delta^{i-1}_{\mathcal{H}}$ following the rules described above. We will prove that at each step of the procedure the following properties are respected:

Suppose that $C_1, \ldots, C_j$ are the existentially quantified concepts appearing in $\mathsf{Sub}(\mathcal{T}, C)$ such that $v_i \in C_k^{\mathcal{J}}$ for all $1 \leq k \leq j$. Note that j is necessarily smaller or equal to e. Then we define $\Delta_{\mathcal{H}}^i$ the following way:

- $\Delta_{\mathcal{H}}^{i-1} \cup \{v_i\} \subseteq \Delta_{\mathcal{H}}^i$;
- For all $1 \leq k \leq j$, if C_k has the form $\exists r.D$, then there must exist $w \in \Delta^J$ such that $(v_i, w) \in r^{\mathcal{J}}$ and such that $w \in D^{\mathcal{J}}$. We then add w to $\Delta_{\mathcal{H}}^i$.
- For all $1 \leq k \leq j$, if C_k is an atomic path constraint $\exists P.c$ of depth $d_k \leq d$, then we can find an instance of the path P, namely some tuple $(v_i, w_1, \ldots, w_{d_k}) \in P^{\mathcal{J}}$ that satisfies the constraint c. We then add $w_1, \ldots, w_{d_k}$ to $\Delta_{\mathcal{H}}^i$.
- If v_i still has $t < n$ successors, we choose $w_{t+1}, \ldots, w_n$ arbitrarily among the successors of ε in $\Delta_{\mathcal{J}}$ that we have not yet added to $\Delta_{\mathcal{H}}^i$, and include them.

Figure 1: Procedure for obtaining $\Delta_{\mathcal{H}}^i$.

(1) $\Delta_{\mathcal{H}}^i \setminus \Delta_{\mathcal{H}}^{i-1}$ only contains nodes of the form $v_i w$, i.e. we are only adding descendants of v_i.

(2) All nodes $w \trianglelefteq v_i$, where $\trianglelefteq$ denotes the same-level traversal, have exactly n successors in $\Delta_{\mathcal{H}}^i$.

(3) All nodes $v_i w \in \Delta_{\mathcal{H}}^i$ such that $|w| = k$ for $1 \leq k \leq d$ have at most $(d-k)e$ successors.

Note that, thanks to rule (2), the choice of v_i is well defined, because v_{i-1} has n successors, which are bigger than v_{i-1} itself according to the same level traversal.

Conditions (1) and (2) are trivially verified for $i = 1$. Condition (3) is easily seen as follows: At each step of the procedure we are either adding single nodes which are successors of ε, or we are adding at most d-many nodes $w_1, \ldots, w_d$ such that $|w_l| = l$. Since we are adding at most e of such d-tuples of nodes, it is clearly satisfied that each one of them has at most e successors, which proves condition (3).

Now suppose that for all $j < i$ v_j and $\Delta_{\mathcal{H}}^j$ respect conditions (1)-(3). Let us prove that v_i and $\Delta_{\mathcal{H}}^{i-1}$ also do.

(1) This is clear from the definition of $\Delta_{\mathcal{H}}^i$.

(2) By induction hypothesis, given any node $w \trianglelefteq v_{i-1}$, this has exactly n successors in $\Delta_{\mathcal{H}}^{i-1}$, and by (1) we have not added any new successors of w to $\Delta_{\mathcal{H}}^i$. Suppose v_i is the successor of some node v. Then $v = v_j$ for some $j < i$, and $v_i \in \Delta_{\mathcal{H}}^j$, because all successive steps of the procedure cannot have added new successors of v_j. By inductive hypothesis on (3) we know that at step j, v_i had at most $(d-1)e$ successors, and, by (1), no new successors of v_i can have been added in steps $j+1, \ldots, i-1$. Then, according to the second and third step of the procedure, we will add at most e-many distinct successors of v_i, obtaining a number of successors for v_i that is smaller or equal to n. The last step of the procedure assures that v_i has exactly n successors.

(3) As noted above, v_i is the successor of some v_j such that $j < i$, and by the inductive hypothesis (1) no new descendants of v_i could have been added to the domains $\Delta_{\mathcal{H}}^{j+1}, \ldots, \Delta_{\mathcal{H}}^{i-1}$. Therefore, by inductive hypothesis (3), we know that for all $1 \leq k \leq d$, and for all nodes $v_j w$ with $|w| = k$, $v_j w$ has at most $(d-k)e$-many successors. Since v_i is a direct successor of v_j, this implies that at step $i-1$, all descendants of v_i of the form $v_i w$ with $|w| = k$ correspond to a node $v_j w'$ with $|w'| = k+1$, and have therefore at most $(d-k-1)e$-many successors. Note that k cannot be larger than $d-1$, because all nodes $v_j w'$ with

$|w'| = d$ have no successors. By the second and third step of the procedure, we will add at most $(d \cdot e)$-many nodes $v_i w_k^j$ for $j = 1, \ldots, e$ and $k = 1, \ldots, d$, such that $|w_k^j| = k$ for all k and j. That is, we are adding to $\Delta^i_{\mathcal{H}}$ at most e-many nodes k levels below v_i for each $k = 1, \ldots, d$, which implies that all nodes $v_i w$ with $|w| = k$ have now at most $(d-k)e$-many successors, as wanted.

Now $\Delta_{\mathcal{H}}$ is defined as the union of all $\Delta^i_{\mathcal{H}}$ for $i \in \mathbb{N}$, and $\cdot^{\mathcal{H}}$ and $\gamma_{\mathcal{H}}$ are defined as the restriction of $\cdot^{\mathcal{J}}$ and $\gamma_{\mathcal{J}}$ to $\Delta_{\mathcal{H}}$. Due to properties (1) and (2), we have guaranteed that $\mathcal{H}$ is an n-tree. It is only left to show that $\mathcal{H} \models_{\mathcal{T}} C$, and this is obtained through a simple induction on the structure of the formula, where we use the fact that, whenever an existentially quantified subconcept of $\mathcal{T}$ and C is satisfied on $\mathcal{J}$, we have included in $\mathcal{H}$ the witnesses that made such concept true.

□

Given this result, we can restrict ourselves from now on to those $\mathcal{D}$-interpretations of the form $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ where for each $u, v \in [1,n]^*$ there exists $r \in \mathsf{N_R}$ such that $(u,v) \in r^{\mathcal{I}}$ if and only if there exists $i \in [1,n]$ such that $v = ui$.

3.2 Strong negation normal form

We show now how, requiring that the concrete domain satisfies a property called *negation closure*, we can obtain a *strong negation normal form*, where negation only appears in front of concept names.

Definition 24. We call a σ-structure $\mathcal{D} = (D, R_1^{\mathcal{D}}, R_2^{\mathcal{D}}, \ldots)$ *negation-closed* if for every $R \in \sigma$ the complement of $R^{\mathcal{D}}$ is effectively definable by a positive existential first-order formula, i.e., if there is a computable function that maps each relation symbol $R \in \sigma$ to a positive existential first-order formula $\varphi_R(x_1, \ldots, x_{\mathsf{ar}(R)})$ (i.e., a formula that is built up from relations of σ using $\wedge$, $\vee$, and $\exists$) such that

$$D^{\mathsf{ar}(R)} \setminus R^{\mathcal{D}} = \{(a_1, \ldots, a_{\mathsf{ar}(R)}) \mid \mathcal{D} \models \varphi_R(a_1, \ldots, a_{\mathsf{ar}(R)})\}.$$

Example 25. Consider the structure $(\mathbb{Z}, <, =, (=_a)_{a \in \mathbb{Z}}, (\equiv_{a,b})_{0 \le a < b})$, where $=_a := \{a\}$ is the unary predicate which holds only for a, and $\equiv_{a,b} := \{a + kb \mid k \in \mathbb{Z}\}$ is a unary predicate expressing the fact that some number is congruent to a modulo b. Such structure is negation-closed, we have in fact:

- $\neg x = y$ if and only if $x < y \vee y < x$,[4]
- $\neg x < y$ if and only if $x = y \vee y < x$,
- $\neg x = a$ if and only if $\exists y\, (y = a \wedge (x < y \vee y < x))$, and
- $\neg x \equiv a \bmod b$ if and only if $x \equiv c \bmod b$ for some $0 \le c < b$ with $a \ne c$:

$$\bigvee_{\substack{0 \le c < b \\ a \ne c}} x \equiv c \bmod b\ .$$

Definition 26. We say that an $\mathcal{ALC}^P(\mathcal{D})$-concept φ is in *strong negation normal form* if it is in negation normal form and if, additionally, all constraints $c(x_1, \ldots, x_k)$ do not contain any negation. Consequently we say that a TBox $\mathcal{T}$ is in strong negation normal form if so is the TBox-concept $C_{\mathcal{T}}$.

[4] To improve readability we write $x = y$ instead of $=(x,y)$, $x = a$ instead of $=_a x$, and so on.

Lemma 27. *If $\mathcal{D} = (D, R_1^{\mathcal{D}}, R_2^{\mathcal{D}}, \dots)$ is negation-closed, given a concept C and a TBox $\mathcal{T}$, one can compute $\widehat{C}$ and $\widehat{\mathcal{T}}$ in strong negation normal form such that C is satisfiable with respect to $\mathcal{T}$ if and only if $\widehat{C}$ is satisfiable with respect to $\widehat{\mathcal{T}}$.*

Proof. We can assume that C and $\mathcal{T}$ are in negation normal form, that is, negation only occurs before a concept name or before a relation from the concrete domain, inside a constraint. Using induction, it suffices to eliminate one negated atomic constraint $\theta = \neg R(S^{i_1}x_1, \dots, S^{i_k}x_k)$ from C and $\mathcal{T}$, where $k = \mathsf{ar}(R)$. Let $d = \max\{i_1, \dots, i_k\}$ be the depth of θ. Since $\mathcal{D}$ is negation-closed, we can compute a positive quantifier-free first-order formula $\psi(y_1, \dots, y_k, z_1, \dots, z_m)$ such that

$$(a_1, \dots, a_k) \notin R^{\mathcal{D}} \iff \mathcal{D} \models \exists z_1 \cdots \exists z_m \psi(a_1, \dots, a_k, z_1, \dots, z_m) \; . \tag{2}$$

Let $s_1, \dots, s_m \in \mathsf{Reg}$ be fresh register variables not occurring in ψ. We define $\widehat{C}$ and $\widehat{\mathcal{T}}$ as obtained from C and $C_{\mathcal{T}}$ by replacing every occurrence of the negated relation θ by

$$\psi(S^{i_1}x_1, \dots, S^{i_k}x_k, S^d s_1, \dots, S^d s_m) \; , \tag{3}$$

i.e., we replace in the positive quantifier-free formula $\psi(y_1, \dots, y_k, z_1, \dots, z_m)$ every occurrence of a variable y_j (respectively, z_j) by $S^{i_j}x_j$ (respectively, $S^d s_j$).

The idea is the following: Given a negated atomic constraint θ, we substitute it with a boolean combination of positive ones, involving the same variables appearing in θ ($S^{i_j}x_j$), but also new ones ($S^d s_1 \dots, S^d s_m$) which will carry the value of those existentially quantified values that occur in the formula ψ. These new variables are "placed" at depth d, so that (considering the tree-like structure of an $\mathcal{ALC}^P(\mathcal{D})$ model) we can use a different tuple of values ($S^d s_1, \dots, S^d s_m$) for each path P on which the constraint containing ψ is evaluated.

Now we want to prove that

$$C \text{ is satisfiable with respect to } \mathcal{T} \iff \widehat{C} \text{ is satisfiable with respect to } \widehat{\mathcal{T}} \; .$$

Proof of $\Longrightarrow$. If C is satisfiable w.r.t. $\mathcal{T}$, then by Lemma 23 there is an n-tree $\mathcal{D}$-interpretation $\mathcal{I} = ([1, n]^*, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ such that $\mathcal{I} \models_{\mathcal{T}} C$. We modify $\mathcal{I}$ and obtain a new n-tree $\mathcal{D}$-interpretation $\mathcal{J} = ([1, n]^*, \cdot^{\mathcal{J}}, \gamma_{\mathcal{J}})$, such that $\mathcal{J} \models_{\widehat{\mathcal{T}}} \widehat{C}$. We redefine $\gamma_{\mathcal{J}}$ on the fresh register variables $s_1, \dots, s_m$ and leave otherwise $\gamma_{\mathcal{J}} = \gamma_{\mathcal{I}}$ as follows: Consider $wv \in \Delta$ such that $|v| = d$ and let v_p denote the prefix of v of length i_p for $1 \leq p \leq k$,

D1. By (2), $(\gamma(wv_1, r_1), \dots, \gamma(wv_k, r_k)) \notin R^{\mathcal{I}}$, if and only if there exist values $b_1, \dots, b_m \in D$ such that
$$\mathcal{D} \models \psi(\gamma_{\mathcal{I}}(wv_1, r_1), \dots, \gamma_{\mathcal{I}}(wv_k, r_k), b_1, \dots, b_m) \; .$$
In this case we set $\gamma_{\mathcal{J}}(wv, s_q) = b_q$ for all $1 \leq q \leq m$.

D2. If $(\gamma_{\mathcal{I}}(wv_1, r_1), \dots, \gamma_{\mathcal{I}}(wv_k, r_k)) \in R^{\mathcal{I}}$, we choose $\gamma_{\mathcal{J}}(wv, s_q) \in D$ arbitrarily for all $1 \leq q \leq m$.

D3. Finally, for all $w \in \Delta$ such that $|w| < d$ we choose $\gamma_{\mathcal{J}}(w, s_q) \in D$ arbitrarily for all $1 \leq q \leq m$.

We finally define $X^{\mathcal{J}} := X^{\mathcal{I}}$ for $X \in \mathsf{N_C} \cup \mathsf{N_R}$ and extend it as usual to all $\mathcal{ALC}^P(\mathcal{D})$-concepts. Note that the resulting interpretation function $\cdot^{\mathcal{J}}$ can differ from $\cdot^{\mathcal{I}}$ because $\gamma_{\mathcal{J}}$ may differ from $\gamma_{\mathcal{I}}$.

Using induction we can now prove the following (HP): for all concepts $E \in \mathsf{Sub}(\mathcal{T}, C)$ and all $v \in \Delta$, $v \in E^{\mathcal{I}} \Rightarrow v \in \widehat{E}^{\mathcal{J}}$. Since $\mathcal{I} \models \mathcal{T}$, we know that $v \in (C_{\mathcal{T}})^{\mathcal{I}}$ for all $v \in \Delta$. Then using (HP) we can deduce that $v \in \widehat{C_{\mathcal{T}}}^{\mathcal{J}}$ for all $v \in \Delta$, that is, $\mathcal{J} \models \widehat{\mathcal{T}}$. Furthermore, since $\mathcal{I} \models_{\mathcal{T}} C$

there exists $v \in \Delta$ such that $v \in C^{\mathcal{I}}$. Using (HP) we can then deduce that $v \in \widehat{C}^{\mathcal{J}}$, which together with the previous observation proves $\mathcal{J} \models_{\widehat{\mathcal{T}}} \widehat{C}$.

The proof of (HP) is easy, given the fact that $\mathcal{I}$ and $\mathcal{J}$ coincide on everything except the values of $\gamma_{\mathcal{I}}$ and $\gamma_{\mathcal{J}}$ on the new variables $s_1 \dots s_m$. The only non-trivial case is the one where E has the form $\exists P.c(S^{j_1}y_1, \dots, S^{j_t}y_t)$ or $\forall P.c(S^{j_1}y_1, \dots, S^{j_t}y_t)$, where P is a path of length n and c contains the negated atomic constraint $\theta = \neg R(S^{i_1}x_1, \dots, S^{i_k}x_k)$. We show how to deal with this case: $v \in E^{\mathcal{I}}$ if and only if for some (or for all) tuples $(v_0, \dots, v_n) \in P^{\mathcal{I}}$ with $v_0 = v$ we have that $\mathcal{D} \models c(\gamma_{\mathcal{I}}(v_{j_1}, y_1), \dots, \gamma_{\mathcal{I}}(v_{j_t}, y_t))$. Since $r^{\mathcal{I}} = r^{\mathcal{J}}$ for all $r \in \mathsf{N_R}$ we know that $(v_0, \dots, v_n) \in P^{\mathcal{I}}$ implies $(v_0, \dots, v_n) \in P^{\mathcal{J}}$. Then we simply need to prove $\mathcal{D} \models \widehat{c}(\gamma_{\mathcal{J}}(v_{j_1}, y_1), \dots, \gamma_{\mathcal{J}}(v_{j_t}, y_t), \gamma_{\mathcal{J}}(v_d, s_1), \dots, \gamma_{\mathcal{J}}(v_d, s_m))$. Now $\widehat{c}$ is a boolean combination of relations on the concrete domain $\mathcal{D}$ in negation normal form, obtained from c by replacing every occurrence of θ with the formula (3). We hence apply a second level of induction on the subformulas of c, and reduce ourselves to proving that $\mathcal{D} \models \neg R(\gamma_{\mathcal{I}}(v_{i_1}, x_1), \dots, \gamma_{\mathcal{I}}(v_{i_k}, x_k))$ implies $\mathcal{D} \models \psi(\gamma_{\mathcal{J}}(v_{i_1}, x_1), \dots, \gamma_{\mathcal{J}}(v_{i_k}, x_k), \gamma_{\mathcal{J}}(v_d, s_1), \dots, \gamma_{\mathcal{J}}(v_d, s_m))$. But this is clear, given how we have chosen the values of $\gamma_{\mathcal{J}}(v_d, s_q)$ for $q = 1 \dots m$ according to D1.

Proof of $\Longleftarrow$. If $\widehat{C}$ is satisfiable w.r.t. $\widehat{\mathcal{T}}$, then there exists a $\mathcal{D}$-interpretation $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ such that $\mathcal{I} \models_{\widehat{\mathcal{T}}} \widehat{C}$. We claim that $\mathcal{I}$ is also a model for C and, in particular, for each element $v \in \Delta$ and for each concept $E \in \mathsf{Sub}(\mathcal{T}, C)$, $v \in \widehat{E}^{\mathcal{I}}$ implies $v \in E^{\mathcal{I}}$.

To prove this, again by induction on the structure of all subconcepts in $\mathsf{Sub}(\mathcal{T}, C)$, after a sequence of trivial steps, we find ourselves with the task to show that

$$\mathcal{D} \models \psi(\gamma_{\mathcal{I}}(v_{i_1}, x_1), \dots, \gamma_{\mathcal{I}}(v_{i_k}, x_k), \gamma_{\mathcal{I}}(v_d, s_1), \dots, \gamma_{\mathcal{I}}(v_d, s_m)) \\ \text{implies } \mathcal{D} \models \neg R(\gamma_{\mathcal{I}}(v_{i_1}, x_1), \dots, \gamma_{\mathcal{I}}(v_{i_k}, x_k)),$$

but this is a direct consequence of (2). □

Example 28. Consider the concrete domain $\mathcal{D} = (\mathbb{Z}, <, =, (=_a)_{a \in \mathbb{Z}})$ and the $\mathcal{ALC}^P(\mathcal{D})$-concept $C = \exists rs.[S^1x < S^2x \wedge \neg S^2x = 3]$. An individual d which belongs to an interpretation of C must necessarily have an r-successor d_1 which has an s-successor d_2, such that the value of x in d_1 is smaller than the value of x in d_2, which in turn must be different than 3. As one can see from Example 25, $\mathcal{D}$ is negation-closed, and we can find an existentially quantified positive first order formula, namely

$$\psi(a) = \exists z(z = 3 \wedge (a < z \vee z < a)) \ ,$$

such that $\neg x = 3$ if and only if $\psi(x)$ holds. The strong negation normal form of C is then

$$\widehat{C} = \exists rs.[S^1x < S^2x \wedge S^2y = 3 \wedge (S^2x < S^2y \vee S^2y < S^2x)] \ .$$

As you can see we have introduced a new register variable y and placed it at depth 2 inside the constraint to hold the value that was existentially quantified in ψ.

Now that we have successfully eliminated negation from inside the constraints, there is one last step to do, in order to obtain a normal form that will be useful in the next section. Observe that if a constraint $c(x_1, \dots, x_k)$ does not contain negation, it is possible to apply distributivity repeatedly and obtain an equivalent constraint in DNF or in CNF[5] which still does not contain negation. Therefore we can assume that all path constraints of the form $\exists P.c$ (respectively $\forall P.c$) are such that the constraint c is in DNF (resp. CNF). Using then the fact that universal quantification commutes with conjunction and that existential quantification commutes with disjunction, we can easily prove the following facts:

[5] Disjunctive or conjunctive normal form.

$$\exists P.\bigvee_{i=1}^{n}(a_1^i \wedge \cdots \wedge a_{m_i}^i) \equiv \bigsqcup_{i=1}^{n} \exists P.(a_1^i \wedge \cdots \wedge a_{m_i}^i), \text{ and}$$
$$\forall P.\bigwedge_{i=1}^{n}(a_1^i \vee \cdots \vee a_{n_i}^i) \equiv \bigsqcap_{i=1}^{n} \forall P.(a_1^i \vee \cdots \vee a_{n_i}^i),$$

where each a_j^i is an atomic constraint. Therefore, given a concept C in strong negation normal form, and applying the above described transformations, we can obtain a new concept C' which is still in strong negation normal form, and is such that all path constraints are of the kind $\exists P.c$ (or $\forall P.c$) where c is a conjunction (resp. disjunction) of atomic constraints. We call this the *constraint normal form* of the concept C.

4 The EHD-method

Here we retrace the steps of the work done in [6, 7] for CTL^* and ECTL^* to show how, assuming the right properties on the concrete domain, we can reduce the satisfiability problem of $\mathcal{ALC}^P(\mathcal{D})$ to the satisfiability problem for BMWB over n-trees.

4.1 The EHD-property

We now introduce one of the central notions involved in the decidability proof: the EHD-*property*. EHD stands for "the existence of a homomorphism is definable". This is a property of a relational structure $\mathcal{A}$, expressing the ability of a logic L to distinguish between those structures $\mathcal{B}$ which can be mapped to $\mathcal{A}$ by a homomorphism ($\mathcal{B} \preceq \mathcal{A}$) and those who cannot. Recall the definition of homomorphism (Definition 3).

Definition 29. Let L be a logic (e.g. MSO). A σ-structure $\mathcal{A}$ has the property EHD(L) if there is a computable function that maps every finite subsignature $\tau \subseteq \sigma$ to an L-sentence φ_τ such that for every countable τ-structure $\mathcal{B}$ we have:

$$\mathcal{B} \preceq \mathcal{A} \;\Leftrightarrow\; \mathcal{B} \models \varphi_\tau \ .$$

One can also formulate a variant of the EHD-property for classes of structures:

Definition 30. A class Γ of relational structures over the common signature σ has the property EHD(L) if there is a computable function that maps every finite subsignature $\tau \subseteq \sigma$ to an L-sentence φ_τ such that for every countable τ-structure $\mathcal{B}$ we have:

$$\exists \mathcal{A} \in \Gamma \text{ s.t. } \mathcal{B} \preceq \mathcal{A} \;\Leftrightarrow\; \mathcal{B} \models \varphi_\tau \ .$$

As we will see later, it turns out that for a negation-closed domain $\mathcal{D}$ with the EHD(BMWB)-property, satisfiability of $\mathcal{ALC}^P(\mathcal{D})$ is decidable. For this reason, in the following we will be mainly interested in structures which have the property EHD(L) where L is BMWB or some fragment of this logic. In this cases, we sometimes omit to specify L and simply write EHD.

Example 31. In [6, 7, 5] several relational structures and classes of relational structures are investigated and found to enjoy the property EHD. Most notably

- the integers with equality-, order-, constants-, and modulo-constraints: $(\mathbb{Z}, =, <, (=_a)_{a\in\mathbb{Z}}, (\equiv_{a,b})_{a<b})$,

- the natural numbers with the same relational signature: $(\mathbb{N}, =, <, (=_a)_{a\in\mathbb{N}}, (\equiv_{a,b})_{0\leq a<b})$,
- the rational numbers $(\mathbb{Q}, =, <, (=_q)_{q\in\mathbb{Q}})$,
- the class of all semi-linear orders (see [5]),
- the class of all ordinal trees (see [5]),
- the class of all trees of height h for some fixed $h \in \mathbb{N}$,
- $(\mathbb{Z}^n, <_{\text{lex}}, =)$ where $<_{\text{lex}}$ is the lexicographic order,
- $\text{Allen}_{\mathbb{Z}}$: the set of intervals over the integers together with Allen's relations, which allow to describe their relative positioning.

It was first shown in [6] that $(\mathbb{Z}, <)$ has the property EHD. Consider any countable $\{<\}$-structure $\mathcal{A} = (A, <)$. For $x, y \in A$ we write $x <^* y$ if there exist $x_1, \ldots, x_n$ such that $x < x_1 < \cdots < x_n < y$ in $\mathcal{A}$. We call then $\{x, x_1, \ldots, x_n, y\}$ a $<$-path between x and y.

It was proved in [6] that $\mathcal{A} \preceq (\mathbb{Z}, <)$ if and only if

- $\mathcal{A}$ does not contain a *cycle*, that is, two elements x, y such that $x <^* y < x$, and
- for every two elements $x, y \in A$, there exists a bound n such that one cannot find a $<$ path between x and y with more than n elements.

This is then expressed in the following BMWB-formulas: $\neg\exists x, y(\mathsf{reach}_<(x, y) \wedge y < x)$ and $\forall x \forall y\, \mathsf{B}X \mathsf{Path}(X, x, y)$. Here $\mathsf{reach}_<$ is the same as in ex. 8, where the edge relation E is replaced by $<$, and $\mathsf{Path}(X, x, y)$ is a formula indicating that the set X is a $<$-path from x to y (see [7, ex. 2]). We see here the bounding quantifier in action, bounding the size of all paths between any two elements.

In [11, 12] concrete domains over the rationals are considered for the logics $\mathbb{Q}$-$\mathcal{SHIQ}$ and $\mathcal{TDL}$. This last one differs from $\mathcal{ALC}^P(\mathcal{D})$ only in the fact that it solely allows feature-paths as connectors to the concrete domains. In both cases, it is stated that adding a unary predicate int, allowing to express that a certain concrete value has to be an integer, would be extremely useful. Decidability of reasoning in these logics under this addition remained an open problem. Here we show that the domain $\mathcal{Q} = (\mathbb{Q}, <, \mathsf{int}, \overline{\mathsf{int}})$, where $\mathsf{int} = \mathbb{Z}$ and $\overline{\mathsf{int}} = \mathbb{Q} \setminus \mathbb{Z}$, has the EHD-property. In [7, lem. 38] it is shown that, whenever a domain $\mathcal{D}$ has the EHD-property, then so does $\mathcal{D}^=$, obtain by adding the equality relation. This proves that $\mathcal{Q}^=$ (which is now negation-closed) has the EHD-property. This will imply that satisfiability of $\mathcal{ALC}^P(\mathcal{Q}^=)$ is decidable.

Proposition 32. *$\mathcal{Q}$ has the* EHD*-property.*

Proof. Consider an arbitrary countable structure $\mathcal{A} = (A, <, \mathsf{int}^{\mathcal{A}}, \overline{\mathsf{int}}^{\mathcal{A}})$. We want to prove that $\mathcal{A}$ allows a homomorphism to $(\mathbb{Q}, <, \mathsf{int}, \overline{\mathsf{int}})$ if and only if

H1 $\mathcal{A}$ does not contain two elements x, y such that $x <^* y < x$,

H2 there exists no x such that $x \in \mathsf{int}^{\mathcal{A}} \cap \overline{\mathsf{int}}^{\mathcal{A}}$,

H3 given any two elements $x, y \in A$, there exists a bound n such that each $<$-path between x and y contains at most n elements from $\mathsf{int}^{\mathcal{A}}$.

In this setting, it is only the number of elements of $\mathsf{int}^{\mathcal{A}}$ that needs to be bounded on all path between any two elements. The reason is that, being $\mathbb{Q}$ dense, we can accommodate any countable amount of numbers in any interval, provided that they are not forced to be integers.

Let us denote from now on $I = \mathsf{int}^{\mathcal{A}}$. Properties H1-H3 are easily expressed in BMWB: acyclicity using $\mathsf{reach}_<$ as above, H2 by $\neg\exists x(\mathsf{int}(x) \wedge \overline{\mathsf{int}}(x))$ and H3 is expressed by the formula $\forall x,y\, \mathsf{B}X[X \subseteq I \wedge \exists Z(X \subseteq Z \wedge \mathsf{Path}(Z,x,y))]$.

Now let us prove our claim. The "only if" implication is straightforward, let us consider the "if" direction. Fix an enumeration $a_0, a_1, a_2, \ldots$ of the countable set A. For $n \geq 0$ let $S_n := \{a \in A \mid \exists i,j \leq n : a_i <^* a \text{ and } a <^* a_j\}$, which has the following properties:

(P1) S_n is convex w.r.t. the partial order $<^*$: If $a, c \in S_n$ and $a <^* b$ and $b <^* c$, then $b \in S_n$.

(P2) For $a \in A \setminus S_n$ all $<$-paths between a and S_n are "one-way", i.e., there do not exist $b, c \in S_n$ such that $b <^* a$ and $a <^* c$. This follows from (P1).

(P3) For all $a \in A \setminus S_n$ there exists a bound $c \in \mathbb{N}$ such that all $<$-paths between a and S_n contain at most c elements $x \in I$. Let $c_n^a \in \mathbb{N}$ be the smallest such bound (hence, we have $c_n^a = 0$ if there exist no $<$-paths between a and S_n, or if all $<$-paths do not intersect I).

To see (P3), assume that there only exist $<$-paths from S_n to a but not the other way round (see (P2)); the other case is symmetric. If there is no bound c such that all $<$-paths from S_n to a have at most c elements from I, then by definition of S_n, there is no bound on the number of elements from I on $<$-paths from $\{a_0, \ldots, a_n\}$ to a. By the pigeon principle, there exists $0 \leq i \leq n$ such that the number of elements from I on $<$-paths from a_i to a is unbounded. But this contradicts property (H2).

We build our homomorphism h inductively. For every $n \geq 0$ we define functions $h_n : S_n \to \mathbb{Z}$ such that the following invariants hold for all $n \geq 0$.

(I1) If $n > 0$ then $h_n(a) = h_{n-1}(a)$ for all $a \in S_{n-1}$

(I2) $h_n(S_n)$ is bounded in $\mathbb{Z}$, i.e., there exist $z_1, z_2 \in \mathbb{Z}$ such that $h_n(S_n) \subseteq [z_1, z_2]$.

(I3) for all $a \in S_n$, $h_n(a) \in \mathbb{Z}$ if and only if $a \in I$.

(I4) h_n is a homomorphism from the substructure $\mathcal{A}_{|S_n}$ to $\mathcal{Q}$.

For $n = 0$ we have $S_0 = \{a_0\}$. We set $h_0(a_0) = 0$ if $a_0 \in I$ (any other integer would be also fine), or $h_0(a_0) = 1/2$ otherwise. Properties (I1)–(I4) are easily verified. For $n > 0$, there are four cases.

Case 1. $a_n \in S_{n-1}$, thus $S_n = S_{n-1}$. We set $h_n = h_{n-1}$. Clearly, (I1)–(I4) hold for n.

Case 2. $a_n \notin S_{n-1}$ and there is no $<$-path from a_n to S_{n-1} or vice versa. We set $h_n(a_n) := 0$ if $a_n \in I$ and $h_n(a_n) = 1/2$ otherwise. Note that in this case $S_n = S_{n-1} \cup \{a_n\}$, and (I1)–(I4) follow easily from the induction hypothesis.

Case 3. $a_n \notin S_{n-1}$ and there exist $<$-paths from a_n to S_{n-1}. Then, by (P2) there do not exist paths from S_{n-1} to a_n. Hence, we have

$$S_n = S_{n-1} \cup \{a \in A \mid \exists b \in S_{n-1} : a_n <^* a <^* b\}.$$

We have to assign a value $h_n(a)$ for all $a \in A \setminus S_{n-1}$ that lie along a path from a_n to S_{n-1}. By (I2) there exist $z_1, z_2 \in \mathbb{Z}$ with $h_{n-1}(S_{n-1}) \subseteq [z_1, z_2]$. Recall the definition of c_{n-1}^a from (P3). For all $a \in S := (S_n \setminus S_{n-1}) \cap I$ we set $h_n(a) := z_1 - c_{n-1}^a$. Since $a \in I$, we have $c_{n-1}^a > 0$, hence $h_n(a) < z_1$.

Let us now call $B = S \setminus I$ and fix an enumeration $b_1, b_2, \ldots$ all elements from B. Note that

$$\forall a, b \in B \ a <^* b \Rightarrow c^a_{n-1} \geq c^b_{n-1}\,. \tag{4}$$

For each $b_i \in B$ we will define h_n as the union over i of the functions $h^i_n(b_i)$ defined inductively as follows, so that

(i) $h^i_n(b_i)$ belongs to the interval $(z_1 - c^{b_i}_{n-1} - 1, z_1 - c^{b_i}_{n-1})$, and

(ii) for all $j, k < i$, $b_j <^* b_k$ implies $h^i_n(b_j) < h^i_n(b_k)$.

For b_0 we set $h^0_n(b_0) = z_1 - c^{b_0}_{n-1} - 1/2$, which respects (i) and (ii).

For step i define $h^i_n(b_j) = h^{i-1}_n(b_j)$ for all $j < i$. Then, define $m_i = \max_{j<i}\{z_1 - c^{b_i}_{n-1} - 1\} \cup \{h^i_n(b_j) \mid b_j <^* b_i\}$ and $M_i = \min_{j<i}\{z_1 - c^{b_i}_{n-1}\} \cup \{h^i_n(b_j) \mid b_i <^* b_j\}$. We then set $h^i_n(b_i) = m_i + |M_i - m_i|/2$. We claim that $z_1 - c^{b_i}_{n-1} - 1 \leq m_i < M_i \leq z_1 - c^{b_i}_{n-1}$. The first and last inequality are trivial consequences of the definition of max and min. The central inequality is given by the inductive hypothesis: in fact since all x, y such that $x \in \{b_j \mid j < i,\ b_j <^* b_i\}$ and $y \in \{b_j \mid j < i,\ b_i <^* b_j\}$ are such that $x <^* y$, applying (ii) we obtain $h^i_n(x) < h^i_n(y)$. Furthermore for such x and y, thanks to (4), we know that $c^x_{n-1} \geq c^{b_i}_{n-1} \geq c^y_{n-1}$, and therefore, thanks to (i), $h^i_n(x) < z_1 - c^{b_i}_{n-1}$ and $h^i_n(y) > z_1 - c^{b_i}_{n-1} - 1$. This proves the claim. We can then use this claim to easily deduce that $h^i_n(b_i)$ respects (i) and (ii).

Now that we have defined h_n on the whole $S_n \setminus S_{n-1}$, let us check that it satisfies (I1)– (I4): Invariant (I1) and (I3) hold by definition of h_n. For (I2) note that $h_n(S_n) \subseteq [z_1 - c^{a_n}_{n-1}, z_2]$.

It remains to show (I4), i.e., that h_n is a homomorphism from $\mathcal{A}_{|S_n}$ to $\mathcal{Q}$. By (I3) we have that $a \in I$ implies $h(a) \in \mathbb{Z}$, as wanted. If $\overline{\mathsf{int}}(a)$ holds in $\mathcal{A}$, then by H2 it cannot be $a \in I$. Therefore also by (I3) $h(a) \notin \mathbb{Z}$.

We have to show that for all $b_1 < b_2 \in S_n$ we have $h(b_1) < h(b_2)$.

- If $b_1, b_2 \in S_{n-1}$, then $h_n(b_1) = h_{n-1}(b_1) < h_{n-1}(b_2) = h_n(b_2)$ by induction hypothesis.
- If $b_1 \in S_n \setminus S_{n-1}$ and $b_2 \in S_{n-1}$, we know that $h_n(b_2) = h_{n-1}(b_2) \geq z_1$ while $h_n(b_1) < z_1$ by construction. This directly implies $h_n(b_1) < h_n(b_2)$.
- If $b_2 \in S_n \setminus S_{n-1}$ and $b_1 \in S_{n-1}$, then $b_1 <^* b_2$ and by assumption b_2 must be on a path from a_n to S_{n-1} which contradicts (P2).
- If both b_1 and b_2 belong to $(S_n \setminus S_{n-1}) \cap I$ then $h_n(b_i) := z_1 - c^{b_i}_{n-1}$ for $i \in \{1, 2\}$ Since $b_1 < b_2$, we have $c^{b_1}_{n-1} > c^{b_2}_{n-1}$. This implies $h_n(b_1) < h_n(b_2)$.
- If both b_1 and b_2 belong to $(S_n \setminus S_{n-1}) \setminus I$, then we know by *(ii)* that for the first $i \in \mathbb{N}$ such that h^i_n is defined on both b_1 and b_2, $h^i_n(b_1) < h^i_n(b_2)$. Since $h_n(b_1) = h^i_n(b_1)$ and $h_n(b_2) = h^i_n(b_2)$, we have what we want.
- If $b_1 \in (S_n \setminus S_{n-1}) \cap I$ and $b_2 \in (S_n \setminus S_{n-1}) \setminus I$, we know that $h_n(b_1) = z_1 - c^{b_1}_{n-1}$ and by (i) $h_n(b_2) > z_1 - c^{b_2}_{n-1} - 1$. We also know that since $b_1 < b_2$, we have $c^{b_1}_{n-1} > c^{b_2}_{n-1}$. Therefore $h_n(b_1) < h_n(b_2)$. The symmetrical case is treated analogously.

Case 4. $a_n \notin S_{n-1}$ and there exist paths from S_{n-1} to a_n. For all $a \in S_n \setminus S_{n-1} = \{a \in A \setminus S_{n-1} \mid a$ belongs to a path from S_{n-1} to $a_n\}$, such that $a \in I$ set $h_n(a) = z_2 + c^a_{n-1}$. The rest of the argument goes analogously to Case 3.

This concludes the construction of h_n. By (I1) limit function $h = \bigcup_{i \in \mathbb{N}} h_i$ exists. By (I4) and $A = \bigcup_{i \in \mathbb{N}} S_i$, h is a homomorphism from $\mathcal{A}$ to $(\mathbb{Z}, <)$.

□

4.2 Satisfiability of $\mathcal{ALC}^P(\mathcal{D})$

Definition 33. Fix a concrete domain $\mathcal{D}$. The satisfiability problem for $\mathcal{ALC}^P(\mathcal{D})$ is the following computational problem: *Given an $\mathcal{ALC}^P(\mathcal{D})$-concept C and a TBox $\mathcal{T}$, is there a $\mathcal{D}$-interpretation $\mathcal{I}$ such that $\mathcal{I} \models_{\mathcal{T}} C$?*

We are now ready to introduce the main result of this work:

Theorem 34. *If a concrete domain $\mathcal{D}$ is negation-closed and has the property* EHD(BMWB)*, the satisfiability problem for $\mathcal{ALC}^P(\mathcal{D})$ is decidable.*

The above Thm. 34 can be applied to all the concrete domains listed in Ex. 31 and the new one from Proposition 32, yielding a good number of decidability results for $\mathcal{ALC}^P(\mathcal{D})$ in the presence of general TBoxes, which strictly improves what was known so far.

The idea behind this theorem is to separate the search of a $\mathcal{D}$-interpretation for a concept C with respect to a TBox $\mathcal{T}$ into two parts: In a first step we look for an ordinary $\mathcal{ALC}$ interpretation (i.e., without the valuation function) that satisfies an *abstracted* version of C and $\mathcal{T}$. That is, we replace each atomic constraint appearing in C and $\mathcal{T}$ with a fresh concept name B and obtain a classical $\mathcal{ALC}$-concept C_a and TBox $\mathcal{T}_a$, where the a stands for abstracted. The fact that C_a is satisfiable with respect to $\mathcal{T}_a$ is clearly not enough to guarantee that C is satisfiable with respect to $\mathcal{T}$: for instance the $\mathcal{ALC}^P(\mathcal{D})$-concept $\exists r.(x < Sx \wedge Sx < x)$ is unsatisfiable, while its abstraction $\exists r.(B_1 \sqcap B_2)$ is satisfiable. So the second step consists in creating alongside the interpretation of the abstracted concept what we call a *constraint graph*, a structure in charge of *remembering* the contribution of the constraints that we abstracted away. It turns out that asking that such constraint graph allows a homomorphism to our concrete domain is enough to guarantee that the constraints are satisfied.

For the following definitions let us fix a signature σ, a negation-closed σ-structure $\mathcal{D}$ with the EHD-property as concrete domain, and an $\mathcal{ALC}^P(\mathcal{D})$-concept C and TBox $\mathcal{T}$, both in constraint normal form, in which only the atomic constraints $\theta_1, \ldots, \theta_n$ occur. Let d_i be the depth of each θ_i, and let $B_1, \ldots, B_n$ be concept names that do not appear in $\mathsf{Sub}(\mathcal{T}, C)$.

Definition 35. Let $E = \exists P.c(S^{i_1}x_1, \ldots, S^{i_k}x_k)$ be an existential path constraint, where $P = r_1 \ldots r_p$ and c is a conjunction of the atomic constraints $\theta_1, \ldots, \theta_m$ with $m \leq n$ where the depths are such that $0 =: d_0 \leq d_1 \leq \cdots \leq d_m \leq d_{m+1} := p$ (if this is not the case it will suffice to reorder the constraints). Then we define the *abstraction* of E as

$$E_a = \exists P_1.(B_1 \sqcap \exists P_2.(B_2 \sqcap \cdots \exists P_m.(B_m \sqcap \exists P_{m+1}.\top) \ldots), \tag{5}$$

where $\exists P_i$ is short for $\exists r_{d_{i-1}+1} \ldots \exists r_{d_i}$. It can happen that $d_i = d_{i+1}$, in which case $\exists P_{i+1}$ is empty.

Symmetrically, if $E = \forall P.c(S^{i_1}x_1, \ldots, S^{i_k}x_k)$ where c is a disjunction of atomic constraints containing $\theta_1, \ldots, \theta_m$ with $0 =: d_0 d_1 \leq \cdots \leq d_m \leq d_{m+1} := p$, we define

$$E_a = \forall P_1.(B_1 \sqcup \forall P_2.(B_2 \sqcup \cdots \forall P_m.(B_m \sqcup \forall P_{m+1} \bot) \ldots), \tag{6}$$

where P_i is defined as above. We define C_a and $\mathcal{T}_a$ as the $\mathcal{ALC}$-concept and TBox obtained by C and $\mathcal{T}$ by replacing every occurrence of a path constraint E by its abstraction E_a.

Notice how we use the fact that the constraints from $\mathcal{ALC}^P(\mathcal{D})$ are local to individuate the "lower" node involved in the constraint (the one at depth d_i) and mark it as belonging to the fresh concept B_i. This way, when navigating a tree-model of the abstracted concept C_a with respect to $\mathcal{T}_a$, we know that all paths of length d_i that end in a node marked with B_i should satisfy the constraint θ_i. This would not work if the constraints were non-local.

Example 36. If $C = \exists.r_1r_2r_3(x = y \wedge x < S^2x \wedge S^1y = S^2x)$, the abstraction C_a is

$$B_1 \sqcap \exists r_1.\exists r_2.(B_2 \sqcap B_3 \sqcap \exists r_3.\top)\,,$$

where we have assigned the new concept names to the atomic constraints in order of appearance.

Definition 37. Given an n-tree $\mathcal{D}$-interpretation $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ such that ${B_1}^{\mathcal{I}} = \cdots = {B_m}^{\mathcal{I}} = \emptyset$, we define the abstraction of $\mathcal{I}$ as the interpretation $\mathcal{I}_a = ([1,n]^*, \cdot^{\mathcal{I}_a})$ where $\cdot^{\mathcal{I}_a}$ is defined as

- $A^{\mathcal{I}_a} = A^{\mathcal{I}}$ for all $A \in (\mathsf{N_C} \setminus \{B_1, \ldots, B_m\})$,
- $r^{\mathcal{I}_a} = r^{\mathcal{I}}$ for all $r \in \mathsf{N_R}$,
- if $\theta_j = R(S^{i_1}x_1, \ldots, S^{i_k}x_k)$ has depth d_j then $u \in {B_j}^{\mathcal{I}_a}$ if and only if
  - $u = wv$ for some $w, v \in [1,n]^*$ with $|v| = d_j$, and
  - $(\gamma(wv_1, x_1), \ldots, \gamma(wv_k, x_k)) \in R^{\mathcal{D}}$,

  where v_t denotes the prefix of v of length i_t.

Hence, the fact that an element wv with $|v| = d_j$ belongs to the interpretation of B_j means that the atomic constraint θ_j is satisfied along every path that starts in node w and descends in the tree down via wv.

Now let $\mathsf{Reg}_{C,\mathcal{T}}$ be the set of register variables occurring in C and $\mathcal{T}$.

Definition 38. Given a tree shaped interpretation $\mathcal{J} = ([1,n]^*, \cdot^{\mathcal{J}})$ where the new concept names $B_1, \ldots, B_m$ can have a non-empty interpretation, we define a countable σ-structure $\mathcal{G}_{\mathcal{J}} = (([1,n]^* \times \mathsf{Reg}_{C,\mathcal{T}}), R_1^{\mathcal{G}}, R_2^{\mathcal{G}}, \ldots)$ (the *constraint graph* of $\mathcal{J}$) as follows: The interpretation $R^{\mathcal{G}}$ of the relation symbol $R \in \sigma$ contains all k-tuples $((wv_1, x_1), \ldots, (wv_k, x_k))$, where $k = \mathsf{ar}(R)$, for which there are $1 \leq j \leq m$ and $v \in [1,n]^{d_j}$ such that $wv \in B_j^{\mathcal{J}}$, and $\theta_j = R(S^{i_1}x_1, \ldots, S^{i_k}x_k)$, where v_t still denotes the prefix of v of length i_t.

The domain of $\mathcal{G}_{\mathcal{J}}$ has one element for each pair (v, x) where v is a member of the domain of $\mathcal{J}$ and x is a register variable appearing in C. When we abstract an atomic constraint θ_i we replace it with its *placeholder* B_i, but any appearance of B_i marks a path where θ_i needs to hold. Such information is stored in the relations of $R^{\mathcal{G}}$. $\mathcal{G}_{\mathcal{J}}$ is called a constraint graph because (when all relations are binary) it can be seen as a graph with different kinds of edges (one for each relation) representing the atomic constraints, and we will use this image in the following.

Example 39. Let $\mathcal{D}$ be a concrete domain having $<$ and $=$ in its signature and suppose the concept names B_1 and B_2 are used to replace the atomic constraints $\theta_1 = (x = y)$ and $\theta_2 = (x < Sx)$ of depth $d_1 = 0$ and $d_2 = 1$, respectively. In Figure 2 we show how to build the constraint graph associated to an ordinary 2-tree interpretation $\mathcal{J}$.

In the next theorem, we illustrate what is the connection between the satisfiability of an $\mathcal{ALC}^P(\mathcal{D})$ concept w.r.t. a TBox, and the satisfiability of its abstraction. Let C and $\mathcal{T}$ be an $\mathcal{ALC}^P(\mathcal{D})$-concept and TBox in constraint normal form, and let $n = d \cdot \#_E(C, \mathcal{T})$ where d is the maximum depth of all constraints appearing in $\mathsf{Sub}(\mathcal{T}, C)$. Then the following holds:

Theorem 40. *C is satisfiable with respect to $\mathcal{T}$ if and only if there exists an ordinary n-tree interpretation $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}})$ such that $\mathcal{I} \models_{\mathcal{T}_a} C_a$ and such that $\mathcal{G}_{\mathcal{I}} \preceq \mathcal{D}$.*

Proof. Let $\theta_1, \ldots, \theta_m$, $d_1, \ldots, d_m$ and $\mathsf{Reg}_{C,\mathcal{T}}$ be defined as before.

($\Rightarrow$) Without loss of generality assume that $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}}, \gamma_{\mathcal{I}})$ is an n-tree $\mathcal{D}$-interpretation such that $\mathcal{I} \models_{\mathcal{T}} C$. Our first claim is that $\mathcal{I}_a \models_{\mathcal{T}_a} C_a$, which we show by induction, proving that for all $v \in [1,n]^*$ and for all subconcepts $E \in \mathsf{Sub}(\mathcal{T}, C)$, $u \in E^{\mathcal{I}}$ implies $u \in E_a^{\mathcal{I}_a}$.

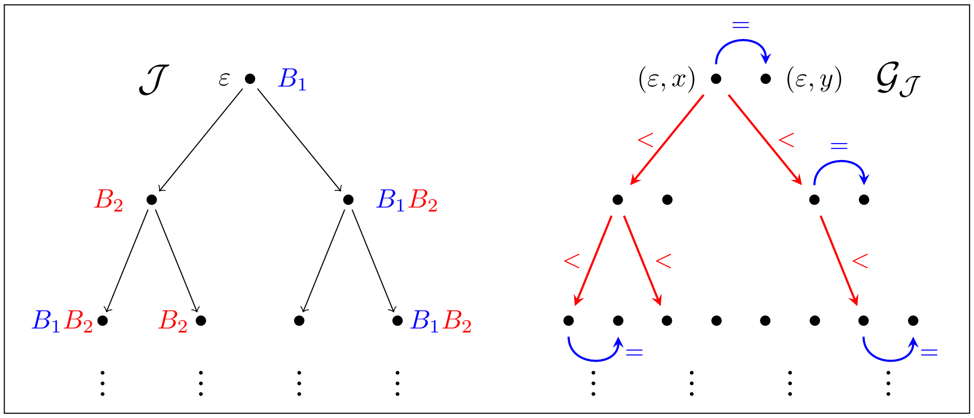

Figure 2: An ordinary 2-tree interpretation where B_1 and B_2 have a non-empty interpretation and its associated constraint graph from Example 39.

- If $E \in \mathsf{Sub}(\mathcal{T}, C)$ is a concept name, then $E_a = E$. Since $E^{\mathcal{I}_a} = E^{\mathcal{I}}$ for all $E \in (\mathsf{N_C} \backslash \{B_1, \dots, B_m\})$, and since $B_1, \dots, B_m \notin \mathsf{Sub}(\mathcal{T}, C)$, we have that $u \in E^{\mathcal{I}} \Leftrightarrow u \in E_a^{\mathcal{I}_a}$. This also proves the case $E = \neg F$ with $F \in \mathsf{N_C}$.

- If $E = F \sqcap G$, then $u \in E^{\mathcal{I}}$ implies $u \in F^{\mathcal{I}}$ and $u \in G^{\mathcal{I}}$. By induction hypothesis we have that $u \in F_a^{\mathcal{I}_a}$ and $u \in G_a^{\mathcal{I}_a}$ which yields $u \in (F_a \sqcap G_a)^{\mathcal{I}_a} = E_a^{\mathcal{I}_a}$.

- If $E = F \sqcup G$, we use the induction hypothesis as in the case above.

- If $E = \exists r.F$ and $u \in E^{\mathcal{I}}$, then we know that there exists an element $v \in [1, n]^*$, such that $(u, v) \in r^{\mathcal{I}}$ and $v \in F^{\mathcal{I}}$. Then $(u, v) \in r^{\mathcal{I}_a}$ by definition of $\mathcal{I}_a$ and $v \in F_a^{\mathcal{I}_a}$ by induction hypothesis. Together we obtain $u \in (\exists r.F_a)^{\mathcal{I}_a} = E_a^{\mathcal{I}_a}$.

- The case $E = \forall r.F$ is treated analogously.

- Let $E = \exists P.c(S^{i_1}x_1, \dots, S^{i_t}x_t)$ with $P = r_1 \cdots r_p$. Since C and $\mathcal{T}$ are in constraint normal form, we can assume that (eventually renaming the atomic constraints) $c = \theta_1 \wedge \dots \wedge \theta_n$ where the depths $d_1, \dots, d_n$ satisfy $0 =: d_0 \leq d_1 \leq \dots \leq d_n \leq d_{n+1} := p$. Since $u \in E^{\mathcal{I}}$, we know that there exists a tuple $(u_0, \dots, u_p) \in P^{\mathcal{I}}$ such that $u_0 = u$ and $\mathcal{D} \models c(\gamma(u_{i_1}, x_1), \dots, \gamma(u_{i_t}, x_t))$. If $\theta_i = R(S^{j_1}y_1, \dots, S^{j_k}y_k)$, this means that $(\gamma(u_{j_1}, y_1), \dots, \gamma(u_{j_k}, y_k)) \in R^{\mathcal{D}}$. By definition of $\mathcal{I}_a$, this implies that $u_{d_i} \in B_i^{\mathcal{I}_a}$. Now, since $(a, b) \in r^{\mathcal{I}}$ implies $(a, b) \in r^{\mathcal{I}_a}$, then $(u_0, \dots, u_p) \in P^{\mathcal{I}_a}$ as well. This, together with the fact that $u_{d_i} \in B_i^{\mathcal{I}_a}$ for $i = 1, \dots, n$, implies that $u \in (\exists P_1.(B_1 \sqcap \exists P_2.(B_2 \sqcap \dots \exists P_n.(B_n \sqcap \exists P_{n+1}.\top) \dots))^{\mathcal{I}_a}$, where $\exists P_i$ is short for $\exists r_{d_{i-1}+1} \dots \exists r_{d_i}$. Which is exactly what we wanted to show.

- Let $E = \forall P.c(S^{i_1}x_1, \dots, S^{i_t}x_t)$ with $P = r_1 \cdots r_p$. We can assume that $c = \theta_1 \vee \dots \vee \theta_n$ where the depths $d_1, \dots, d_n$ satisfy $0 =: d_0 \leq d_1 \leq \dots \leq d_n \leq d_{n+1} := p$. Given $u \in E^{\mathcal{I}}$, we want to prove that $u \in E_a^{\mathcal{I}_a}$, where E_a is defined as $\forall P_1.(B_1 \sqcup \forall P_2.(B_2 \sqcup \dots \forall P_n.(B_n \sqcup \forall P_{n+1} \bot) \dots)$ with $P_i = r_{d_{i-1}+1} \cdots r_{d_i}$. Towards a contradiction suppose $u \notin E_a^{\mathcal{I}_a}$, then u belongs to the interpretation in $\mathcal{I}_a$ of $\neg E_a = \exists P_1.(\neg B_1 \sqcap \exists P_2.(\neg B_2 \sqcap \dots \exists P_n.(\neg B_n \sqcap \exists P_{n+1}.\top) \dots)$. This implies that there exists a path $(v_0, \dots, v_p) \in P^{\mathcal{I}_a}$ such that $v_{d_i} \in (\neg B_i)^{\mathcal{I}_a}$ for all $i = 1, \dots, n$. Given the fact that, by definition, for all role names $r \in \mathsf{N_R}$ $(a, b) \in r^{\mathcal{I}}$ if and only if $(a, b) \in r^{\mathcal{I}_a}$, we know that $(v_0, \dots, v_p) \in P^{\mathcal{I}}$. Then, since $u \in E^{\mathcal{I}}$, this implies that $\mathcal{D} \models c(\gamma(v_{i_1}, x_1), \dots, \gamma(v_{i_t}, x_t))$. In particular, i must exist such that $\theta_i = R(S^{j_1}y_1, \dots, S^{j_k}y_k)$ and $(\gamma(v_{j_1}, y_1), \dots, \gamma(v_{j_k}, y_k)) \in R^{\mathcal{D}}$. By Definition 37, this implies that $v_{d_i} \in B_i$. We have reached a contradiction.

The second claim is that $\mathcal{G}_{\mathcal{I}_a} \preceq \mathcal{D}$. More specifically, we want to prove that the valuation function $\gamma_{\mathcal{I}} : ([1,n]^* \times \mathsf{Reg}_{C,\mathcal{T}}) \rightarrow D$ is a homomorphism. For this, suppose that a tuple $\big((u_1,x_1),\ldots,(u_k,x_k)\big) \in R^{\mathcal{G}}$. By Definition 38 this means that there exist $j \in \{1,\ldots,m\}$ and $wv \in (B_j)^{\mathcal{I}_a}$ such that θ_j has the form $R(S^{i_1}x_1,\ldots,S^{i_k}x_k)$ with depth d_j and such that $v = v_1 \cdots v_{d_j}$ and $u_t = wv_{i_t}$ for all $t = 1 \ldots k$. By Definition 37, this means that $(\gamma_{\mathcal{I}}(u_1,x_1),\ldots,\gamma_{\mathcal{I}}(u_t,x_t)) \in R^{\mathcal{D}}$, as wanted.

($\Leftarrow$) Now we want to show that, given an ordinary n-tree interpretation $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}})$ such that $\mathcal{I} \models_{\mathcal{T}_a} C_a$ and a homomorphism h from $\mathcal{G}_{\mathcal{I}}$ to $\mathcal{D}$, we can construct a $\mathcal{D}$-interpretation $\mathcal{J}$ such that $\mathcal{J} \models_{\mathcal{T}} C$. Let us define $\mathcal{J} = ([1,n]^*, \cdot^{\mathcal{J}}, h)$ where $\cdot^{\mathcal{J}}$ coincides with $\cdot^{\mathcal{I}}$ on all concept names and role names, and is extended to all concepts using the valuation function h. We prove by induction that, for all concepts $E \in \mathsf{Sub}(\mathcal{T}, C)$ and for all $u \in [1,n]^*$, $u \in (E_a)^{\mathcal{I}}$ implies $u \in E^{\mathcal{J}}$.

- If $E \in \mathsf{N_C}$, then $E_a = E$ and since $\cdot^{\mathcal{J}} = \cdot^{\mathcal{I}}$ on concept names, $u \in (E_a)^{\mathcal{I}}$ if and only if $u \in E^{\mathcal{J}}$. This also proves the case $E = \neg F$ with $F \in \mathsf{N_C}$.
- If $E = F \sqcup G$ or $E = F \sqcap G$, we can easily use the induction hypothesis.
- If $E = \exists r.F$ or $E = \forall r.F$, then we can use the induction hypothesis plus the fact that $r^{\mathcal{I}} = r^{\mathcal{J}}$ to show that $u \in E_a^{\mathcal{I}}$ implies $u \in E^{\mathcal{J}}$.
- Suppose $E = \exists P.c(S^{i_1}x_1,\ldots,S^{i_k}x_k)$ where $P = r_1 \cdots r_p$ is a role-path of length p and c is a conjunction of atomic constraints $\theta_1 \wedge \cdots \wedge \theta_n$ with depths $d_1,\ldots,d_n$ such that $0 =: d_0 \leq d_1 \leq \cdots \leq d_n \leq d_{n+1} := p$. Then $E_a = \exists P_1.(B_1 \sqcap \exists P_2.(B_2 \sqcap \ldots \exists P_n.(B_n \sqcap \exists P_{n+1}.\top)\ldots)$ with $P_i = r_{d_{i-1}+1} \cdots r_{d_i}$. If $u \in (E_a)^{\mathcal{I}}$ then there exists a tuple $(u_0,\ldots,u_p) \in P^{\mathcal{I}}$ with $u_0 = u$ and such that $u_{d_i} \in (B_i)^{\mathcal{I}}$ for $i = 1,\ldots,n$. Fix $i \in \{1,\ldots,n\}$, if θ_i has the form $R(S^{j_1}y_1,\ldots,S^{j_t}y_t)$, according to Definition 38, this means that $((u_{j_1},y_1),\ldots,(u_{j_t},y_t)) \in R^{\mathcal{G}}$. Now, since h is a homomorphism from $\mathcal{G}_{\mathcal{I}}$ to $\mathcal{D}$, we have $(h(u_{j_1},y_1),\ldots,h(u_{j_t},y_t)) \in R^{\mathcal{D}}$, which means that $\mathcal{D} \models R(h(u_{j_1},y_1),\ldots,h(u_{j_t},y_t))$. Since this is true for an arbitrary $i \in \{1,\ldots,n\}$ this holds true for the conjunction $\theta_1 \wedge \cdots \wedge \theta_n$, that is $\mathcal{D} \models c(h(u_{i_1},x_1),\ldots,h(u_{i_k},x_k))$. Also, since $r^{\mathcal{I}} = r^{\mathcal{J}}$, we know that $(u_0,\ldots,u_p) \in P^{\mathcal{J}}$. This means that $u \in E^{\mathcal{J}}$, as wanted.
- If $E = \forall P.c(S^{i_1}x_1,\ldots,S^{i_k}x_k)$ with $P = r_1 \cdots r_p$. As usual, we can assume that $c = \theta_1 \vee \cdots \vee \theta_n$ where the depths $d_1,\ldots,d_n$ satisfy $0 =: d_0 \leq d_1 \leq \cdots \leq d_n \leq d_{n+1} := p$. Then $E_a = \forall P_1.(B_1 \sqcup \forall P_2.(B_2 \sqcup \ldots \forall P_n.(B_n \sqcup \forall P_{n+1}\bot)\ldots)$ with $P_i = r_{d_{i-1}+1} \cdots r_{d_i}$. Suppose that $u \in (E_a)^{\mathcal{I}}$, we want to prove that $u \in E^{\mathcal{J}}$. Towards a contradiction, suppose that $u \notin E^{\mathcal{J}}$. This means that there exists an instance of the path P, a tuple $(u_0,\ldots,u_p) \in P^{\mathcal{J}}$ with $u_0 = u$, such that $\mathcal{D} \not\models c(h(u_{i_1},x_1),\ldots,h(u_{i_k},x_k))$, that is $\neg\theta_1 \wedge \cdots \wedge \neg\theta_n$ is satisfied along the path $(u_0,\ldots,u_p)$. Since $r^{\mathcal{I}} = r^{\mathcal{J}}$ for all role names, $(u_0,\ldots,u_p)$ also belongs to $P^{\mathcal{I}}$. Because $u \in (E_a)^{\mathcal{I}}$, we have that there exists some $i \in \{1,\ldots,n\}$ such that $u_{d_i} \in (B_i)^{\mathcal{I}}$. By Definition 38, if $\theta_i = R(S^{j_1}y_1,\ldots,S^{j_t}y_t)$ then $((u_{j_1},y_1),\ldots,(u_{j_t},y_t)) \in R^{\mathcal{G}}$. Since h is a homomorphism, $(h(u_{j_1},y_1),\ldots,h(u_{j_t},y_t)) \in R^{\mathcal{D}}$. This means that θ_i is satisfied on $(u_0,\ldots,u_p)$, which is in contrast with what derived above. Therefore u must belong to $E^{\mathcal{J}}$, as wanted.

$\square$

We are now almost ready to give the proof of the main result of this work. We only need a few additional results:

Definition 41. Given an n-tree ordinary interpretation $\mathcal{I} = ([1,n]^*, \cdot^{\mathcal{I}})$ we define an n-tree $T(\mathcal{I})$ over the signature $\{S\} \cup \mathsf{N_C} \cup \mathsf{N_R}$, where $\mathsf{N_C}$ and $\mathsf{N_R}$ are seen as unary predicates whose interpretation is given by: $A^{T(\mathcal{I})} = A^{\mathcal{I}}$ for each $A \in \mathsf{N_C}$ and $r^{T(\mathcal{I})} = \{xi \in [1,n]^* \mid (x,xi) \in r^{\mathcal{I}}\}$ for all $r \in \mathsf{N_R}$.

Remark 42. The only difference between an n-tree interpretation $\mathcal{I}$ and its induced n-tree $T = T(\mathcal{I})$ is the fact that roles are turned into unary predicates such that, if a pair $(x, y) \in r^{\mathcal{I}}$, now $y \in r^T$. In particular, if we define $\mathcal{G}_T = (([1,n]^* \times \mathsf{Reg}_{C,\mathcal{T}}), R_1^{\mathcal{G}}, R_2^{\mathcal{G}}, \dots)$ (the *constraint graph* of T) in exactly the same way as in Definition 38, only substituting the interpretation $\mathcal{J}$ with T, what we obtain is that $\mathcal{G}_{\mathcal{I}} = \mathcal{G}_T$.

Lemma 43. *Given C and $\mathcal{T}$ an ordinary $\mathcal{ALC}$-concept and TBox in negation normal form, we can write a* FO *formula φ over the signature $\{S\} \cup \mathsf{N_C} \cup \mathsf{N_R}$, where all elements of $\mathsf{N_C} \cup \mathsf{N_R}$ are seen as unary symbols, such that for any given n-tree interpretation $\mathcal{I} = ([0,1]^*, \cdot^{\mathcal{I}})$ we have $\mathcal{I} \models_{\mathcal{T}} C$ if and only if $T(\mathcal{I}) \models \varphi$.*

Proof. The method is similar to what is described in [2, Chapter 3], with the only difference that here we are only allowed to use unary predicates. Roles and features are then seen as unary predicates added to the second node of the relation, which can be done only due to the fact that we are considering tree-shaped models. We define two translations π_x and π_y which inductively map $\mathcal{ALC}$ concepts to FO formulas with only one free variable, x or y respectively:

- $\pi_i(A) := A(i)$ for each $A \in \mathsf{N_C}$ and $i = x, y$;
- $\pi_i(\neg A) := \neg A(i)$ for each $A \in \mathsf{N_C}$ and $i = x, y$;
- $\pi_i(D \sqcap E) := \pi_i(D) \wedge \pi_i(E)$ for $i = x, y$;
- $\pi_i(D \sqcup E) := \pi_i(D) \vee \pi_i(E)$ for $i = x, y$;
- $\pi_i(\exists r.D) := \exists j.S(i,j) \wedge r(j) \wedge \pi_j(D)$ for $(i,j) = (x,y)$ or (y,x);
- $\pi_i(\forall r.D) := \forall j.(S(i,j) \wedge r(j)) \rightarrow \pi_j(D)$ for $(i,j) = (x,y)$ or (y,x);

Now let R be the set of role names appearing in C and $\mathcal{T}$, and let $F \subseteq R$ be feature names. Keeping in mind that the root ε of a tree is definable in first order logic, we define

$$\psi_R := \forall (x \neq \varepsilon).\ \bigvee_{r \in R} r(x)\ \wedge \bigwedge_{r,s \in R, r \neq s} \neg(r(x) \wedge s(x))$$

$$\psi_F := \forall x. \forall (y \neq z).\ S(x,y) \wedge S(x,z) \rightarrow \bigwedge_{f \in F} \neg(f(y) \wedge f(z))\,.$$

ψ_R enforces that each pair of elements (x, y), where y is a successor of x, is assigned a unique role name. ψ_F ensures that the functionality of the features is respected. Then we can prove easily that given a tree-shaped interpretation $\mathcal{I}$ and a TBox $\mathcal{T} = \{C_1, \dots, C_t\}$, $\mathcal{I} \models_{\mathcal{T}} C$ if and only if $T(\mathcal{I})$ is a model for the following FO formula

$$\varphi = \exists x.\pi_x(C) \wedge \forall x.(\pi_x(C_1) \wedge \dots \wedge \pi_x(C_t)) \wedge \psi_R \wedge \psi_F\ .$$

□

Lemma 44. *Let $C, \mathcal{T}$ and φ be as in Lemma 43. Given an n-tree T over the relational signature $\{S\} \cup \mathsf{N_R} \cup \mathsf{N_F}$ that satisfies φ we can build an n-tree interpretation $\mathcal{I}$ such that $T = T(\mathcal{I})$ and such that $\mathcal{I} \models_{\mathcal{T}} C$.*

Sketch of proof. The fact that $T \models \varphi$ means in particular that $T \models \psi_R \wedge \psi_F$, which guarantees that each node of the tree T is assigned at most one role name, and that the functionality of the features is respected. We can therefore safely define $A^{\mathcal{I}} = A^T$ for all $A \in \mathsf{N_C}$ and $r^{\mathcal{I}} = \{(x,y) \in ([1,n]^*)^2 \mid S(x,y) \text{ and } y \in r^T\}$ for all $r \in \mathsf{N_R}$ and obtain a tree shaped interpretation. It is easy to see that $T(\mathcal{I}) = T$, and $\mathcal{I} \models_{\mathcal{T}} C$ can be proved by structural induction. □

Now we show a useful property of BMWB, which is also needed to prove our main result.

Definition 45. Let $k \in \mathbb{N}$ and let $\mathcal{A} = (A, R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \dots)$ be a structure over the signature σ that does not contain relation symbols $\sim, P_1, P_2, \dots, P_k$ ($\sim$ is binary and all P_i are unary). The *k-copy of* $\mathcal{A}$, denoted by $\mathcal{A}^{\times k}$, is the $(\sigma \cup \{\sim, P_1, P_2, \dots, P_k\})$-structure with the domain $(A \times \{1, 2, \dots, k\})$ and

- for all $R \in \sigma$ if R has arity m,

$$R^{\mathcal{A}^{\times k}} = \{((a_1, i), (a_2, i), \dots, (a_m, i)) \mid (a_1, a_2, \dots, a_m) \in R^{\mathcal{A}}, 1 \leq i \leq k\} ,$$

- $\sim^{\mathcal{A}^{\times k}} = \{((a, i_1), (a, i_2)) \mid a \in A, 1 \leq i_1, i_2 \leq k\}$, and
- for each $1 \leq m \leq k$, ${P_m}^{\mathcal{A}^{\times k}} = \{(a, m) \mid a \in A\}$.

Given a structure $\mathcal{A}$, the k-copy operation creates a new structure, $\mathcal{A}^{\times k}$, which contains k many copies of $\mathcal{A}$: there are k disjoint substructures of $\mathcal{A}^{\times k}$ (identifiable through the predicates $P_1, \dots, P_k$) which, seen as σ-structures, are isomorphic to $\mathcal{A}$. The additional binary predicate $\sim$ relates all those members of $\mathcal{A}^{\times k}$ which are a duplicate of the same element in $\mathcal{A}$.

The following proposition states that BMWB is *compatible* with the k-copy operation, i.e., whatever property we can specify on $\mathcal{A}^{\times k}$ using BMWB can also be recognized by BMWB directly on $\mathcal{A}$.

Proposition 46 (Prop. 2.26 of [4])**.** *Let $k \in \mathbb{N}$ be some number, $\mathcal{A}$ some infinite structure over the signature σ, and $\tau = \sigma \cup \{\sim, P_1, P_2, \dots, P_k\}$ an extension of σ by one fresh binary relation symbol $\sim$ and k fresh unary relation symbols $P_1, \dots, P_k$. Given a* BMWB*-sentence φ over τ, we can compute a* BMWB*-sentence φ^k over σ such that $\mathcal{A}^{\times k} \models \varphi$ if and only if $\mathcal{A} \models \varphi^k$.*

Let $\tau \subseteq \sigma$, we say a τ-structure $\mathcal{A}$ with domain A is FO*-interpretable* in a σ-structure $\mathcal{B}$ with domain B, if there exists a FO-formula φ such that $A \cong \{b \in B \mid \mathcal{B} \models \varphi(b)\}$, and for each $R \in \tau$ of arity k, there exists a FO-formula φ_R such that $R^{\mathcal{A}} \cong \{(b_1, \dots, b_k) \in B^k \mid \mathcal{B} \models \varphi_R(b_1, \dots, b_k)\}$. Intuitively, the fact that $\mathcal{A}$ is interpretable in $\mathcal{B}$ means that we can *describe* $\mathcal{A}$ inside $\mathcal{B}$ using FO logic.

Lemma 47. *Suppose $\mathsf{Reg}_{C,\mathcal{T}} = \{x_1, \dots, x_k\}$, then for an n-tree T over the signature $\{S\} \cup \mathsf{N_C} \cup \mathsf{N_R}$, the structure $\mathcal{G}_T$ is* FO*-interpretable in $T^{\times k}$.*

Proof. The domains of $\mathcal{G}_T$ and $T^{\times k}$, $([1, n]^* \times \{x_1, \dots, x_k\})$ and $([1, n]^* \times \{1, 2, \dots, k\})$ respectively, are trivially in a bijection through the mapping $f : (v, x_k) \mapsto (v, k)$. We then extend the bijection f to tuples of elements of $([1, n]^* \times \{x_1, \dots, x_k\})$ as $f(a_1, \dots, a_t) = (f(a_1), \dots, f(a_t))$. We claim that the relations $R_1^{\mathcal{G}}, R_2^{\mathcal{G}}, \dots$ from $\mathcal{G}_T$ can be represented in $T^{\times k}$ using first order logic, let us describe how: Suppose the relation $R \in \sigma$ of arity t is used to form one of the atomic constraints $\theta = R(S^{i_1} y_1, \dots, S^{i_t} y_t)$, with $y_1, \dots, y_t \in \{x_1, \dots, x_k\}$ and $d = \max\{i_1, \dots, i_t\}$. Then we know that a tuple $\big((v_1, y_1), \dots, (v_t, y_t)\big)$ belongs to $R^{\mathcal{G}}$ if (1) there exist elements $w_0, w_1, \dots, w_d \in [1, n]^*$ such that $w_{i_l} = v_l$ for $l = 1, \dots, t$ and $S(w_{j-1}, w_j)$ holds in T for $j = 1, \dots, d$, and (2) $v_d \in B_j^T$. We would like to identify the tuples in $T^{\times k}$ in bijection through f with those tuples in $\mathcal{G}_T$ satisfying conditions (1) and (2). These are the ones that satisfy the following FO formula

$$\varphi_\theta(a_1, \dots, a_t) = \exists b_0 \dots \exists b_d \bigwedge_{j=1,\dots,d} S(b_{j-1}, b_j) \wedge \bigwedge_{l=1,\dots,t} b_{i_l} \sim a_l \wedge \bigwedge_{i=1,\dots,t} P_{z_i}(a_i)$$

where $i_1, \dots, i_t$ are the same indices appearing in $\theta = R(S^{i_1} y_1, \dots, S^{i_t} y_t)$ and z_i is such that $y_i = x_{z_i}$. Once we have defined φ_{θ_j} for the atomic constraints $\theta_1, \dots, \theta_n$, which appear in C

and $\mathcal{T}$, we can state the following: if the relation R of arity t is used in all and only $\theta_{j_1}, \ldots, \theta_{j_k}$, then $\bar{a} = (a_1, \ldots, a_t) \in R^{\mathcal{G}}$ if and only if $\varphi_R(f(\bar{a})) = \varphi_{\theta_{j_1}}(f(\bar{a})) \vee \cdots \vee \varphi_{\theta_{j_k}}(f(\bar{a}))$ holds. If the relation $R \in \sigma$ of arity t is not used in any of the atomic constraints $\theta_{j_1}, \ldots, \theta_{j_k}$, then there will be no tuple in $\mathcal{G}_T$ which belongs to $R^{\mathcal{G}}$. Therefore $(a_1, \ldots, a_t) \in R^{\mathcal{G}}$ if and only if $\varphi_R(f(\bar{a})) = \bot$ holds.

□

Corollary 48 (of Lemma 47)**.** *If α is a* BMWB*-formula over the signature σ, we can write a* BMWB *formula α' over the signature $\{S\} \cup \mathsf{N_C} \cup \mathsf{N_R} \cup \{\sim, P_1, P_n\}$ such that $\mathcal{G}_T \models \alpha$ if and only if $T^{\times k} \models \alpha'$.*

Sketch of proof. Since $\mathcal{G}_T$ is FO-interpretable in $T^{\times k}$, and since FO is a fragment of BMWB, we can easily obtain the formula α' from α. This is done by replacing any occurrence of an atomic relation $R(a_1, \ldots, a_t)$ by the formula $\varphi_R(a_1, \ldots, a_t)$ defined in the proof of Lemma 47. □

We are now ready to give the proof of our main result.

Proof of Theorem 34. Let C and $\mathcal{T}$ be an $\mathcal{ALC}^P(\mathcal{D})$-concept and TBox respectively. Let $n = d \cdot \#_E(\mathcal{T}, C)$ where d is the maximum depth of all constraints that appear in $\mathsf{Sub}(\mathcal{T}, C)$. Due to Lemma 27 we can assume without loss of generality, that C and $\mathcal{T}$ are in constraint normal form. By Theorem 40, we have to check, whether there is an ordinary n-tree interpretation $\mathcal{I}$ such that

$$\mathcal{I} \models_{\mathcal{T}_a} C_a \text{ and } \mathcal{G}_{\mathcal{I}} \preceq \mathcal{D} \; .$$

Let $\tau \subseteq \sigma$ be the finite subsignature consisting of all relation symbols that occur in C and $\mathcal{T}$. Note that $\mathcal{G}_{\mathcal{I}}$ is actually a countable τ-structure. Since the concrete domain $\mathcal{D}$ has the property EHD(BMWB), one can compute from τ a BMWB-sentence α such that for every countable τ-structure $\mathcal{B}$ we have $\mathcal{B} \models \alpha$ if and only if $\mathcal{B} \preceq \mathcal{D}$. Our new goal is to decide whether there is an ordinary n-tree interpretation $\mathcal{I}$ such that

$$\mathcal{I} \models_{\mathcal{T}_a} C_a \text{ and } \mathcal{G}_{\mathcal{I}} \models \alpha \; . \tag{7}$$

Now $\mathcal{T}_a$ and C_a are ordinary $\mathcal{ALC}$-concepts. We can then use Lemma 43, Lemma 44 and Remark 42, and obtain a FO formula φ such that: If C_a is satisfied with respect to $\mathcal{T}_a$ by some n-tree interpretation $\mathcal{I}$, then φ is satisfied by an n-tree T such that $\mathcal{G}_{\mathcal{I}} = \mathcal{G}_T$. Also, if φ is satisfied by some n-tree T, then there exists an n-tree interpretation $\mathcal{I}$ such that $\mathcal{I} \models_{\mathcal{T}_a} C_a$ and such that $\mathcal{G}_T = \mathcal{G}_{\mathcal{I}}$.

Then finding $\mathcal{I}$ such that (7) holds is equivalent to finding an n-tree T such that

$$T \models \varphi \text{ and } \mathcal{G}_T \models \alpha \; .$$

By Corollary 48, we can find a BMWB-formula β such that $\mathcal{G}_T \models \alpha$ if and only if $T^{\times k} \models \beta$. But we also know, due to Proposition 46, that we can compute a formula β^k such that $T^{\times k} \models \beta$ if and only if $T \models \beta^k$. At this point we have to check whether there exists an n-tree T such that

$$T \models \varphi \wedge \beta^k \; ,$$

where $\psi \wedge \beta^k$ is a BMWB-sentence. By Theorem 10 this is decidable, which completes the proof.

□

5 Undefined concrete features

In the original definition of $\mathcal{ALC}$ with concrete domains from [1], concrete features do not need to be defined for each individual of the interpretation: the valuation function γ is a partial function, assigning values from the concrete domain to some - not necessarily all! - of the pairs (v, x) where v is an individual and x is a concrete feature (or register variable).

For instance, a concrete feature *boarding_priority* could be defined only for the passengers of a specific airline. In this case, writing $\forall(\mathit{boarding\_priority} = 1) \sqsubseteq \mathsf{Board\_First}$ means that if the concrete feature *boarding_priority* is defined for an individual, then the fact that it has value 1 implies that the individual will board first. The same will happen if *boarding_priority* is undefined for this individual (because the universal quantifier is trivially satisfied) but not if it is defined and it holds a value different than 1.

In our framework, instead, the valuation function needs to be defined for every pair (v, x), more in the flavor of the *attributes* used by Toman and Weddell (see [17]). The difference is in our case not crucial. One could define a dummy value, for instance 1.000, and use it for all the passengers of an airline that does not assign boarding priority. For instance we could write $\forall\mathsf{flies\_with}.\neg\mathsf{AIR1} \sqsubseteq (\mathit{boarding\_priority} = 1.000)$ to make sure that all the individuals that do not fly with AIR1 (or do not fly at all) are assigned the dummy boarding priority.

In some situations, though, it is not clear whether a concrete feature is bounded, and which values it can or cannot assume. Think for example of a tax identification number, or a numerical id assigned to all participants of a summer camp. One should then identify a specific dummy value to use for each concrete feature, and tailor a solution for the specific situation. It could be then interesting to add a default dummy value to our concrete domain in the following way: Instead of using $(\mathbb{Z}, <, =, (=_a)_{a\in\mathbb{Z}}, (\equiv_{a,b})_{a<b\in\mathbb{Z}})$, we consider the domain $\mathcal{Z}_u = (\mathbb{Z} \cup \{u\}, \mathsf{und}, <, =, (=_a)_{a\in\mathbb{Z}}, (\equiv_{a,b})_{a<b\in\mathbb{Z}})$, where und is a unary predicate which only holds for the freshly introduced domain element u, while the other relations remain unchanged and do not involve u. Notice that the $\mathcal{Z}_u$ still negation-closed, since the complement of und can be defined as $\{x \mid x = 0 \vee x > 0 \vee x < 0\}$, let us write def instead of $\neg\mathsf{und}$ to increase readability. Using $\mathcal{Z}_u$ as concrete domain, one can leave the possibility for a concrete feature to be undefined. Consider the following TBox: $\{\mathsf{Staff} \sqsubseteq \mathsf{und}(\mathit{camp\_id}), \mathsf{Participant} \sqsubseteq \mathsf{def}(\mathit{camp\_id}), (0 < \mathit{camp\_id} < 200) \equiv (\mathit{lunch\_turn} = 1), (200 < \mathit{camp\_id} < 400) \equiv (\mathit{lunch\_turn} = 2)\}$. This simple ontology regulates the lunch breaks at a summer camp, where all participants are assigned a *camp_id*, but the staff member are not. All individuals with camp id between 0 and 199 (which excludes the staff members) have the first lunch turn, while the ones with camp id between 200 and 400 have the second lunch turn.

To ensure that $\mathcal{ALC}^P(\mathcal{Z}_u)$ has a decidable satisfiability problem we can prove the following:

Theorem 49. *$\mathcal{Z}_u$ has the* EHD*-property.*

Proof. Let τ be a finite subsignature of $\{\mathsf{und}, <, =, (=_a)_{a\in\mathbb{Z}}, (\equiv_{a,b})_{a<b\in\mathbb{Z}}\}$ and let $\mathcal{A}$ be a τ-structure with domain A. Let us refer with $\mathcal{Z}_+$ to the structure $(\mathbb{Z}, <, =, (=_a)_{a\in\mathbb{Z}}, (\equiv_{a,b})_{a<b\in\mathbb{Z}})$. Define $U = \{x \mid \exists y.\, \mathsf{reach}_=(x, y) \wedge \mathsf{und}(y)\}$, where $\mathsf{reach}_=$ is the MSO formula expressing reachability through $=$-edges (see Ex. 8). Our claim is the following: $\mathcal{A} \preceq \mathcal{Z}_u$ if and only if

(1) $\mathcal{A}_{|A\setminus U} \preceq \mathcal{Z}_+$,

(2) $(U \times A) \cap (<^{\mathcal{A}} \cup (<^{\mathcal{A}})^{-1}) = \emptyset$,

(3) $(U \cap (=_a)^{\mathcal{A}}) = \emptyset$ for all $=_a$ in τ, and

(4) $(U \cap (\equiv_{a,b})^{\mathcal{A}}) = \emptyset$ for all $=_{a,b}$ in τ.

Facts (2)-(4) are easily expressible in MSO. For instance (2) can be expressed as $x \in U \rightarrow \neg\exists y.(x < y \vee y < x)$, where $x \in U$ is defined as $\exists y.\, \mathsf{reach}_=(x, y) \wedge \mathsf{und}(y)$. Regarding (1), in

[7] we proved that $\mathcal{Z}_+$ has the EHD-property. Therefore, we can compute a BMWB-formula ψ such that $\mathcal{A} \preceq \mathcal{Z}_+$ if and only if $\mathcal{A} \models \psi$. Now, since U is MSO-definable in $\mathcal{A}$, we can easily compute a formula ψ' such that $\mathcal{A} \models \psi'$ if and only if $\mathcal{A}_{|A\setminus U} \models \psi$ if and only if $\mathcal{A}_{|A\setminus U} \preceq \mathcal{Z}_+$.

We just need to prove that the claim holds. Let us start with the *if* direction. Suppose (1)-(4) hold. Due to (1) we can find a homomorphism h from $\mathcal{A}_{|A\setminus U}$ to $\mathcal{Z}_+$. We define $h' : A \to \mathbb{Z} \cup \{u\}$ as $h'(x) = u$ for all $x \in U$ and $h' = h$ otherwise, and we want to prove that this is a homomorphism from $\mathcal{A}$ to $\mathcal{Z}_u$. Consider a pair $(a,b) \in <^{\mathcal{A}}$. Due to (2), we know that $a, b \in (A \setminus U)$, therefore $(h'(a), h'(b)) = (h(a), h(b)) \in <^{\mathcal{Z}_+}$ and since $<^{\mathcal{Z}_+} = <^{\mathcal{Z}_u}$, this is what we wanted. The same kind of reasoning can be applied to the relations $=_a$ and $\equiv_{b,c}$ that belong to τ. Now suppose that $(a,b) \in (=^{\mathcal{A}})$, if $a, b \in A \setminus U$, then $h'(a) = h(a) = h(b) = h'(b)$. On the other hand, if $a \in U$, then the definition of U guarantees that $b \in U$ as well. We have then $h'(a) = u = h'(b)$. Finally, if $a \in \mathsf{und}^{\mathcal{A}}$, we know that $a \in U$ and therefore $h'(a) = u$, as wanted.

For the *only if* direction, assume that there exists a homomorphism h from $\mathcal{A}$ to $\mathcal{Z}_u$, we want to show that (1)-(4) hold. First of all we prove that for all $a \in U$, $h(a) = u$. Suppose $a \in U$, then there exists $b \in U$ that is reachable from a through $=$-edges, such that $b \in \mathsf{und}^{\mathcal{A}}$. There must exist then $a = a_0, \ldots, a_n = b$ such that $(a_{i-1}, a_i) \in =^{\mathcal{A}}$ for $i = 1, \ldots, n$. Then we know that, since h is a homomorphism, $h(a) = h(b) = u$, as wanted. We know then that $h(A \setminus U) \subseteq \mathbb{Z}$, and that all relations are preserved by h. Therefore h is a homomorphism from $\mathcal{A}_{|A\setminus U}$ to $\mathcal{Z}_+$, i.e. (1) holds. To prove (2) we work towards a contradiction: Suppose w.l.o.g. $(a,b) \in (U \times A) \cap <^{\mathcal{A}}$, then $h(a) < h(b)$ in $\mathbb{Z}$, but $h(a) = u$, and $(u, x) \in <^{\mathcal{Z}_u}$ is false for all $x \in A$. The proofs of (3) and (4) use the same reasoning. □

6 Conclusions

We have introduced a novel way to integrate concrete domains in $\mathcal{ALC}$, via *path constraints*. The resulting logic, $\mathcal{ALC}^P(\mathcal{D})$, is of incomparable expressiveness with the several variants of $\mathcal{ALC}(\mathcal{D})$ that are present in the literature. We have seen, however, how on the domains that we are interested in, our logic is strictly more expressive: We allow not only feature-paths, but also full role-paths, to connect abstract individuals and their concrete attributes.

We exploit the path-structure of the constraints to show that $\mathcal{ALC}^P(\mathcal{D})$ is compatible with the EHD-method from [6] and show the very general result: satisfiability for $\mathcal{ALC}^P(\mathcal{D})$ is decidable w.r.t. general TBoxes, if the concrete domain $\mathcal{D}$ is negation-closed and has the EHD-property. This solves the problem that has been open for some time (see [12]), whether reasoning in $\mathcal{ALC}$ with non-dense concrete domains such as the natural numbers or the integers would be decidable in the presence of general TBoxes, since these domains enjoy our required properties. Such domains did not satisfy the ω-admissibility criterion that was formulated in [15]. In this sense, we prove that ω-admissibility is not a necessary condition to guarantee the decidability of reasoning over a concrete domain in the presence of general TBoxes.

We could have easily chosen a more expressive DL than $\mathcal{ALC}$ as underlying logic. In principle we could add any concept constructor preserving the tree model property, and that can be then translated to MSO over trees with one successor and unary predicates only (see Lemma 43). Examples of such constructors would be transitive roles, role hierarchy and qualified number restriction.

The main open question remains the complexity. The EHD method reduces our problem to satisfiability of WMSO+B, which is decidable [3]. Here the authors do not provide complexity bounds for their decision procedure for the logic. On the other hand, the WMSO+B-formulas that need to be checked for decidability are fixed and depend solely on the concrete domain. Roughly speaking, once we fix our domain $\mathcal{D}$, the EHD method transforms a given $\mathcal{ALC}^P(\mathcal{D})$-TBox and -concept into *constraint normal form* which already blows up the size. This in turn get transformed into an MSO-formula φ (which is clearly non optimal). We then have to decide whether a conjunction of φ and a fixed $\mathsf{WMSO{+}B}$-formula ψ (which depends on $\mathcal{D}$) is decidable. Analyzing this procedure would very hardy lead to tight complexity bounds. In our opinion the EHD-method is more of an admissibility criterion, which provides easy conditions on a concrete domain $\mathcal{D}$ to establish whether reasoning with it remains decidable or not.

Also, it would be interesting to know if one can add constant predicates of the form $(=_q)_{q\in\mathbb{Q}}$ to the domain $\mathcal{Q}$ from Prop. 32 and prove that the resulting structure still has the EHD-property. We conjecture that a method similar to the one presented in [7] for constant predicates over the integers could apply to this case.

Another follow-up question is whether the EHD method can be adapted to show decidability for *fuzzy* concrete domains. It was shown in [16] that ω-admissibility remains a sufficient condition for decidability of satisfiability even if predicate membership is given by a membership degree. It would be interesting to show a similar result for the EHD-property.

References

[1] F. Baader and P. Hanschke. A scheme for integrating concrete domains into concept languages. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence, IJCAI-91*, pages 452–457, Sydney (Australia), 1991.

[2] F. Baader, I. Horrocks, and U. Sattler. Description logics. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 135–179. Elsevier, 2007.

[3] M. Bojańczyk and S. Toruńczyk. Weak MSO+U over infinite trees. In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, volume 14 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 648–660, Dagstuhl, Germany, 2012. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[4] Claudia Carapelle. *On the satisfiability of temporal logics with concrete domains*. PhD thesis, Universität Leipzig, 2015. http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-190987.

[5] Claudia Carapelle, Shiguang Feng, Alexander Kartzow, and Markus Lohrey. Satisfiability of ECTL* with Tree Constraints. In Lev D. Beklemishev and Daniil V. Musatov, editors, *Computer Science – Theory and Applications*, volume 9139 of *Lecture Notes in Computer Science*, pages 94–108. Springer International Publishing, 2015.

[6] Claudia Carapelle, Alexander Kartzow, and Markus Lohrey. Satisfiability of CTL* with Constraints. In Pedro R. D'Argenio and Hernán C. Melgratti, editors, *CONCUR 2013 Concurrency Theory*, volume 8052 of *Lecture Notes in Computer Science*, pages 455–469. Springer Berlin Heidelberg, 2013.

[7] Claudia Carapelle, Alexander Kartzow, and Markus Lohrey. Satisfiability of ECTL* with Constraints. *Journal of Computer and System Sciences*, 2016. To appear.

[8] Volker Haarslev, Ralf Möller, and Michael Wessel. The description logic $\mathcal{ALCNH}_{R+}$ extended with concrete domains: A practically motivated approach. In R. Goré, A. Leitsch, and T. Nipkow, editors, *In Proceedings of the First International Joint Conference on Automated Reasoning, IJCAR*, volume 2083 of *Lecture Notes in Computer Science*, pages 29–44. Springer, 2001.

[9] Ian Horrocks and Ulrike Sattler. Ontology reasoning in the $\mathcal{SHOQ}(\mathcal{D})$ description logic. In B. Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI*, pages 199–204, 2001.

[10] C. Lutz. Nexptime-complete description logics with concrete domains. LTCS-Report LTCS-00-01, LuFG Theoretical Computer Science, RWTH Aachen, Germany, 2000. See http://www-lti.informatik.rwth-aachen.de/Forschung/Reports.html.

[11] C. Lutz. Adding numbers to the $\mathcal{SHIQ}$ description logic—First results. In *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR'02)*, pages 191–202. Morgan Kaufman, 2002.

[12] C. Lutz. Description logics with concrete domains—a survey. In *Advances in Modal Logic 2002 (AiML 2002)*, Toulouse, France, 2002. Final version appeared in Advanced in Modal Logic Volume 4, 2003.

[13] C. Lutz. Combining interval-based temporal reasoning with general tboxes. *Artificial Intelligence*, 152(2):235–274, 2004.

[14] Carsten Lutz. NExpTime-complete description logics with concrete domains. *ACM Transactions on Computational Logic*, 5(4):669–705, 2004.

[15] Carsten Lutz and Maja Miličić. A tableau algorithm for description logics with concrete domains and general TBoxes. *Journal of Automated Reasoning*, 38(1-3):227–259, 2007.

[16] Dorian Merz, Rafael Peñaloza, and Anni-Yasmin Turhan. Reasoning in $\mathcal{ALC}$ with fuzzy concrete domains. In Carsten Lutz and Michael Thielscher, editors, *Proceedings of 37th edition of the German Conference on Artificial Intelligence (KI'14)*, volume 8736 of *Lecture Notes in Artificial Intelligence*, pages 171–182. Springer Verlag, 2014.

[17] David Toman and Grant E. Weddell. Applications and extensions of PTIME description logics with functional constraints. In *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 948–954, 2009.

[18] W3C OWL Working Group. OWL 2 web ontology language document overview. W3C Recommendation, 27th October 2009. http://www.w3.org/TR/2009/REC-owl2-overview-20091027/.