Finite Entailment of Local Queries in the $\mathcal{Z}$ Family of Description Logics

Bartosz Bednarczyk,[1,2] **Emanuel Kieroński**[2]

[1] Computational Logic Group, TU Dresden
[2] University of Wrocław, Faculty of Mathematics and Computer Science
{ bartosz.bednarczyk, emanuel.kieronski }@cs.uni.wroc.pl

Abstract

In the last few years the field of logic-based knowledge representation took a lot of inspiration from database theory. A vital example is that the finite model semantics in description logics (DLs) is reconsidered as a desirable alternative to the classical one and that query entailment has replaced knowledge-base satisfiability (KBSat) checking as the key inference problem. However, despite the considerable effort, the overall picture concerning finite query answering in DLs is still incomplete. In this work we study the complexity of finite entailment of local queries (conjunctive queries and positive boolean combinations thereof) in the $\mathcal{Z}$ family of DLs, one of the most powerful KR formalisms, lying on the verge of decidability. Our main result is that the DLs $\mathcal{ZOQ}$ and $\mathcal{ZOI}$ are finitely controllable, *i.e*. that their finite and unrestricted entailment problems for local queries coincide. This allows us to reuse recently established upper bounds on querying these logics under the classical semantics. While we will not solve finite query entailment for the third main logic in the $\mathcal{Z}$ family, $\mathcal{ZIQ}$, we provide a generic reduction from the finite entailment problem to the finite KBSat problem, working for $\mathcal{ZIQ}$ and some of its sublogics. Our proofs unify and solidify previously established results on finite satisfiability and finite query entailment for many known DLs.

1 Introduction

In the last few years the field of logic-based knowledge representation took a lot of inspiration from database theory. In particular, finite model semantics in description logics (DLs) is reconsidered as a desirable alternative to classical one and query entailment has replaced knowledge base satisfiability checking as the key inference problem (Gogacz et al. 2020). Under classical semantics, that is when arbitrary models are admitted, the conjunctive query (CQ) entailment problem for DLs is already quite well understood (Glimm et al. 2008). The situation is different if we admit only finite models. For *finite* query entailment of local queries (CQs and their positive boolean combinations, known as PEQs), despite the considerable effort, the overall picture is still incomplete.

The main goal of this paper is to contribute to clarifying this picture by considering the $\mathcal{Z}$ (a.k.a. $\mathcal{ALCH}b^{\mathsf{Self}}_{\mathsf{reg}}$) family of DLs (Calvanese, Eiter, and Ortiz 2009), one of the most powerful KR formalisms, lying on the verge of decidability (Rudolph 2016). The logics from the $\mathcal{Z}$ family incorporate many different features, the most important of which is the ability to write regular expressions as building blocks for complex roles. The most expressive member of the $\mathcal{Z}$ family is $\mathcal{ZOIQ}$, offering nominals ($\mathcal{O}$), inverse roles ($\mathcal{I}$) and number restrictions ($\mathcal{Q}$). One of the motivations behind $\mathcal{ZOIQ}$ was that another expressive DL $\mathcal{SROIQ}$ (Horrocks, Kutz, and Sattler 2006) can be reduced to $\mathcal{ZOIQ}$ (with an exponential blow-up). $\mathcal{SROIQ}$ is of practical interest, as it serves as a base for OWL 2 the W3C Web Ontology Language (Grau et al. 2008). Querying with *regular* query languages in undecidable for $\mathcal{ZOIQ}$ and the decidability of its knowledge base (KB) satisfiability problem (KBSat) is still open (Rudolph 2016). By removing one of the constructors $\mathcal{I}$, $\mathcal{Q}$, $\mathcal{O}$, we get, resp., the logics $\mathcal{ZOQ}$, $\mathcal{ZOI}$, $\mathcal{ZIQ}$ known to have better model-theoretic properties and ExpTime (resp. 2ExpTime) KBSat (resp. PEQ entailment) problems, see: (Calvanese, Eiter, and Ortiz 2009; Bednarczyk and Rudolph 2019). The proofs of these facts rely on the *quasi-forest model property*, which admits automata-theoretic approach (a reduction to the so-called fully enriched automata).

Not much is known about finite model reasoning in the $\mathcal{Z}$ family. The only result we are aware of is that $\mathcal{ZOI}$ has the finite model property (Calvanese, Ortiz, and Simkus 2016), *i.e*. that every satisfiable KB also has a finite model, and thus the unrestricted and finite satisfiability coincide. It is also easy to show that this property fails for $\mathcal{ZIQ}$ (and even for $\mathcal{ALCIF}$). Regarding finite query entailment, to the best of our knowledge, nothing is known even about $\mathcal{Z}$.

Our Contribution. Our most important result is showing that $\mathcal{ZOQ}$ and $\mathcal{ZOI}$ are *finitely controllable*, *i.e*. for any $\mathcal{ZOQ}$ or $\mathcal{ZOI}$ KB $\mathcal{K}$ and any PEQ q it holds that $\mathcal{K}$ entails q over all models iff $\mathcal{K}$ entails q over all finite models. This allows one to reuse the existing algorithms for unrestricted model semantics (Calvanese, Eiter, and Ortiz 2009; Bednarczyk and Rudolph 2019) to infer that finite entailment of local queries is 2ExpTime-complete. Note that our results cover all DLs between $\mathcal{ALC}$ and $\mathcal{ALCH}b\mathcal{O}^{\mathsf{Self}}_{\mathsf{reg}}(\mathcal{I}/\mathcal{Q})$ in a uniform way.

The proof is based on a finite model construction which starts from an (infinite) quasi-forest model, distinguishes in this model some finite pattern fragments, and carefully forms a new finite model out of some finite number of copies of those fragments. An important tool used in this process are *automaton roles*, a concept applied earlier in the context of PDL and modal logics, see (Nguyen 2020, Sec. 2.2).

We do not solve the challenging problem of finite PEQ entailment for $\mathcal{ZIQ}$ here. However, we demonstrate that this problem can be effectively reduced to finite KBSat for $\mathcal{ZIQ}$, making the finite KBSat *the* main problem to study. We apply our methods to derive new results for extensions of $\mathcal{ALC}$ with transitive closure of roles (Jung, Lutz, and Zeume 2020) and to lift some known results about finite CQ entailment for various DLs to the setting of local queries.

Related Work. As we already mentioned, the topic of finite CQ over DL ontologies has started receiving more attention only recently. On the positive side, the decidability of finite CQ entailment in Horn DLs was shown in (Ibáñez-García, Lutz, and Schneider 2014). Another positive example is the series of papers on CQ entailment for the $\mathcal{S}$ family of logics (Gogacz, Ibáñez-García, and Murlak 2018; Gogacz et al. 2019; Danielski and Kieronski 2019). On the negative side (Rudolph 2016) proves undecidability of finite querying in $\mathcal{SHOIQ}$. Worth mentioning are also some related works going beyond DLs, in particular admitting relations of arity greater than 2: the guarded fragment, which captures DLs up to $\mathcal{ALCIH}b^{\mathsf{Self}}$, was shown to be finitely controllable in (Bárány, Gottlob, and Otto 2014), finite model reasoning over existential rules is studied in (Gogacz and Marcinkowski 2017; Gottlob, Manna, and Pieris 2018; Amendola, Leone, and Manna 2017), and finite satisfiability of the unary negation fragment (in which one can directly express CQs) with transitive relations in (Danielski and Kieronski 2019).

2 Preliminaries

In this section we provide basic definitions concerning the $\mathcal{Z}$ (a.k.a. $\mathcal{ALCH}b^{\mathsf{Self}}_{\mathsf{reg}}$) family of DLs and the query entailment.

2.1 The $\mathcal{Z}$ Family of Description Logics

First we briefly describe the syntax and semantics of the very expressive DL $\mathcal{ZOIQ}$ and its relevant sublogics (Calvanese, Eiter, and Ortiz 2009). For brevity and due to the space limit we present $\mathcal{ZOIQ}$-KBs in the simple normal forms, involving in particular automaton roles instead of regular expressions used originally. They are polynomially computable from original $\mathcal{ZOIQ}$-KBs (cf. *e.g.* (Jung, Lutz, and Zeume 2019, p.14)).

Fix infinite and mutually-disjoint sets $\mathbf{N_I}$, $\mathbf{N_C}$, $\mathbf{N_R}$, of *individual names*, *concept names*, and *role names*. The following grammar defines *atomic concepts* B, *concepts* C, *atomic roles* r, *simple roles* s, with $\mathrm{o} \in \mathbf{N_I}$, $\mathrm{A} \in \mathbf{N_C}$, $p \in \mathbf{N_R}$ and a nondeterministic automata (NFA) $\mathbb{A}$ over the alphabet $\Sigma = \{p, p^- \mid p \in \mathbf{N_R}\} \cup \{id(\mathrm{B}) \mid \mathrm{B} \in \mathbf{N_C}\}$:

$$\begin{aligned} \mathrm{B} &::= \mathrm{A} \mid \top \mid \bot \mid \{\mathrm{o}\} \\ \mathrm{C} &::= \mathrm{B} \mid \neg\mathrm{C} \mid \mathrm{C}\sqcap\mathrm{C} \mid \mathrm{C}\sqcup\mathrm{C} \mid \forall s.\mathrm{C} \mid \exists s.\mathrm{C} \mid \\ &\qquad \geqslant n s.\mathrm{C} \mid \leqslant n s.\mathrm{C} \mid \exists s.\mathsf{Self} \mid \exists p_{\mathbb{A}}.\mathrm{C} \mid \forall p_{\mathbb{A}}.\mathrm{C} \\ r &::= p \mid p^- \\ s &::= r \mid s\cap s \mid s\cup s \mid s\setminus s \end{aligned}$$

The $p_{\mathbb{A}}$ in the above grammar are *automaton roles*. Given an automaton $\mathbb{A}$ and its state $\mathtt{s}$ we denote by $\mathbb{A}_{\mathtt{s}}$ the automaton obtained from $\mathbb{A}$ by changing its initial state to $\mathtt{s}$. An *assertion* is of the form $\mathrm{C}(\mathrm{o})$, $r(\mathrm{o}_1,\mathrm{o}_2)$, $\neg r(\mathrm{o}_1,\mathrm{o}_2)$ for $\mathrm{C}\in\mathbf{N_C}$ and $\mathrm{o},\mathrm{o}_1,\mathrm{o}_2\in\mathbf{N_I}$. A *general concept inclusion* (GCI) has the form $\mathrm{C}_1 \sqsubseteq \mathrm{C}_2$. We use $\mathrm{C}_1 \equiv \mathrm{C}_2$ in place of two GCIs $\mathrm{C}_1 \sqsubseteq \mathrm{C}_2$ and $\mathrm{C}_2 \sqsubseteq \mathrm{C}_1$. A $\mathcal{ZOIQ}$ *knowledge base* (KB) $\mathcal{K} = (\mathcal{A},\mathcal{T})$ consists of a finite set $\mathcal{A}$ (called *ABox*) of assertions and a finite set $\mathcal{T}$ (called *TBox*) of GCIs of the form:

$$\mathrm{A}\equiv\{\mathrm{o}\}, \qquad \mathrm{A}\equiv\neg\mathrm{B}, \qquad \mathrm{A}\equiv\mathrm{B}\sqcup\mathrm{B}',$$
$$\mathrm{A}\equiv\geqslant ns.\mathrm{B}, \qquad \mathrm{A}\equiv\exists p_{\mathbb{A}}.\mathrm{B}, \qquad \mathrm{A}\equiv\exists s.\mathsf{Self}$$

where $\mathrm{A},\mathrm{B},\mathrm{B}'$ are concept names, s is a simple role, $p_{\mathbb{A}}$ is an automaton role, and for every definition $\mathrm{A}\equiv\exists p_{\mathbb{A}}.\mathrm{B}$ in $\mathcal{T}$ and for every state $\mathtt{s}$ of $\mathbb{A}$ we have that $\mathcal{T}$ contains also $\mathrm{A}_{\mathtt{s}}\equiv\exists p_{\mathbb{A}_{\mathtt{s}}}.\mathrm{B}$ for some fresh concept name $\mathrm{A}_{\mathtt{s}}$.

The semantics of $\mathcal{ZOIQ}$ is defined via *interpretations* $\mathcal{I}=(\Delta^{\mathcal{I}},\cdot^{\mathcal{I}})$ composed of a non-empty set $\Delta^{\mathcal{I}}$ called the *domain of* $\mathcal{I}$ and a function $\cdot^{\mathcal{I}}$ mapping individual names to elements of $\Delta^{\mathcal{I}}$ (we call $\mathrm{o}^{\mathcal{I}}$ for any $\mathrm{o}\in\mathbf{N_I}$ *nominal elements*), concept names to subsets of $\Delta^{\mathcal{I}}$, and role names to subsets of $\Delta^{\mathcal{I}}\times\Delta^{\mathcal{I}}$. A structure is an interpretation with a partial assignment of names. This mapping is extended to concepts, and to simple and automaton roles (we omit the cases for concept intersection, universal restriction, $\leqslant n\ s.\mathrm{C}$, and role union/difference).

Name	Syntax	Semantics
top/bottom	$\top/\bot$	$\Delta^{\mathcal{I}}/\emptyset$
nominal	$\{o\}$	$\{o^{\mathcal{I}}\}$
negation	$\neg\mathrm{C}$	$\Delta^{\mathcal{I}}\setminus\mathrm{C}^{\mathcal{I}}$
concept union	$\mathrm{C}_1\sqcup\mathrm{C}_2$	$\mathrm{C}_1^{\mathcal{I}}\cup\mathrm{C}_2^{\mathcal{I}}$
existential restr.	$\exists t.\mathrm{C}$	$\{\mathrm{d}\mid\exists\mathrm{e}.(\mathrm{d},\mathrm{e})\in t^{\mathcal{I}}\wedge\mathrm{e}\in\mathrm{C}^{\mathcal{I}}\}$
number restr.	$\geqslant n\ s.\mathrm{C}$	$\{\mathrm{d}\mid\#\{y\in\mathrm{C}^{\mathcal{I}}\mid(\mathrm{d},\mathrm{e})\in s^{\mathcal{I}}\}\geq n\}$
Self concept	$\exists s.\mathsf{Self}$	$\{\mathrm{d}\mid(\mathrm{d},\mathrm{d})\in s^{\mathcal{I}}\}$
inverse role	r^-	$\{(\mathrm{d},\mathrm{e})\mid(\mathrm{e},\mathrm{d})\in r^{\mathcal{I}}\}$
role intersection	$s_1\cap s_2$	$s_1^{\mathcal{I}}\cap s_2^{\mathcal{I}}$
automaton $\mathbb{A}$-role	$p_{\mathbb{A}}$	$\{(\mathrm{d},\mathrm{e})\mid\exists\text{path }\mathrm{d}\rightsquigarrow\mathrm{e}\text{ matching }\mathbb{A}\}$

To define the semantics of automaton roles we first introduce a handy notion of paths *matching* an automaton. We say that a path $\mathrm{d}=\mathrm{d}_1,\mathrm{d}_2,\ldots,\mathrm{d}_k=\mathrm{e}$ between d and e in an interpretation $\mathcal{I}$ *matches* an automaton $\mathbb{A}$ if there exist $\sigma_1,\sigma_2,\ldots,\sigma_{k-1}\in\Sigma$ satisfying, for all indices $1\leq i<k$, either (i) $\sigma_i=r$ for some (possibly inverted) role r and $(\mathrm{d}_i,\mathrm{d}_{i+1})\in\sigma_i^{\mathcal{I}}$ or (ii) $\sigma_i=id(\mathrm{B})$ for some concept name B, $\mathrm{d}_i=\mathrm{d}_{i+1}$, and $d_i\in\mathrm{B}^{\mathcal{I}}$, such that the word $\sigma_1\sigma_2\ldots\sigma_{k-1}$ is accepted by $\mathbb{A}$. Now, with such an $\mathbb{A}$ we naturally associate the role $p_{\mathbb{A}}$ interpreted as the set of all pairs of domain elements (d,e) with a path from d to e matching $\mathbb{A}$.

Next, we define *satisfaction* of assertions and GCIs.

Axiom α	$\mathcal{I}\models\alpha$, if	
$\mathrm{C}\sqsubseteq\mathrm{D}$	$\mathrm{C}^{\mathcal{I}}\subseteq\mathrm{D}^{\mathcal{I}}$	TBox $\mathcal{T}$
$\mathrm{C}(\mathrm{o})$	$\mathrm{o}^{\mathcal{I}}\in\mathrm{C}^{\mathcal{I}}$	ABox $\mathcal{A}$
$r(\mathrm{o}_1,\mathrm{o}_2)$	$(\mathrm{o}_1^{\mathcal{I}},\mathrm{o}_2^{\mathcal{I}})\in r^{\mathcal{I}}$	
$\neg r(\mathrm{o}_1,\mathrm{o}_2)$	$(\mathrm{o}_1^{\mathcal{I}},\mathrm{o}_2^{\mathcal{I}})\notin r^{\mathcal{I}}$	

We say that an interpretation $\mathcal{I}$ *satisfies* a KB $\mathcal{K}=(\mathcal{A},\mathcal{T})$ (or $\mathcal{I}$ is a *model* of $\mathcal{K}$, written: $\mathcal{I}\models\mathcal{K}$) if it satisfies all assertions in $\mathcal{A}$ and all axioms of $\mathcal{T}$. A KB $\mathcal{K}$ is called *(finitely)* *satisfiable* if it has a (finite) model and *(finitely) unsatisfiable* otherwise. From $\mathcal{ZOIQ}$, we obtain $\mathcal{ZIQ}$ by disallowing nominal concepts $\{o\}$, $\mathcal{ZOQ}$, by disallowing role inverses $(\cdot)^-$ (also in the alphabet of automaton roles) and $\mathcal{ZOI}$ by disallowing numb. restrictions $(\geqslant ns.\mathrm{C})$, $(\leqslant ns.\mathrm{C})$ for $n{\neq}1$. Finally, $\mathsf{ind}(\mathcal{K})$ is the set of all individual names from $\mathcal{K}$.

Given $\mathcal{I}$ and $\Delta_0 \subseteq \Delta^{\mathcal{I}}$ we denote by $\mathcal{I}\restriction\Delta_0$ the *restriction* of $\mathcal{I}$ to Δ_0, that is the structure with the domain Δ_0, mapping each concept name A to $\Delta_0 \cap \mathrm{A}^{\mathcal{I}}$, each role name p to $p^{\mathcal{I}} \cap \Delta_0 \times \Delta_0$, and mapping each individual name o to $\mathrm{o}^{\mathcal{I}}$ if $\mathrm{o}^{\mathcal{I}} \in \Delta_0$, leaving $\mathrm{o}^{\mathcal{I}}$ undefined otherwise. For $\mathrm{d} \in \Delta^{\mathcal{I}}$ we define its *atomic type* as the isomorphism type[1] of $\mathcal{I}\restriction(\{\mathrm{d}\} \cup \mathsf{Nom}_{\mathcal{I}})$, where $\mathsf{Nom}_{\mathcal{I}} = \{\mathrm{o}^{\mathcal{I}} : \mathrm{o} \in \mathbf{N_I}\}$.
The *(finite) KB satisfiability problem* (short: (finite) KBSat) for a DL $\mathcal{L}$ is the problem of deciding if a given $\mathcal{L}$-KB is (finitely) satisfiable. If every satisfiable $\mathcal{L}$-KB has a finite model then we say that $\mathcal{L}$ has the *finite model property* (short: fmp). For $\mathcal{L}$ with fmp, the KBSat and finite KBSat coincide.

2.2 Queries

Local queries (known as PEQs) are boolean queries generated with the grammar $q, q' ::= r(x, y) \mid \mathrm{A}(z) \mid q \vee q' \mid q \wedge q'$, where $r \in \mathbf{N_R}$, $\mathrm{A} \in \mathbf{N_C}$ and x, y, z are variables from some countably infinite set $\mathbf{N_V}$. PEQs without disjunction are called conjunctive queries (CQs) and the disjunctions of CQs are called UCQs. Note that every PEQ can be converted into an equivalent UCQ (of possibly exponential size). We use $|q|$ to denote the number of atoms in q, and $\mathrm{Var}(q)$ to denote the set of its variables. We define entailment of PEQs in the usual way (Ortiz and Simkus 2012), *i.e.* $\mathcal{I} \models q$ if there is a variable assignment $\eta : \mathrm{Var}(q) \to \Delta^{\mathcal{I}}$ such that q evaluates to true under η.

We say that a PEQ q is *(finitely) entailed* by a KB $\mathcal{K}$, written: $\mathcal{K} \models_{(\mathrm{fin})} q$, if every (finite) model of $\mathcal{K}$ satisfies q. When $\mathcal{I} \models \mathcal{K}$ but $\mathcal{I} \not\models q$, we call $\mathcal{I}$ a *countermodel* for $\mathcal{K}$ and q; if such $\mathcal{I}$ is finite, we call it a *finite countermodel*. The *(finite) query entailment* problem for a logic $\mathcal{L}$ is defined as follows: given $\mathcal{L}$ KB $\mathcal{K}$ and a PEQ q verify if $\mathcal{K}$ (finitely) entails q. A DL $\mathcal{L}$ is *finitely controllable* (short: fc), if for every $\mathcal{K}$ and every PEQ q if there is a countermodel for $\mathcal{K}$ and q then there is also a finite one. For $\mathcal{L}$ with fc, the PEQ entailment and the finite PEQ entailment coincide. Fc implies fmp but not vice versa. Cf. (Gutiérrez-Basulto, Ibáñez-García, and Jung 2017).

Conjunctive queries themselves may be seen as structures: for a query q we define the structure $\mathcal{I}_q = (\mathrm{Var}(q), \cdot^{\mathcal{I}_q})$ satisfying $(x, y) \in r^{\mathcal{I}_q}$ iff $r(x, y) \in q$ and $x \in \mathrm{A}^{\mathcal{I}_q}$ iff $\mathrm{A}(x) \in q$. CQ is tree-shaped if its query structure is a tree. A *homomorphism* from $\mathcal{I}$ to $\mathcal{J}$ is a function that maps every element of $\Delta^{\mathcal{I}}$ to some element from $\Delta^{\mathcal{J}}$ and it preserves concept and role names. Since CQs can be seen as structures, their matches can be seen as homomorphisms.

2.3 Quasi-Forest Models

In the analysis of satisfiability and the query entailment problem for sublogics of $\mathcal{ZOIQ}$ a crucial role is played by the following model-theoretic notion of a *quasi-forest model*.

[1] Structures have equal isomorphism type iff are isomorphic.

Definition 2.1 ((Calvanese, Eiter, and Ortiz 2009))**.** *Let $\mathcal{K}$ be a KB. A model $\mathcal{I}$ of $\mathcal{K}$ is a* quasi-forest model *if:*

- *its domain $\Delta^{\mathcal{I}}$ is a forest,* i.e. *a non-empty prefix-closed subset of* nodes *from $(\mathbb{N}_+)^+$ (where $\mathbb{N}_+$ denotes the set of positive integers and the superscript $^+$ indicates that we consider sequences of positive length),*
- *the set of* roots *of $\mathcal{I}$, denoted with $Roots_{\mathcal{I}} = \Delta^{\mathcal{I}} \cap \mathbb{N}_+$ is finite and equal to $\{\mathrm{o}^{\mathcal{I}} \mid \mathrm{o} \in \mathbf{N_I}\} = \{\mathrm{o}^{\mathcal{I}} \mid \mathrm{o} \in \mathsf{ind}(\mathcal{K})\}$,*
- *and for every $\mathrm{d}, \mathrm{e} \in \Delta^{\mathcal{I}}$ with $(\mathrm{d}, \mathrm{e}) \in p^{\mathcal{I}}$, for some atomic role p, either (i) $\{\mathrm{d}, \mathrm{e}\} \cap Roots_{\mathcal{I}} \neq \emptyset$, or (ii) $\mathrm{d} = \mathrm{e}$, or (iii) d is a* child *of* e, i.e. *it satisfies $\mathrm{d} = \mathrm{e} \cdot n$ for some $n \in \mathbb{N}_+$, or (iv) e is a child of* d.

We always assume that if $\mathrm{d}{\cdot}n$ belongs to $\Delta^{\mathcal{I}}$ for some $n \in \mathbb{N}_+$ then for all positive integers $n' < n$ we have $d{\cdot}n' \in \Delta^{\mathcal{I}}$. This is w.l.o.g., as we always can simply rename elements. Thus, we call $\mathrm{d} \cdot n$ the n-th child *of* d. *We refer to the number of children of d as its* degree. *Moreover, w.l.o.g. if a role/concept name is not present in $\mathcal{K}$, then $\mathcal{I}$ interprets it as $\emptyset$.*

A *descendant* of $\mathrm{d} \in \Delta^{\mathcal{I}}$ is any node of the form $\mathrm{d}w \in \Delta^{\mathcal{I}}$ for $w \in (\mathbb{N}_+)^+$. By $\mathit{Subtree}_{\mathcal{I}}(\mathrm{d})$ we denote the set consisting of d and its descendants. Its induced substructure is *the subtree of* d. Sometimes we also look at descendants at a certain distance from d. Hence, we denote with $\mathit{Subtree}^{\leq n}_{\mathcal{I}}(\mathrm{d})$ the subset of $\Delta^{\mathcal{I}}$ consisting of d and all of its descendants of the form $\mathrm{d}w$ for some $w \in (\mathbb{N}_+)^+$ of length at most n.

A KB $\mathcal{K}$ has the *quasi-forest model property* (short: qfmp) if $\mathcal{K}$ is either unsatisfiable or it has a quasi-forest model. A DL $\mathcal{L}$ has the qfmp if every $\mathcal{L}$-KB $\mathcal{K}$ has the qfmp. Similarly, we say that a KB $\mathcal{K}$ has *quasi-forest controllability* (with respect to PEQs; short: qfc), if for every PEQ q, if there is a countermodel for $\mathcal{K}$ and q then there is also a quasi-forest countermodel. A DL $\mathcal{L}$ has qfc if every $\mathcal{L}$-KB $\mathcal{K}$ has the qfc. Obviously, qfc implies qfmp but not vice-versa.

The following fact about the model-theoretic properties of the $\mathcal{Z}$ family of DLs is well-known:

Theorem 2.2 (Proposition 3.3 from (Calvanese, Eiter, and Ortiz 2009))**.** *The logics $\mathcal{ZOQ}$, $\mathcal{ZOI}$ and $\mathcal{ZIQ}$ have qfc.*

There are some additional properties that hold for quasi-forest (counter)models for $\mathcal{ZOI}$/$\mathcal{ZOQ}$/$\mathcal{ZIQ}$-KBs. First, we can assume that degree of each node in $\mathcal{I}$ is finite (*cf.* (Calvanese, Eiter, and Ortiz 2009, p. 719)). Second, in the case of $\mathcal{ZOQ}$ we assume that there are no backward edges in $\mathcal{I}$, except, possibly, some edges leading to the roots (in particular, for all $\mathrm{d} \cdot n, \mathrm{d} \in \Delta^{\mathcal{I}}$ non-nominal, there is no $p \in \mathbf{N_R}$ s.t. $(\mathrm{d} \cdot n, \mathrm{d}) \in p^{\mathcal{I}}$). This cannot be assumed in $\mathcal{ZOI}$. Instead, in $\mathcal{ZOI}$ we assume that whenever d needs a witness path for some concept $\exists p_{\mathbb{A}}$ it has a *downward witness path*, *i.e.* a path in which for any two consecutive nodes d, e we have that either e is a root of $\mathcal{I}$, e is a child of d or $d = e$, cf., e.g. (Oritz 2010).

Regarding the complexity of the standard reasoning problems for the $\mathcal{Z}$ family of DLs, the current state of the art is as follows (Thm. 3.11 (Calvanese, Eiter, and Ortiz 2009), Thm. 8 (Bednarczyk and Rudolph 2019)):

Theorem 2.3. *The KBSat problem and the CQ/UCQ/PEQ entailment problem for $\mathcal{ZOI}$, $\mathcal{ZOQ}$ and $\mathcal{ZIQ}$ are, respectively,* ExpTime*-complete and* 2ExpTime*-complete.*

3 Finite Controllability of $\mathcal{ZOQ}$ and $\mathcal{ZOI}$

This section is devoted to proving our main result, namely:

Theorem 3.1. *$\mathcal{ZOQ}$ and $\mathcal{ZOI}$ are finitely controllable.*

As an immediate corollary, due to Theorem 2.3, we have:

Corollary 3.2. *The finite KB satisfiability problems for $\mathcal{ZOQ}$ and $\mathcal{ZOI}$ are* ExpTime*-complete, while their finite PEQ entailment problems are* 2ExpTime*-complete (both even under binary encodings in number restrictions).*

In DB theory, one of the most prominent reasoning problems is query containment, which received also some attention from the DL community (Calvanese, Ortiz, and Simkus 2011; Bienvenu, Lutz, and Wolter 2012). Arguing along the lines of (Calvanese, Eiter, and Ortiz 2009) we can lift our findings to *finite* query containment problem. Missing definitions are in (Calvanese, Eiter, and Ortiz 2009). Up to our knowledge it is the first result on *finite containment* of regular queries in the local ones, in the DL setting.

Corollary 3.3. *For a given positive conjunctive regular path query q, a PEQ q' and a $\mathcal{ZOQ}$ / $\mathcal{ZOI}$ KB $\mathcal{K}$, it is decidable in* 2ExpTime *if $\mathcal{K} \models q \subseteq q'$ over all* finite interpretations.

We proceed with the proof of Thm. 3.1. Until the end we fix a satisfiable $\mathcal{ZOQ}$/$\mathcal{ZOI}$ KB $\mathcal{K}=(\mathcal{A}, \mathcal{T})$, a PEQ q and an interpretation $\mathcal{I}$ being a countermodel for $\mathcal{K}$ and q. W.l.o.g. we assume that q is UCQ $q = \bigvee_i q_i$ and we let K be the maximal number of variables across the q_i. We assume that $\mathcal{I}$ is a quasi-forest model of $\mathcal{K}$, possessing all the properties mentioned in Section 2.3. Moreover, we also assume that $\mathcal{K}$ has the empty ABox (*i.e.* that its ABox is internalised in its TBox $\mathcal{T}$ with nominals) and that all concept/role names from q appear also in $\mathcal{T}$ (they can be inserted to $\mathcal{T}$ in some dummy way, if necessary). We will find a finite countermodel $\mathcal{J}$ for $\mathcal{K}$ and q by distinguishing certain substructures of $\mathcal{I}$ and carefully linking together some number of their copies. The construction and its correctness proof will be nearly the same for $\mathcal{ZOQ}$ and $\mathcal{ZOI}$; we will pinpoint the minor differences.

Recall that the atomic type of $\mathrm{d} \in \Delta^{\mathcal{I}}$, denoted $atp_{\mathcal{I}}(\mathrm{d})$ is the isomorphism type of $\mathcal{I}\restriction(\{\mathrm{d}\} \cup \mathsf{Nom}_{\mathcal{I}})$. For quasi-forest models we also want to know how the structure consisting of the elements being at most K steps below d looks (recall that K is the maximal number of variables in the q_i). Formally, the *downward type* of d, written: $dtp_{\mathcal{I}}(\mathrm{d})$, is the isomorphism type of $\mathcal{I}\restriction(Subtree_{\mathcal{I}}^{\leq K}(\mathrm{d}) \cup \mathsf{Nom}_{\mathcal{I}})$. We will call the downward types of the nominal elements the *nominal (downward) types*. We denote the sets of downward and nominal types appearing in $\mathcal{I}$ with $DTP_{\mathcal{I}}$ and, resp., $NTP_{\mathcal{I}}$. The said sets are finite due to finite branching of $\mathcal{I}$ and the fact that the interpretation of roles/concept names outside $\mathcal{T}$ is empty.

3.1 Preparing Building Blocks

Our next step is to define substructures of $\mathcal{I}$ called *components* whose copies will serve as basic building blocks in the construction of $\mathcal{J}$.

For each downward type $\pi \in DTP_{\mathcal{I}}$ we fix a domain element $\mathrm{d}_\pi \in \Delta^{\mathcal{I}}$ having this downward type. We are going to select a finite subset Δ_π of $Subtree_{\mathcal{I}}(\mathrm{d}_\pi)$, which will become the domain of component $\mathcal{C}_\pi$. Δ_π will be *sibling*- and *parent-closed* (up to d_π), *i.e.* for each element from such a subset all of its siblings belong to it, and if an element d is not equal d_π then its parent belongs to the subset. We create them as follows. First, we require that $Subtree_{\mathcal{I}}^{\leq K}(\mathrm{d}_\pi)$ is contained in Δ_π. Second, for every GCI $\mathrm{A} \equiv \exists p_{\mathbb{A}}.\mathrm{B}$ from $\mathcal{K}$ with $\mathrm{A} \in \mathbf{N_C}$, whenever $\mathrm{d}_\pi \in \mathrm{A}^{\mathcal{I}}$ holds we choose a downward witness path $\rho = \mathrm{d}_1(= \mathrm{d}_\pi), \mathrm{d}_2, \ldots, \mathrm{d}_k$ for d_π and $\exists p_{\mathbb{A}}.\mathrm{B}$, with the minimal possible number of nominal elements on it. We consider two cases:

- If ρ is nominal-free, make d_k a member of Δ_π.
- Otherwise, take the smallest j such that d_j is a nominal element and make d_{j-1} a member of Δ_π.

Finally, extend all the Δ_π in a minimal way to make them sibling- and parent-closed. The *component* for the downward type π is the interpretation $\mathcal{C}_\pi := \mathcal{I}\restriction\Delta_\pi$. Components are finite trees and hence we speak about their *leaves*. However, d_π will be called the *origin* of $\mathcal{C}_\pi$, rather than its root, to avoid confusion with the roots of $\mathcal{I}$.

3.2 Assembling a Finite Model $\mathcal{J}$

The interpretation $\mathcal{J}$ will be composed of a finite number of copies of the components $\mathcal{C}_\pi$, carefully connected to preserve satisfaction of $\mathcal{K}$ and non-satisfaction of q. We will use sub- and super-scripts to distinguish such copies. As expected, a unique copy of $\mathcal{C}_\pi$ for every nominal type π from $NTP_{\mathcal{I}}$ will be used—this guarantees the uniqueness of nominals. We remark that the scheme of joining components is somehow similar to the one used recently in (Danielski and Kieronski 2019) to build small finite models for finitely-satisfiable formulae in the extension of the UNFO with transitivity.

Let L be the maximal number of leaves across all the components and let M be the maximal number of children of a node in $\mathcal{I}$ (note that L and M are finite). The domain of $\mathcal{J}$ is

$$\Delta^{\mathcal{J}} := \bigcup_{\pi \in NTP_{\mathcal{I}}} \Delta_\pi \quad \cup \quad \bigcup \Delta_{\pi,\ell,m}^{\pi',b},$$

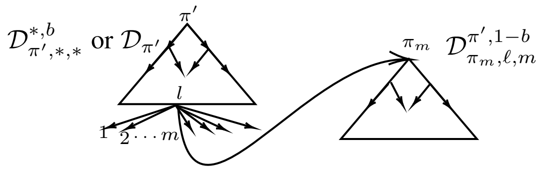
where the second union ranges over $\pi \in DTP_{\mathcal{I}} \setminus NTP_{\mathcal{I}}$, $\pi' \in DTP_{\mathcal{I}}$, $1 \leq \ell \leq L$, $1 \leq m \leq M$ and $b \in \{0,1\}$. The presence of various indices may look cryptic but it will be clarified soon, when describing the "linking process". The above sums are disjoint and the elements of the decorated $\Delta_{\pi,*,*}^{*,*}$ are just disjoint copies of the corresponding Δ_π. For every π, π', ℓ, m, b as above, we make $\mathcal{D}_{\pi,\ell,m}^{\pi',b} := \mathcal{J}\restriction\Delta_{\pi,\ell,m}^{\pi',b}$ isomorphic to $\mathcal{C}_\pi$; similarly for every $\pi \in NTP_{\mathcal{I}}$ we make $\mathcal{D}_\pi := \mathcal{J}\restriction\Delta_\pi$ isomorphic to $\mathcal{C}_\pi$.

We naturally define the *pattern function* $\mathfrak{f} : \Delta^{\mathcal{J}} \to \Delta^{\mathcal{I}}$ that maps an element from $\Delta^{\mathcal{J}}$ to the element in $\Delta^{\mathcal{I}}$ which it is a copy of. Note that $\mathfrak{f}(\mathrm{o}^{\mathcal{J}}) = \mathrm{o}^{\mathcal{I}}$ for all $\mathrm{o} \in \mathbf{N_I}$.

What remains to be done is to define the roles between the elements from different components of $\mathcal{J}$.

- Connections involving nominal elements: Roles between $\mathrm{d} \in \Delta^{\mathcal{J}}$ and $\mathrm{o}^{\mathcal{J}} \in \mathsf{Nom}_{\mathcal{J}}$ are defined according to $\mathfrak{f}$. For all atomic roles r we put $(\mathrm{d}, \mathrm{o}^{\mathcal{J}}) \in r^{\mathcal{J}}$ iff $(\mathfrak{f}(\mathrm{d}), \mathfrak{f}(\mathrm{o}^{\mathcal{J}})) \in r^{\mathcal{I}}$. This way $\mathfrak{f}$ preserves atomic types.
- Linking different components: For every component, its leaves will be connected to the origins of other components, in order to provide leaves with their local witnesses. Our strategy is as follows. For every type $\pi \in$

$DTP_{\mathcal{I}} \setminus NTP_{\mathcal{I}}$ the origin of $\mathcal{D}^{\pi',1-b}_{\pi,\ell,m}$ will serve as the m-th "child" of the ℓ-th leaf of any of the components $\mathcal{D}^{*,b}_{\pi',*,*}$, or, in the case of $b = 1$ and $\pi' \in NTP_{\mathcal{I}}$, of the component $\mathcal{D}_{\pi'}$.[2] This explains our naming scheme.



Formally, let d' be the ℓ-th leaf of some component $\mathcal{D}^{*,b}_{\pi',*,*}$ or $\mathcal{D}_{\pi'}$ (in the latter case set $b := 1$), let $\mathrm{d}_1, \ldots, \mathrm{d}_k$ be the naturally ordered list of the children of $\mathfrak{f}(\mathrm{d}')$ in $\mathcal{I}$. Let $\pi_1, \ldots, \pi_k$ be their downward types. For all $m = 1, \ldots, k$ we join d' with the origin of $\mathcal{D}^{\pi',1-b}_{\pi_m,\ell,m}$ in the same way (*i.e.* by the same atomic roles) as $\mathfrak{f}(\mathrm{d}')$ is joined with d_m in $\mathcal{I}$. Repeat this step for all leaves in all components.

- For any pair of elements which we haven't explicitly connected in the above steps, we leave it unconnected (we do not join them by any atomic role).

This finishes the definition of $\mathcal{J}$. We prove correctness next.

3.3 Preservation of Knowledge Base Satisfiability

We show $\mathcal{J} \models \mathcal{K}$. $\mathcal{K}$ is in normal form with empty ABox, thus we consider the GCIs:

1. $\mathrm{A} \equiv \{\mathrm{o}\}$, $\mathrm{A} \equiv \neg\mathrm{B}$, $\mathrm{A} \equiv \mathrm{B} \sqcup \mathrm{B}'$, and $\mathrm{A} \equiv \exists s.\mathsf{Self}$. All of them are preserved in $\mathcal{J}$ since atomic types of any $\mathrm{d} \in \Delta^{\mathcal{J}}$ and $\mathfrak{f}(\mathrm{d})$ in $\mathcal{I}$ coincide.

2. $\mathrm{A} \equiv\ \geqslant n\ s.\mathrm{B}$. Here we consider the case of $\mathcal{ZOQ}$, where we only count the "forward successors" of a given element. Take $\mathrm{d} \in \Delta^{\mathcal{J}}$. If d is not a leaf of its component, then it "sends forward edges" only to its children and, possibly, to nominal elements. $\mathcal{J}$ restricted to these elements is isomorphic to $\mathcal{I}$ restricted to $\mathfrak{f}(\mathrm{d})$, its children and the nominal elements of $\mathcal{I}$. If d is a leaf then in our construction (1st item of the linking process) we link d with nominals in exactly the same way as $\mathfrak{f}(\mathrm{d})$ is linked to them in $\mathcal{I}$. Moreover if $\mathfrak{f}(\mathrm{d})$ has ℓ children $\mathrm{e}_1, \ldots \mathrm{e}_\ell$ in $\mathcal{I}$ then d is connected to exactly ℓ other domain elements (origins) $\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_\ell$ in a way that $(\mathrm{d}, \mathrm{d}_i)$ are in exactly the same atomic (and thus also simple) roles in $\mathcal{J}$ as $(\mathfrak{f}(\mathrm{d}), e_i)$ are in $\mathcal{I}$ — see the 2nd point of the construction. Due to the space limit the case of $\mathcal{ZOI}$ is left for the reader; recall that in this case we have $n = 1$.

3. $\mathcal{J} \models \mathrm{A} \sqsubseteq \exists p_{\mathbb{A}}.\mathrm{B}$. As a first step, we show by induction on k that for any origin e in $\mathcal{J}$ and any GCI $\mathrm{C} \sqsubseteq \exists p_{\mathbb{A}'}.\mathrm{D}$ such that $\mathrm{e} \in \mathrm{C}^{\mathcal{J}}$: if $\mathrm{d} = \mathfrak{f}(\mathrm{e})$ has a downward witness path for $\exists p_{\mathbb{A}'}.\mathrm{D}$ with at most k occurrences of nominal elements on it (not counting the first occurrence of e, in case e is nominal) then $\mathrm{e} \in (\exists p_{\mathbb{A}'}.\mathrm{D})^{\mathcal{J}}$. Having the above proved, we take any e in $\mathcal{J}$ such that $\mathrm{e} \in \mathrm{A}^{\mathcal{J}}$ and show that $\mathrm{e} \in (\exists p_{\mathbb{A}}.\mathrm{B})^{\mathcal{J}}$. If e is an origin then this follows from the above property. If not, let $\mathcal{D}$ be the component e belongs to, $\mathrm{d} = \mathfrak{f}(\mathrm{e})$, and $\mathcal{C}$ the component d belongs to. Note that $\mathcal{C}$ and $\mathcal{D}$ are isomorphic. Since $\mathfrak{f}$ retains atomic types we infer $\mathrm{d} \in \mathrm{A}^{\mathcal{J}}$. Take any downward witness path ρ for d and $\exists p_{\mathbb{A}}.\mathrm{B}$. If ρ is fully contained in $\mathcal{C}$ then its isomorphic copy, starting from e is contained in $\mathcal{D}$ and then $\mathrm{e} \in (\exists p_{\mathbb{A}}.\mathrm{B})^{\mathcal{J}}$. Otherwise, $\rho = \rho'\mathrm{d}'\rho''$, where d' is the first element not belonging to $\mathcal{C}$ (note that d' may be a nominal). Then an isomorphic copy ζ' of ρ', starting at e, belongs to $\mathcal{D}$. Assume that $\mathbb{A}$, when accepting ρ and reading (first time) the role-letter leading to d' enters a state $\mathtt{s}$. In $\mathcal{K}$ we have a GCI $\mathrm{A}_{\mathtt{s}} \equiv \exists p_{\mathbb{A}_{\mathtt{s}}}.\mathrm{B}$. Clearly $\mathrm{d}' \in (\exists p_{\mathbb{A}_{\mathtt{s}}}.\mathrm{B})^{\mathcal{I}}$ and thus $\mathrm{d}' \in \mathrm{A}_{\mathtt{s}}^{\mathcal{I}}$. By our scheme of joining the components the last element of ζ' is joined to some origin e' in $\mathcal{J}$ isomorphically to how the last element of ρ' is joined to d' in $\mathcal{I}$. Moreover the downward types of d' and $\mathfrak{f}(\mathrm{e}')$ are the same. This means that the atomic types of d', e' are equal, so $\mathrm{e}' \in \mathrm{A}_{\mathtt{s}}^{\mathcal{J}}$. Recalling that e' is an origin we have already proved that $\mathrm{e}' \in (\exists p_{\mathbb{A}_{\mathtt{s}}}.\mathrm{B})^{\mathcal{J}}$. Let ζ'' be a witness path for e' and $\exists p_{\mathbb{A}_{\mathtt{s}}}.\mathrm{B}$. It is readily verified that the path $\zeta'\zeta''$ is a witness path for e' and $\exists p_{\mathbb{A}}.\mathrm{B}$, so $\mathrm{e} \in (\exists p_{\mathbb{A}}.\mathrm{B})^{\mathcal{J}}$.

4. $\mathcal{J} \models \exists p_{\mathbb{A}}.\mathrm{B} \sqsubseteq \mathrm{A}$. The proof goes by induction on k where the inductive assumption states: for any $\mathrm{C}, \mathrm{D}, \exists p_{\mathbb{A}'}$ such that $\mathcal{T}$ contains $\mathrm{C} \equiv \exists p_{\mathbb{A}'}.\mathrm{D}$ and any $\mathrm{d} \in (\exists p_{\mathbb{A}'}.\mathrm{D})^{\mathcal{J}}$ for which there is a witness path ζ for $\exists p_{\mathbb{A}'}.\mathrm{D}$ on which the number of component changes is $\leq k$ it holds that $\mathrm{d} \in \mathrm{C}^{\mathcal{J}}$.

3.4 Preservation of Query (Non)entailment

Towards contradiction, assume $\mathcal{J} \models q$. Thus, there is a CQ q_i such that $\mathcal{J} \models q_i$ and let η be a match witnessing it. Let Δ_{q_j} be the image of η. We will define a homomorphism from $\mathcal{J}\restriction\Delta_{q_j}$ to $\mathcal{I}$, which will provide us with a match of q_j in $\mathcal{I}$, contradicting our initial assumption that $\mathcal{I} \not\models q$.

For convenience we treat separately the connections among non-nominal elements of Δ_{q_j} and connections involving at least one nominal element. Let us denote with $\Delta^*_{q_j}$ the set $\Delta_{q_j} \setminus \mathsf{Nom}_{\mathcal{J}}$ and let $G^*_{q_j}$ be the Gaifman graph of $\mathcal{J}\restriction\Delta^*_{q_j}$ (*i.e.* the graph, whose nodes are the domain elements, and an undirected edge between a pair of nodes is present if the nodes are connected by some atomic role). Note that the edges of $G^*_{q_j}$ correspond to parent-child connections inside components of $\mathcal{J}$ or connections between leaves of components and (non-nominal) origins of some other components. Let $G_1, \ldots, G_m$ be the connected components of $G^*_{q_j}$. We next construct a homomorphism $\mathfrak{h}_k$ from $\mathcal{J}\restriction G_k$ into $\Delta^{\mathcal{I}}$, for $k = 1, \ldots, m$. Setting additionally $\mathfrak{h}_0 : \mathsf{Nom}_{\mathcal{J}} \cap \Delta_{q_j} \to \mathsf{Nom}_{\mathcal{I}}$ in the natural way: $\mathfrak{h}_0(\mathrm{o}^{\mathcal{J}}) := \mathrm{o}^{\mathcal{I}}$, we will get the desired homomorphism $\mathfrak{h} = \bigcup_{k=0}^{m} \mathfrak{h}_k$ from Δ_{q_j} into $\Delta^{\mathcal{I}}$.

Consider a single G_k. Call the components of $\mathcal{J}$ containing nodes of G_k *active* (for G_k). If there is only one active component then as $\mathfrak{h}_k$ we take the restriction of $\mathfrak{f}$ to G_k. Otherwise call an active component *upper* (resp. *lower*) if at least one of its leaves (resp. its origin) belongs to G_k. Observe that for an active component it cannot be the case that both its origin and a leaf belong to G_k (since the path leading from the origin to a leaf has at least $K{+}1$ nodes and G_k is connected and has at most K nodes), and that each active component is either upper or lower (since different components may be

[2] The index b is not crucial here, but using it allows us to avoid *e.g.* the need to link leaves of a component with its origin.

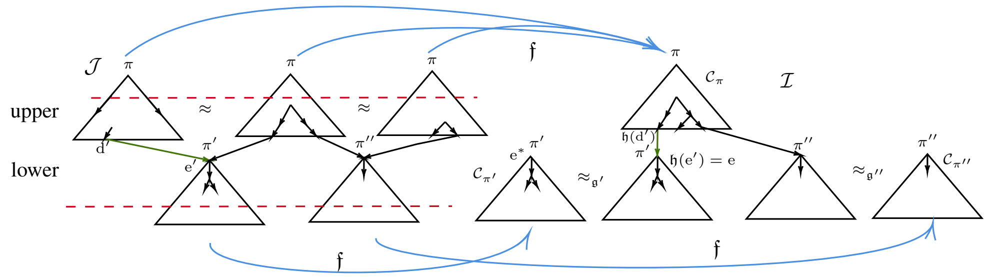


Figure 1: Constructing the homomorphism $\mathfrak{h}_k$. For the upper components $\mathfrak{h}_k$ coincides with f, for the lower ones it coincides with $\mathfrak{g} \cdot \mathfrak{f}$, for the appropriate $\mathfrak{g}$.

joined only by edges between leaves and origins). The origins of the lower components are not nominal elements. By our strategy of joining the components it must be the case that either all the upper components are of the form $\mathcal{D}^{*,b}_{\pi,*,*}$ for some fixed $\pi \in \mathit{DTP}_{\mathcal{I}} \setminus \mathit{NTP}_{\mathcal{I}}$ and $b \in \{0,1\}$, meaning that they are all isomorphic to $\mathcal{C}_\pi$ or there is only one upper component $\mathcal{D}_\pi$ for some $\pi \in \mathit{NTP}_{\mathcal{I}}$, which is then isomorphic to $\mathcal{C}_\pi$. For every d' of G_k belonging to an upper component we set $\mathfrak{h}_k(\mathrm{d}') = \mathfrak{f}(\mathrm{d}')$. Note that the images of all such d' are members of the single component $\mathcal{C}_\pi$ in $\mathcal{I}$. It remains to define $\mathfrak{h}_k$ for the elements of the lower components.

Consider any lower component $\mathcal{D}^{\pi,b}_{\pi',i,j}$ and its origin e' (of type π'). Let d' be the i-th leaf of an upper component. In our process d' has been connected to e' and this connection is isomorphic to the connection between the i-th leaf of $\mathcal{C}_\pi$ and its j-th child e in $\mathcal{I}$. Denoting e^* the origin of $\mathcal{C}_{\pi'}$, by the definition of downward types we have that $\mathcal{I}\restriction \mathit{Subtree}^{\leq K}_{\mathcal{I}}(e)$ is isomorphic to the upper part $\mathit{Subtree}^{\leq K}_{\mathcal{I}}(e^*)$ of $\mathcal{C}_{\pi'}$. Let $\mathfrak{g} : \mathit{Subtree}^{\leq K}_{\mathcal{I}}(e^*) \to \mathit{Subtree}^{\leq K}_{\mathcal{I}}(e)$ be the appropriate isomorphism. For all d from $\mathcal{D}^{\pi,b}_{\pi',i,j}$ we set $\mathfrak{h}_k(\mathrm{d}) := \mathfrak{g}(\mathfrak{f}(\mathrm{d}))$. Let $\mathfrak{h} := \bigcup_{k=0}^{m} \mathfrak{h}_k$.

We explain that $\mathfrak{h}$ is indeed a homomorphism. Concept preservation follows from the fact that $\mathfrak{f}$ and the $\mathfrak{g}$s used in the definition of the $\mathfrak{h}_k$ preserve atomic types and that the nominal elements of $\mathcal{J}$ have the same atomic types as the corresponding nominal elements of $\mathcal{I}$. Assume now that $(\mathrm{d}', \mathrm{e}') \in r^{\mathcal{J}}$ for some $\mathrm{d}', \mathrm{e}' \in \Delta_{q_j}$ and we will show $(\mathfrak{h}(\mathrm{d}'), \mathfrak{h}(\mathrm{e}')) \in r^{\mathcal{I}}$. We consider the following three cases:

1. $\mathrm{d}', \mathrm{e}' \in \Delta^*_{q_j}$ and they belong to the same component. Then $(\mathfrak{h}(\mathrm{d}'), \mathfrak{h}(\mathrm{e}')) \in r^{\mathcal{I}}$ holds, since $\mathfrak{f}$ acts as a partial isomorphism when restricted to a single component and $\mathfrak{g}$ is a partial isomorphism in $\mathcal{I}$.
2. $\mathrm{d}', \mathrm{e}' \in \Delta^*_{q_j}$ and they belong to different components. Then d' is a leaf of a component and e' is an origin of another component, or vice versa; (w.l.o.g. we focus on former case). Looking at Fig. 1 we can see that $\mathfrak{h}(\mathrm{e}')(= \mathfrak{f}(\mathfrak{g}(\mathrm{e}'))$ for the appropriate $\mathfrak{g}$) is a child of $\mathfrak{h}(\mathrm{d}')(= \mathfrak{f}(\mathrm{d}'))$. By the construction of $\mathcal{J}$ the connection between d' and e' in $\mathcal{J}$ is isomorphic to the connection between $\mathfrak{h}(\mathrm{d}')$ and $\mathfrak{h}(\mathrm{e}')$ in $\mathcal{I}$ (from which the claim follows).
3. At least one of $\mathrm{d}', \mathrm{e}' \in \Delta_{q_j} \setminus \Delta^*_{q_j}$ $(\subseteq \mathsf{Nom}_{\mathcal{J}})$. Follows from the fact that when defining $\mathcal{J}$ we always join every element d' with the nominal elements in $\mathcal{J}$ in the same way as $\mathfrak{f}(\mathrm{d}')$ is joined with the corresponding nominals in $\mathcal{I}$, and that $\mathfrak{f}(\mathrm{d}')$ and $\mathfrak{g}(\mathfrak{f}(\mathrm{d}'))$ are joined with nominals in $\mathcal{I}$ in the same way (equal downward types!).

4 Querying $\mathcal{ZIQ}$ and its Sublogics

The third of the main members of the $\mathcal{Z}$ family, $\mathcal{ZIQ}$, does not have fmp. Recall that its (general) KBSat and query entailment are decidable and ExpTime-, resp., 2ExpTime-complete, but the decidability of the corresponding finite problems is open. Here we provide a reduction from its finite PEQ entailment problem to doubly exponentially many instances of its finite KBSat problem. This would be sufficient to give the optimal 2ExpTime-upper bound on PEQ entailment for $\mathcal{ZIQ}$, provided that its finite KBSat is not harder than general KBSat. In fact, the reduction will work not only for $\mathcal{ZIQ}$ but also for some of its sublogics, which will lead to some new complexity results. We now describe two steps:

1. A uniform reduction from finite PEQ entailment to finite KBSat for any $\mathcal{L}$ that is *locally of treewidth-one* (ltw1).
2. An appropriate model transformation[3] that transforms any finite countermodel for a $\mathcal{ZIQ}$-KB and a PEQ q into one that is ltw1. This generalises the existing methods presented in (Baader, Bednarczyk, and Rudolph 2019) and (Ibáñez-García, Lutz, and Schneider 2014).

4.1 From Finite PEQ Entailment to Finite KBSat

For $k \geq 0$ we define the *k-neighbourhood* of d in $\mathcal{I}$, denoted $\mathsf{Nbd}^k_{\mathcal{I}}(\mathrm{d})$, as the restriction of $\mathcal{I}$ to elements reachable from d in $\mathcal{I}$ by an *undirected* path of length $\leq k$. An N-rooted *tw1-forest* $\mathcal{I}$ is a quasi-forest with partial assignment of individual names that satisfies: (i) $\{\mathrm{o}^{\mathcal{I}} \mid \mathrm{o} \in \mathsf{N}\}$ are the roots of $\mathcal{I}$, and (ii) if $(\mathrm{d}, \mathrm{e}) \in r^{\mathcal{I}}$ and one of d, e is a root of $\mathcal{I}$, then the other element is either its child or is also a root of $\mathcal{I}$. A single-root tw1-forest is a *tw1-tree*. For $n \in \mathbb{N}$ and $\mathsf{N} \subseteq \mathbf{N_I}$, we say that $\mathcal{I}$ is (n, N)-*ltw1* if, for

[3] We thank Ian Pratt-Hartmann for discussions.

any $\mathrm{d} \in \Delta^{\mathcal{I}}$, $\mathsf{Nbd}^n_{\mathcal{I}}(\mathrm{d})$ is either a tw1-tree or an N'-rooted tw1-forest, where $\mathsf{N}' = \{\mathrm{o} \in \mathsf{N} \mid \mathrm{o}^{\mathsf{Nbd}^n_{\mathcal{I}}(\mathrm{d})} \text{ is defined}\}$.[4]

Definition 4.1 (coverable/locally-treewidth-one DL)**.** *Let $\mathcal{L}$ be a DL and let $\mathcal{K}$ be an $\mathcal{L}$-KB. We say that $\mathcal{K}$ is (finitely) ltw1-coverable iff for any (finite) model $\mathcal{I} \models \mathcal{K}$ and every $n \in \mathbb{N}$ there is a (finite)* $(n, \mathsf{ind}(\mathcal{K}))$-lwt1 model $\mathcal{J} \models \mathcal{K}$ *that* covers $\mathcal{I}$, i.e. *any n-neighbourhood of $\mathcal{J}$ can be homomorphically-mapped to $\mathcal{I}$. $\mathcal{L}$ is (finitely) ltw1 iff $\mathcal{ALCI}^{\cap}_{\mathsf{Self}} \subseteq \mathcal{L}$ and all (finitely) satisfiable $\mathcal{L}$-KBs are (finitely) ltw1-coverable.*

Let $\mathcal{L}$ be finitely ltw1. Fix an $\mathcal{L}$-KB $\mathcal{K}$ and a PEQ q. We outline how to decide $\mathcal{K} \models_{\text{fin}} q$ using KBSat as a subroutine. Our approach is a modification of techniques of (Glimm et al. 2008), adjusted to work for finite (rather than unrestricted) entailment and for PEQs (rather than just CQs).

1. Using the known rolling-up technique we show that for any tw1-tree-shaped CQ $\hat{q}$ there is an $\mathcal{ALCI}^{\cap}_{\mathsf{Self}}$-concept $\mathrm{Match}_{\hat{q}}$ such that $\mathrm{d} \in (\mathrm{Match}_{\hat{q}})^{\mathcal{I}}$ iff there exists a homomorphism from the query structure $\mathcal{I}_{\hat{q}}$ to $\mathcal{I}$ mapping d to the root of $\hat{q}$. This is why we need that $\mathcal{ALCI}^{\cap}_{\mathsf{Self}} \subseteq \mathcal{L}$.
2. When a CQ $\hat{q}$ with $|\hat{q}| \leq |q|$ matches a $(|q|, \mathsf{ind}(\mathcal{K}))$-ltw1 $\mathcal{I}$, then the match induces a CQ $\hat{q}'$ (obtained by identifying some of the variables from $\hat{q}$) and a *splitting* Π of the variables of $\hat{q}'$ that is *compatible* with $\mathcal{I}$. A splitting Π is a tuple composed of sets of variables that partition $\mathrm{Var}(\hat{q}')$ into trees T, subtrees S_i and roots R, the name assignment $\mathsf{ind}(\mathcal{K}) \to \mathsf{R}$, and a root assignment mapping each S_i to some variable from R. The query $\hat{q}$ restricted to any of the S_i or to T forms a tw1-tree. Compatibility ensures that Π actually induces a match in $\mathcal{I}$, which is achieved by means of the Match concepts and by specifying concepts/role connections between the named individuals from $\mathsf{ind}(\mathcal{K})$. We prove that $\mathcal{I} \models \hat{q}$ iff such $\hat{q}'$ and a splitting Π compatible with $\mathcal{I}$ exists.
3. A *spoiler* for $\hat{q}'$ and Π is an $\mathcal{ALCI}^{\cap}_{\mathsf{Self}}$-KB containing an axiom preventing Π from being compatible with any $\mathcal{I}$. A super-spoiler is an $\mathcal{ALCI}^{\cap}_{\mathsf{Self}}$-KB preventing matches for all the possible $\hat{q}'$ and splittings Π, that may arise for $\hat{q}$. We show that there is a countermodel for $\mathcal{K}$ and $\hat{q}$ iff for some super-spoiler $\mathcal{K}^{s*}_{\hat{q}}$ a finite model for $\mathcal{K}^{s*}_{\hat{q}} \cup \mathcal{K}$ exists.
4. We convert the input PEQ q into a disjunction of exponentially many CQs of polynomial size (both bounds in terms of $|q|$). We consider all possible tuples of super-spoilers (one per each disjunct $\hat{q}$). For each such tuple we test finite KBSat for its union with $\mathcal{K}$. We answer that $\mathcal{K} \models_{\text{fin}} q$ iff all these tests are negative.

As super-spoilers are of at most exponential size, the above procedure requires at most doubly exponentially many KBSat calls, with exponentially bounded inputs (in $|q|$). This yields the following (the lower bound is from (Lutz 2007)).

Theorem 4.2. *For any finitely ltw1 $\mathcal{L}$, $\mathcal{L}$-KB $\mathcal{K}$ and PEQ q, we can decide $\mathcal{K} \models_{\text{fin}} q$ by doubly-exponentially many (in $|\mathcal{K}|+|q|$) checks of finite KBSat of $\mathcal{L}$-KBs of size exponential (in $|\mathcal{K}|+|q|$). If $\mathcal{L}$ has* ExpTime*-complete finite KBSat, then its finite PEQ entailment is* 2ExpTime*-complete.*

[4] Trees in tw1-forests have bidirectional edges and possibly self-loops. Ltw1-forests are finite approximants of tw1-forests: they are not necessarily tw1-forests but they *locally* look like them.

4.2 $\mathcal{ZIQ}$ is Finitely LTW1

Let $\mathcal{K}$ be a $\mathcal{ZIQ}$-KB with a finite model $\mathcal{I}$. For $n \geq 2$, a *cycle* of length $n+1$ in $\mathcal{I}$ is a sequence of distinct elements $\mathrm{d}_0, \mathrm{d}_1, \ldots, \mathrm{d}_n$ such that (assuming $\mathrm{d}_{n+1} = \mathrm{d}_0$) for all $0 \leq i \leq n$ there is a simple role r_i s.t. $(\mathrm{d}_i, \mathrm{d}_{i+1}) \in r_i^{\mathcal{I}}$. An element d is *named* if $\mathrm{d} = \mathrm{o}^{\mathcal{I}}$ for some $\mathrm{o} \in \mathsf{ind}(\mathcal{K})$. Otherwise d is *anonymous*. A cycle $\mathrm{d}_0, \mathrm{d}_1, \ldots, \mathrm{d}_n$ is *anonymous* if $\mathrm{d}_1, \ldots, \mathrm{d}_{n-1}$ are anonymous and if d_n is named then so is d_0. The *anonymous girth* of $\mathcal{I}$, denoted $\mathsf{agirth}(\mathcal{I})$, is the length of its smallest anonymous cycle (∞ if there is no such cycles). We will show that for any $G > 0$ there is a finite $\mathcal{J} \models \mathcal{K}$ with $\mathsf{agirth}(\mathcal{J}) \geq G$ that covers $\mathcal{I}$. This implies that $\mathcal{J}$ is $(G, \mathsf{ind}(\mathcal{K}))$-ltw1 and proves that $\mathcal{K}$ is finitely ltw1.

We will construct finite interpretations $\mathcal{J}_0, \mathcal{J}_1, \ldots$ with origin homomorphism $\mathsf{orig} : \bigcup_i \Delta^{\mathcal{J}_i} \to \Delta^{\mathcal{I}}$ satisfying:

- For any $\mathrm{d} \in \Delta^{\mathcal{J}_i}$, we have $\mathrm{d} \in \mathrm{C}^{\mathcal{J}}$ iff $\mathsf{orig}(\mathrm{d}) \in \mathrm{C}^{\mathcal{I}}$ for all C appearing in $\mathcal{K}$, and orig maps isomorphically the 1-neighbourhood of d to the 1-neighbourhood of $\mathsf{orig}(\mathrm{d})$.
- If $\mathsf{agirth}(\mathcal{J}_i) = g \geq G$ we stop. Otherwise $\mathsf{agirth}(\mathcal{J}_{i+1}) \geq g$ and $\mathcal{J}_{i+1}$ contains less anonymous cycles of length g than $\mathcal{J}_i$, and no anonymous cycles of length $< g$.

Let $M = |\Delta^{\mathcal{I}}| \cdot (1 + |\Delta^{\mathcal{I}}| + \ldots + |\Delta^{\mathcal{I}}|^{G+1}) + 1$. The domain of $\mathcal{J}_0$ is composed of M copies of $\mathcal{I}$. Roles and concepts in $\mathcal{J}_0$ are interpreted in each copy as in $\mathcal{I}$, and the individual names are assigned to the first copy. The origin function orig simply sends an element $\mathrm{d}' \in \Delta^{\mathcal{J}}$ to its pattern in $\Delta^{\mathcal{I}}$. Note that our requirements are satisfied by $\mathcal{J}_0$.

Suppose $\mathcal{J}_{i-1}$ is defined and $\mathsf{agirth}(\mathcal{J}_{i-1}) = g < G$. We initially put $\mathcal{J}_i := \mathcal{J}_{i-1}$, take its shortest anonymous cycle $\mathrm{d}_0, \mathrm{d}_1, \ldots, \mathrm{d}_n$ and select two consecutive elements c, d (w.l.o.g. d is anonymous) on this cycle. By the choice of M, there are c', d' satisfying $\mathsf{orig}(\mathrm{c}) = \mathsf{orig}(\mathrm{c}')$ and $\mathsf{orig}(\mathrm{d}) = \mathsf{orig}(\mathrm{d}')$ that are not reachable from c, d nor any named elements in $\leq G$ steps, and for any simple role s we have $(\mathrm{c}, \mathrm{d}) \in s^{\mathcal{J}_{i-1}}$ iff $(\mathrm{c}', \mathrm{d}') \in s^{\mathcal{J}_{i-1}}$. Hence, we can twist the roles between the elements, *e.g.* $r^{\mathcal{J}_i} = (r^{\mathcal{J}_{i-1}} \setminus \{(\mathrm{c}, \mathrm{d}), (\mathrm{c}', \mathrm{d}')\}) \cup \{(\mathrm{c}, \mathrm{d}'), (\mathrm{c}', \mathrm{d})\}$ for $r^{\mathcal{J}_i}$ containing (c, d) (and similarly for roles containing (d, c)). The other roles remain as in $\mathcal{J}_{i-1}$. We can show, that $\mathcal{J}_i$ indeed satisfies the required conditions and no new cycle of length $\leq g$ is added. This finishes the construction of $\mathcal{J}_i$.

As $\mathcal{J}$ we take the last interpretation in the sequence.

Theorem 4.3. *Any $\mathcal{ALCI}^{\cap}_{\mathsf{Self}} \subseteq \mathcal{L} \subseteq \mathcal{ZIQ}$ is finitely ltw1.*

We give two applications of Theorem 4.2 and Theorem 4.3. It was recently shown (Jung, Lutz, and Zeume 2020) that the finite KBSat of certain fragments of $\mathcal{ZIQ}$, called $\mathcal{ALCHIF}^{1/2}_{\text{reg}}$, is decidable in $(1/2)$NExpTime. The algorithm is automata-based and hence allows us to accommodate conjunctions of roles and self-loops easily.

Corollary 4.4. *Finite PEQ-entailment over $\mathcal{ALCHIF}^{1/2}_{reg}$-KBs is decidable in* $(2/3)$NExpTime.

The second application is lifting the existing results on CQ entailment to the setting of UCQs and PEQs (the lower bound for PEQs holds already for $\mathcal{ALC}$ (Ortiz and Simkus 2014)).

Corollary 4.5. *Any $\mathcal{L}$ satisfying $\mathcal{ALC} \subseteq \mathcal{L} \subseteq \mathcal{ALCIH}b^{\mathsf{Self}}\mathcal{Q}$ has* 2ExpTime*-complete finite PEQ entailment problem.*

Acknowledgements

Bartosz Bednarczyk is supported by the European Research Council through the Consolidator Grant No. 771779 DeciGUT. Emanuel Kieroński was supported by the Polish National Science Centre grant 2018/31/B/ST6/03662.

References

Amendola, G.; Leone, N.; and Manna, M. 2017. Finite model reasoning over existential rules. *Theory Pract. Log. Program. 2017*.

Baader, F.; Bednarczyk, B.; and Rudolph, S. 2019. Satisfiability Checking and Conjunctive Query Answering in Description Logics with Global and Local Cardinality Constraints. In *DL 2019*.

Bárány, V.; Gottlob, G.; and Otto, M. 2014. Querying the Guarded Fragment. *LMCS 2014*.

Bednarczyk, B.; and Rudolph, S. 2019. Worst-Case Optimal Querying of Very Expressive Description Logics with Path Expressions and Succinct Counting. *IJCAI 2019*.

Bienvenu, M.; Lutz, C.; and Wolter, F. 2012. Query Containment in Description Logics Reconsidered. In *KR 2012*.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2009. Regular Path Queries in Expressive Description Logics with Nominals. *IJCAI 2009*.

Calvanese, D.; Ortiz, M.; and Simkus, M. 2011. Containment of Regular Path Queries under Description Logic Constraints. In *IJCAI 2011*.

Calvanese, D.; Ortiz, M.; and Simkus, M. 2016. Verification of Evolving Graph-structured Data under Expressive Path Constraints. In *ICDT 2016*.

Danielski, D.; and Kieronski, E. 2019. Finite Satisfiability of Unary Negation Fragment with Transitivity. In *MFCS 2019*.

Glimm, B.; Lutz, C.; Horrocks, I.; and Sattler, U. 2008. Conjunctive Query Answering for the Description Logic SHIQ. *J. Artif. Intell. Res. 2008*.

Gogacz, T.; Gutiérrez-Basulto, V.; Gutowski, A.; Ibáñez-García, Y.; and Murlak, F. 2020. On Finite Entailment of Non-Local Queries in Description Logics. In *KR 2020*.

Gogacz, T.; Gutiérrez-Basulto, V.; Ibáñez-García, Y.; Jung, J. C.; and Murlak, F. 2019. On Finite and Unrestricted Query Entailment beyond SQ with Number Restrictions on Transitive Roles. In *IJCAI 2019*.

Gogacz, T.; Ibáñez-García, Y. A.; and Murlak, F. 2018. Finite Query Answering in Expressive Description Logics with Transitive Roles. In *KR 2018*.

Gogacz, T.; and Marcinkowski, J. 2017. Converging to the chase - A tool for finite controllability. *J. Comput. Syst. Sci. 2017*.

Gottlob, G.; Manna, M.; and Pieris, A. 2018. Finite Model Reasoning in Hybrid Classes of Existential Rules. In *IJCAI 2018*.

Grau, B. C.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P. F.; and Sattler, U. 2008. OWL 2: The next step for OWL. *J. Web Semant.*

Gutiérrez-Basulto, V.; Ibáñez-García, Y. A.; and Jung, J. C. 2017. Number Restrictions on Transitive Roles in Description Logics with Nominals. In *AAAI 2017*.

Horrocks, I.; Kutz, O.; and Sattler, U. 2006. The Even More Irresistible SROIQ. In *KR 2006*.

Ibáñez-García, Y. A.; Lutz, C.; and Schneider, T. 2014. Finite Model Reasoning in Horn Description Logics. In *KR 2014*.

Jung, J. C.; Lutz, C.; and Zeume, T. 2019. Decidability and Complexity of ALCOIF with Transitive Closure (and More). In *DL 2019*.

Jung, J. C.; Lutz, C.; and Zeume, T. 2020. On the Decidability of Expressive Description Logics with Transitive Closure and Regular Role Expressions. In *KR 2020*.

Lutz, C. 2007. Inverse Roles Make Conjunctive Queries Hard. In *DL 2007*.

Nguyen, L. A. 2020. ExpTime Tableaux with Global Caching for Hybrid PDL. *J. Autom. Reason. 2020*.

Oritz, M. 2010. *Query answering in expressive description logics: techniques and complexity results*. Ph.D. thesis, TU Wien, AT.

Ortiz, M.; and Simkus, M. 2012. Reasoning and Query Answering in Description Logics. *Reasoning Web 2012*.

Ortiz, M.; and Simkus, M. 2014. Revisiting the Hardness of Query Answering in Expressive Description Logics. In *RR 2014*.

Rudolph, S. 2016. Undecidability Results for Database-Inspired Reasoning Problems in Very Expressive Description Logics. In *KR 2016*.