Query Conservative Extensions in Horn Description Logics with Inverse Roles

Jean Christoph Jung, Carsten Lutz, Mauricio Martel, and **Thomas Schneider**
Fachbereich Informatik, Universität Bremen, Germany

Abstract

We investigate the decidability and computational complexity of query conservative extensions in Horn description logics (DLs) with inverse roles. This is more challenging than without inverse roles because characterizations in terms of unbounded homomorphisms between universal models fail, blocking the standard approach to establishing decidability. We resort to a combination of automata and mosaic techniques, proving that the problem is 2EXPTIME-complete in Horn-$\mathcal{ALCHIF}$ (and also in Horn-$\mathcal{ALC}$ and in $\mathcal{ELI}$). We obtain the same upper bound for deductive conservative extensions, for which we also prove a CONEXPTIME lower bound.

1 Introduction

In the past years, access of incomplete data mediated by description logic (DL) ontologies has gained increasing importance [Poggi *et al.*, 2008; Bienvenu and Ortiz, 2015]. The main idea is to specify domain knowledge and semantics of the data in the ontology, resulting in more complete answers to queries. Significant research activity has led to efficient algorithms and tools for a wide range of DLs such as DL-Lite [Calvanese *et al.*, 2007], more expressive Horn-DLs [Eiter *et al.*, 2012; Trivela *et al.*, 2015; Bienvenu *et al.*, 2016], and "full Boolean" DLs such as $\mathcal{ALC}$ [Kollia and Glimm, 2013; Zhou *et al.*, 2015].

In contrast to query answering, which is by now well-understood, there is a need to develop reasoning services for ontology engineering that are tailored towards query-centric applications and support tasks such as ontology versioning and module extraction from ontologies. For example, if one wants to safely replace an ontology with a new version or with a smaller subset of itself (a module), then the new ontology should preserve the answers to all queries over all ABoxes (which store the data) [Kontchakov *et al.*, 2010]. The same guarantee ensures that one can safely replace an ontology with another version in an application [Konev *et al.*, 2012]. In both cases, ontologies need to be tested not for their logical equivalence, but for giving the same answers to relevant queries over relevant datasets.

This requirement can be formalized using conservative extensions. In the following, we use the DL term *TBox* instead of *ontology*. A TBox $\mathcal{T}_2 \supseteq \mathcal{T}_1$ is a (Γ, Σ)-*query conservative extension* of a TBox $\mathcal{T}_1$, where Γ and Σ are signatures of concept/role names relevant for data and queries, respectively, if all Σ-queries give the same answers w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$, for every Γ-ABox. Note that the subset relationship $\mathcal{T}_2 \supseteq \mathcal{T}_1$ is natural when replacing a TBox with a module, but not in versioning, so we might not want to insist on it. In this more general case, $\mathcal{T}_1$ and $\mathcal{T}_2$ are called (Γ, Σ)-*query inseparable*. Conservativity and inseparability of *TBoxes*, as defined above, are useful when knowledge is considered static and data changes frequently. Variants of these notions for *knowledge bases (KBs)*, which consist of a TBox and an ABox, can be used for applications with static data [Wang *et al.*, 2014; Arenas *et al.*, 2016].

We also consider the basic notion of query entailment: $\mathcal{T}_1$ (Γ, Σ)-*query entails* $\mathcal{T}_2$ if all Σ-queries give *at least* the answers w.r.t. $\mathcal{T}_1$ that they give w.r.t. $\mathcal{T}_2$, on any Γ-ABox. Query inseparability and conservativity are special cases of entailment: inseparability is bidirectional entailment and conservativity is entailment with the assumption that $\mathcal{T}_1 \subseteq \mathcal{T}_2$. It thus suffices to prove upper bounds for query entailment and lower bounds for conservative extensions. As a query language, we concentrate on conjunctive queries (CQs); since we work with Horn-DLs and quantify over the queries, this is equivalent to using unions of CQs (UCQs) or positive existential queries (PEQs). CQ entailment has been studied for various DLs [Kontchakov *et al.*, 2009; Lutz and Wolter, 2010; Konev *et al.*, 2012; Botoeva *et al.*, 2016c], also in the KB version [Botoeva *et al.*, 2016b; Botoeva *et al.*, 2016c] and for OBDA specifications [Bienvenu and Rosati, 2015], see also the survey [Botoeva *et al.*, 2016a]. Nevertheless, there is still a notable gap in our understanding of this notion: query entailment between TBoxes is poorly understood in Horn DLs with inverse roles, often considered a crucial feature, for which there do not seem to be any available results. This is for a reason: it has been observed in [Botoeva *et al.*, 2016a; Botoeva *et al.*, 2016b] that standard techniques for Horn DLs without inverse roles fail when inverse roles are added.

In fact, for Horn-DLs without inverse roles query entailment can be characterized by the extence of homomorphisms between universal models [Lutz and Wolter, 2010; Botoeva *et al.*, 2016a]. The resulting characterizations provide an important foundation for decision procedures, often based on tree automata [Botoeva *et al.*, 2016a]. In the presence of inverse roles, however, such characterizations are only correct if we require the existence of n-*bounded* homomorphisms, *for any* n [Botoeva *et al.*, 2016a; Botoeva *et al.*, 2016b]. It is not obvious how the existence of such infinite families of bounded homomorphisms can be verified using tree automata (or related techniques) and, consequently, decidability results for query conservative extensions in Horn-DLs with inverse roles are difficult to obtain. The only result we are aware of concerns inseparability *of KBs*, and it is proved using intricate game-theoretic techniques.

In this paper, we develop decision procedures for query entailment and related problems in Horn DLs with inverse roles. The main idea is to provide a more refined characterization, mixing unbounded and bounded homomorphisms and using unbounded homomorphisms only in places where this is strictly necessary. We can then deal with the "unbounded part" using tree automata while the "bounded part" is addressed by precomputing relevant information using a mosaic technique. In this way, we establish decidability and a 2EXPTIME upper bound for query entailment (and thus inseparability and conservativity) in Horn-$\mathcal{ALCHIF}$. Together with lower bounds from [Botoeva *et al.*, 2016c], we get 2EXPTIME-completeness for all fragments of Horn-$\mathcal{ALCHIF}$ that contain $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$.

We additionally study the case of deductive entailment between TBoxes, i.e., the question whether $\mathcal{T}_1$ entails at least the same concept and role inclusions as well as functionality assertions over Σ as $\mathcal{T}_2$. This problem too has not previously been studied for Horn DLs with inverse roles. We consider $\mathcal{ELHIF}_\bot$-TBoxes and show that deductive entailment is equivalent to a restricted version of query entailment. We obtain a model theoretic characterization, a decision procedure, and a 2EXPTIME upper complexity bound. We also give a coNEXPTIME lower bound.

Omitted proofs can be found in the long version here: www.informatik.uni-bremen.de/tdki/research/papers.html

2 Preliminaries

2.1 Horn-$\mathcal{ALCHIF}$

We introduce Horn-$\mathcal{ALCHIF}$, a member of the Horn-$\mathcal{SHIQ}$ family of DLs whose reasoning problems have been widely studied [Hustadt *et al.*, 2007; Krötzsch *et al.*, 2007; Eiter *et al.*, 2008; Kazakov, 2009; Lutz and Wolter, 2012; Ibáñez-García *et al.*, 2014]. Let $\mathsf{N_C}, \mathsf{N_R}, \mathsf{N_I}$ be sets of concept, role, and individual names. A *role* is either a role name r or an *inverse role* r^-. As usual, we identify $(r^-)^-$ and r, allowing to switch between roles names and their inverses easily. A *concept inclusion (CI)* is of the form $L \sqsubseteq R$, where L and R are concepts defined by the syntax rules

$$R, R' ::= \top \mid \bot \mid A \mid \neg A \mid R \sqcap R' \mid \neg L \sqcup R \mid \exists r.R \mid \forall r.R$$

$$L, L' ::= \top \mid \bot \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L$$

with A ranging over concept names and r over roles. A *role inclusion (RI)* is of the form $r \sqsubseteq s$ with r, s roles and a *functionality assertion (FA)* is of the form $\mathsf{func}(r)$ with r a role. $\mathcal{ELI}_\bot$*-concepts* are expressions that are built according to the syntax rule for L above, but do not use "$\sqcup$".

A *Horn-$\mathcal{ALCHIF}$ TBox* $\mathcal{T}$ is a set of CIs, RIs, and FAs. An $\mathcal{ELHIF}_\bot$ *TBox* is a set of $\mathcal{ELI}_\bot$-CIs, RIs, and FAs. To avoid dealing with rather messy technicalities that do neither seem to be very illuminating from a theoretical viewpoint nor too useful from a practical one,[1] we generally assume that functional roles cannot have any subroles, that is, $r \sqsubseteq s \in \mathcal{T}$ implies $\mathsf{func}(s) \notin \mathcal{T}$. We conjecture that our main results also hold without that restriction. An *ABox* $\mathcal{A}$ is a non-empty set of *concept and role assertions* of the form $A(a)$ and $r(a, b)$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $a, b \in \mathsf{N_I}$. We write $\mathsf{ind}(\mathcal{A})$ for the set of individuals in $\mathcal{A}$.

The semantics is defined as usual in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ complying with the standard name assumption, i.e., $a^\mathcal{I} = a$ for all $a \in \mathsf{N_I}$ [Baader *et al.*, 2017]. An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if it satisfies all inclusions and assertions in it, and likewise for ABoxes. $\mathcal{A}$ is *consistent* with $\mathcal{T}$ if $\mathcal{T}$ and $\mathcal{A}$ have a common model.

A *signature* Σ is a set of concept and role names. A Σ-ABox is an ABox that uses only concept and role names from Σ, and likewise for Σ-$\mathcal{ELI}_\bot$-concepts and other syntactic objects.

Generally and without further notice, we work with Horn-$\mathcal{ALCHIF}$ TBoxes that are in a certain nesting-free normal form, that is, they contain only CIs of the form

$$\top \sqsubseteq A, \ A \sqsubseteq \bot, \ A_1 \sqcap A_2 \sqsubseteq B, \ A \sqsubseteq \exists r.B, \ A \sqsubseteq \forall r.B,$$

where A, B, A_1, A_2 are concept names and r, s are roles. It is well-known that every Horn-$\mathcal{ALCHIF}$ TBox $\mathcal{T}$ can be converted into a TBox $\mathcal{T}'$ in normal form (introducing additional concept names) such that $\mathcal{T}$ is a logical consequence of $\mathcal{T}'$ and every model of $\mathcal{T}$ can be extended to one of $\mathcal{T}'$ by interpreting the additional concept names, see e.g. [Bienvenu *et al.*, 2016]. As a consequence, all results obtained in this paper for TBoxes in normal form lift to the general case.

2.2 Query Conservative Extensions and Entailment

A *conjunctive query (CQ)* is of the form $q(\mathbf{x}) = \exists \mathbf{y}\, \varphi(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are tuples of variables and $\varphi(\mathbf{x}, \mathbf{y})$ is a conjunction of *atoms* of the form $A(v)$ or $r(v, w)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $v, w \in \mathbf{x} \cup \mathbf{y}$. We call $\mathbf{x}$ *answer variables* and $\mathbf{y}$ *quantified variables* of q. A CQ q is *tree-shaped* if the undirected graph $(\mathbf{x} \cup \mathbf{y}, \ \{\{v, w\} \mid r(v, w) \text{ is an atom in } q\})$ is a tree; tree-shaped CQs are thus connected and may contain multi-edges. A tree-shaped CQ q is *strongly tree-shaped* or an *stCQ* if the root is the one and only answer variable and q has no multi-edges, i.e., no two atoms $r_1(z_1, z_2), r_2(z_1', z_2')$ with $r_1 \neq r_2$ and $\{z_1, z_2\} = \{z_1', z_2'\}$.

A *match* of q in an interpretation $\mathcal{I}$ is a function $\pi : \mathbf{x} \cup \mathbf{y} \to \Delta^\mathcal{I}$ such that $\pi(v) \in A^\mathcal{I}$ for every atom $A(v)$ of q and $(\pi(v), \pi(w)) \in r^\mathcal{I}$ for every atom $r(v, w)$ of q. We write $\mathcal{I} \models q(a_1, \ldots, a_n)$ if there is a match of q in $\mathcal{I}$ with $\pi(x_i) = a_i$ for all $i < n$. A tuple $\mathbf{a}$ of elements from $\mathsf{N_I}$ is a *certain answer* to q over an ABox $\mathcal{A}$ given a TBox $\mathcal{T}$, written $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$, if $\mathcal{I} \models q(\mathbf{a})$ for all models of $\mathcal{T}$ and $\mathcal{A}$.

[1]E.g., out of 439 available ontologies in BioPortal [Matentzoglu and Parsia, 2017], only 21 ($\leq$ 4.8%) contain the described pattern. A significant fraction of the occurrences of the pattern appear to be due to modeling mistakes.

Definition 1 *Let Γ, Σ be signatures and $\mathcal{T}_1, \mathcal{T}_2$ Horn-$\mathcal{ALCHIF}$ TBoxes. We say that $\mathcal{T}_1$ (Γ, Σ)*-CQ entails *$\mathcal{T}_2$, written $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, if for all Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, all Σ-CQs $q(\mathbf{x})$ and all tuples $\mathbf{a} \subseteq \mathsf{ind}(\mathcal{A})$, $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ implies $\mathcal{T}_1, \mathcal{A} \models q(\mathbf{a})$. If in addition $\mathcal{T}_1 \subseteq \mathcal{T}_2$, we say that $\mathcal{T}_2$* is a (Γ, Σ)-CQ conservative extension *of $\mathcal{T}_1$. If $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and vice versa, then $\mathcal{T}_1$ and $\mathcal{T}_2$ are* (Γ, Σ)-CQ inseparable.

We also consider (Γ, Σ)-*stCQ entailment*, denoted $\models^{\mathrm{stCQ}}_{\Gamma,\Sigma}$ and defined in the obvious way by replacing CQs with stCQs.

If $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ because $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ but $\mathcal{T}_1, \mathcal{A} \not\models q(\mathbf{a})$ for some Γ-ABox $\mathcal{A}$ consistent with both $\mathcal{T}_i$, Σ-CQ $q(\mathbf{x})$ and $\mathbf{a}$, we call the triple $(\mathcal{A}, q, \mathbf{a})$ a *witness* to non-entailment.

Example 2 *Let $\mathcal{T}_1 = \{PhDStud \sqsubseteq \exists advBy.Prof, advBy \sqsubseteq adv^-\}$ and $\mathcal{T}_2 = \mathcal{T}_1 \cup \{\mathsf{func}(advBy)\}$, $\Sigma = \{Prof\}$ and $\Gamma = \{PhDStud, adv\}$. Then $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ because of the witness $(\{PhDStud(john), adv(mary, john)\}, Prof(x), mary)$.*

If we drop from Definition 1 the condition that $\mathcal{A}$ must be consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$, then we obtain an alternative notion of CQ entailment that we call *CQ entailment with inconsistent ABoxes*. While this new notion trivially implies CQ entailment in the original sense, the converse fails.

Example 3 *Let $\mathcal{T}_1 = \emptyset$, $\mathcal{T}_2 = \{A_1 \sqcap A_2 \sqsubseteq \bot\}$ and $\Gamma = \{A_1, A_2\}$, $\Sigma = \{B\}$. Then $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ but $\mathcal{T}_1$ does not (Γ, Σ)-CQ entail $\mathcal{T}_2$ with inconsistent ABoxes.*

The following lemma relates the two notions of CQ entailment. *CQ evaluation* is the problem to decide, given a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, a CQ q, and a tuple $\mathbf{a} \in \mathsf{ind}(\mathcal{A})$, whether $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$.

Lemma 4 *CQ entailment with inconsistent ABoxes can be decided in polynomial time given access to oracles deciding CQ entailment and CQ evaluation.*

Consequently and since CQ evaluation is in ExpTime in Horn-$\mathcal{ALCHIF}$ [Eiter *et al.*, 2008], all complexity results obtained in this paper also apply to CQ entailment with inconsistent ABoxes.

It is easy to see that $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ if there is a Γ-role r and a Σ-role s with $\mathcal{T}_2 \models r \sqsubseteq s$ but $\mathcal{T}_1 \not\models r \sqsubseteq s$. We write $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ if there are no such r and s. Clearly, $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ can be decided via $|\Gamma| \cdot |\Sigma|$ many Horn-$\mathcal{ALCHIF}$ subsumption tests, thus in ExpTime [Tobies, 2001]. It is thus safe to assume $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ when deciding CQ entailment, which we will generally do from now on to avoid dealing with special cases.

2.3 Deductive Conservative Extensions

Another natural notion of entailment is deductive entailment, which generalizes the notion of deductive conservative extensions [Ghilardi *et al.*, 2006; Lutz *et al.*, 2007; Konev *et al.*, 2009; Lutz and Wolter, 2010], and which separates two TBoxes in terms of concept and role inclusions and functionality assertions, instead of ABoxes and queries.

Definition 5 *Let Σ be a signature and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be $\mathcal{ELHIF}_\bot$ TBoxes. We say that $\mathcal{T}_1$* Σ-deductively entails *$\mathcal{T}_2$, written $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$, if for all Σ-$\mathcal{ELI}_\bot$-concept inclusions α and all Σ-RIs and Σ-FAs α: $\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$. If additionally $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then we say that $\mathcal{T}_2$* is a Σ-deductive conservative extension *of $\mathcal{T}_1$. If $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$ and vice versa, then $\mathcal{T}_1$ and $\mathcal{T}_2$ are* Σ-deductively inseparable.

Although closely related, it is not difficult to see that deductive and query entailment are orthogonal.

Example 6 *(1) Let $\mathcal{T}_1$, $\mathcal{T}_2$ be as in Example 3 and $\Sigma = \{A_1, A_2, B\}$. Then $\mathcal{T}_1 \models^{\mathrm{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$, but $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$.*

(2) Let $\mathcal{T}_1 = \emptyset$ and $\mathcal{T}_2 = \{A \sqsubseteq \exists r.B\}$, and $\Sigma = \{A, B\}$. Then $\mathcal{T}_1 \models^{\mathrm{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$, but $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Sigma,\Sigma} \mathcal{T}_2$ as witnessed by $(\{A(a)\}, \exists x B(x), a)$. However, $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$.

Nevertheless, the two notions are sufficiently closely related so that we have the following.

Lemma 7 *In $\mathcal{ELHIF}_\bot$, deductive entailment can be decided in polynomial time given access to oracles for stCQ entailment and stCQ evaluation.*

2.4 Homomorphisms and the Universal Model

For interpretations $\mathcal{I}_1, \mathcal{I}_2$ and a signature Σ, a Σ-*homomorphism* from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a total function $h : \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ such that (1) $h(a) = a$ for all $a \in \mathsf{N_I}$, (2) $h(d) \in A^{\mathcal{I}_2}$ for all $d \in A^{\mathcal{I}_1}$, $A \in \mathsf{N_C} \cap \Sigma$, and (3) $(h(d), h(d')) \in r^{\mathcal{I}_2}$ for all $(d, d') \in r^{\mathcal{I}_1}$, $r \in \mathsf{N_R} \cap \Sigma$. If there is a Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$, we write $\mathcal{I}_1 \to_\Sigma \mathcal{I}_2$.

Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox in normal form and $\mathcal{A}$ an ABox consistent with $\mathcal{T}$. A *type for* $\mathcal{T}$ is a set $t \subseteq \mathsf{sub}(\mathcal{T}) \cap \mathsf{N_C}$ such that $\mathcal{T} \models \bigsqcap t \sqsubseteq A$ implies $A \in t$ for all concept names A. For $a \in \mathsf{ind}(\mathcal{A})$, let $\mathsf{tp}_\mathcal{T}(a) = \{A \mid \mathcal{T}, \mathcal{A} \models A(a)\}$ be the *type of a relative to* $\mathcal{T}$. When $a \in \mathsf{ind}(\mathcal{A})$, t, t' are types for $\mathcal{T}$, and r is a role, we write

- $a \rightsquigarrow^{\mathcal{T},\mathcal{A}}_r t$ if $\mathcal{T}, \mathcal{A} \models \exists r. \bigsqcap t(a)$ and t is maximal with this condition, and
- $t \rightsquigarrow^{\mathcal{T}}_r t'$ if $\mathcal{T} \models \bigsqcap t \sqsubseteq \exists r. \bigsqcap t'$ and t' is maximal with this condition.

A *path for $\mathcal{A}$ and $\mathcal{T}$* is a finite sequence $\pi = a r_0 t_1 \cdots t_{n-1} r_{n-1} t_n$, $n \geq 0$, with $a \in \mathsf{ind}(\mathcal{A})$, $r_0, \ldots, r_{n-1}$ roles, and $t_1, \ldots, t_n$ types for $\mathcal{T}$ such that

(i) $a \rightsquigarrow^{\mathcal{T},\mathcal{A}}_{r_0} t_1$ and, if $\mathsf{func}(r_0) \in \mathcal{T}$, then there is no $b \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}, \mathcal{A} \models r_0(a, b)$;

(ii) for every $1 \leq i < n$, we have $t_i \rightsquigarrow^{\mathcal{T}}_{r_i} t_{i+1}$ and, if $\mathsf{func}(r) \in \mathcal{T}$, then $r_{i-1} \neq r_i^-$.

When $n > 0$, we use $\mathsf{tail}(\pi)$ to denote t_n. Let Paths be the set of all paths for $\mathcal{A}$ and $\mathcal{T}$. The *universal model* $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ of $\mathcal{T}$ and $\mathcal{A}$ is defined as follows:

$$
\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} &= \mathsf{Paths} \\
A^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} &= \{a \in \mathsf{ind}(\mathcal{A}) \mid \mathcal{T}, \mathcal{A} \models A(a)\} \cup \\
&\quad\ \{\pi \in \Delta^{\mathcal{I}} \setminus \mathsf{ind}(\mathcal{A}) \mid \mathcal{T} \models \textstyle\bigsqcap \mathsf{tail}(\pi) \sqsubseteq A\} \\
r^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} &= \{(a, b) \in \mathsf{ind}(\mathcal{A})^2 \mid s(a, b) \in \mathcal{A}, \mathcal{T} \models s \sqsubseteq r\} \cup \\
&\quad\ \{(\pi, \pi s t) \mid \pi s t \in \mathsf{Paths} \text{ and } \mathcal{T} \models s \sqsubseteq r\} \cup \\
&\quad\ \{(\pi s t, \pi) \mid \pi s t \in \mathsf{Paths} \text{ and } \mathcal{T} \models s^- \sqsubseteq r\}
\end{aligned}
$$

It is standard to prove that $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is indeed a model of $\mathcal{T}$ and $\mathcal{A}$ and that it is universal in the sense that for every model $\mathcal{I}$

of $\mathcal{T}$ and $\mathcal{A}$, we have $\mathcal{I} \to \mathcal{I}_{\mathcal{T},\mathcal{A}}$. Consequently, $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ iff $\mathcal{I}_{\mathcal{T},\mathcal{A}} \models q(\mathbf{a})$, for all CQs $q(\mathbf{x})$ and tuples $\mathbf{a}$ of individuals.

We also need universal models of a TBox $\mathcal{T}$ and a type t, instead of an ABox. More precisely, we define $\mathcal{I}_{\mathcal{T},t} = \mathcal{I}_{\mathcal{T},\mathcal{A}_t}$ where $\mathcal{A}_t = \{A(a) \mid A \in t\}$ for a fixed $a \in \mathsf{N_I}$.

3 Model-theoretic Characterization

We aim to provide a model-theoretic characterization of query entailment that will be the basis for our decision procedure later on. The first step towards this characterization consists in showing that non-entailment is always witnessed by tree-shaped ABoxes and tree-shaped CQs with at most one answer variable. Here, tree-shaped ABoxes $\mathcal{A}$ are defined analogously to tree-shaped CQs: $\mathcal{A}$ is *tree-shaped* if the undirected graph $G_\mathcal{A} = (\mathsf{ind}(\mathcal{A}), \{\{a, b\} \mid r(a, b) \in \mathcal{A}\})$ is a tree.

Lemma 8 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\text{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. If $\mathcal{T}_1 \not\models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is a witness $(\mathcal{A}, q, \mathbf{a})$ where $\mathcal{A}$ and q are tree-shaped and $|\mathbf{a}| \leq 1$, i.e., q has at most one answer variable. If $\mathcal{T}_1 \not\models^{\text{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is such a witness where additionally q is an stCQ.*

Our goal is to characterize query entailment in terms of homomorphisms between (universal) models. Homomorphisms are natural because answers to CQs are preserved under homomorphisms (both on interpretations and on ABoxes). In fact, they are preserved even under bounded homomorphisms if the bound is not smaller than the number of variables in the CQ.

Let $\mathcal{I}_1, \mathcal{I}_2$ be interpretations, $d \in \Delta^{\mathcal{I}_1}$, and $n \geq 0$. We say that there is an *n-bounded Σ-homomorphism* from $\mathcal{I}_1$ to $\mathcal{I}_2$, written $\mathcal{I}_1 \to^n_\Sigma \mathcal{I}_2$, if for any subinterpretation $\mathcal{I}'_1$ of $\mathcal{I}_1$ with $|\Delta^{\mathcal{I}'_1}| \leq n$, we have $\mathcal{I}'_1 \to_\Sigma \mathcal{I}_2$. Moreover, we write $\mathcal{I}_1 \to^{\mathsf{fin}}_\Sigma \mathcal{I}_2$ if $\mathcal{I}_1 \to^n_\Sigma \mathcal{I}_2$ for any n. The following characterization follows from the definition of CQ entailment, Lemma 8, and the connection between CQs and suitably bounded homomorphisms.

Lemma 9 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\text{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.*

Ideally, we would like to use Lemma 9 as a basis for a decision procedure based on tree automata. To this end, it is useful that the ABox $\mathcal{A}$ and models $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ in the lemma are tree-shaped. What is problematic is that Lemma 9 speaks about unbounded homomorphisms, for *any* bound (corresponding to the unbounded size of CQs), since it does not seem possible to verify such a condition using automata. We would thus like to replace bounded homomorphisms with unbounded ones, which does not compromise the characterization in the case of Horn-DLs without inverse roles [Lutz and Wolter, 2010; Botoeva *et al.*, 2016c]. However, this is not true already for $\mathcal{ELI}$ TBoxes [Botoeva *et al.*, 2016a]:

Example 10 *Let $\mathcal{T}_1 = \{A \sqsubseteq \exists s.B,\ B \sqsubseteq \exists r^-.B\}$, $\mathcal{T}_2 = \{A \sqsubseteq \exists s.B,\ B \sqsubseteq \exists r.B\}$, $\Gamma = \{A\}$, and $\Sigma = \{r\}$. Then both $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ contain an infinite r-path; the r-path in $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ has a final element while the one in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ does not. Hence $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$, but $\mathcal{T}_1 \models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ (see Thm. 11 below).*

We now show that it is possible to refine Lemma 9 so that it makes a much more careful statement in which bounded homomorphisms are *partly* replaced by unbounded ones. It is then possible to check the unbounded homomorphism part of the characterization using tree automata as desired, and to deal with unbounded homomorphisms using a mosaic technique that "precompiles" relevant information about unbounded homomorphisms to be used in the automaton construction.

We start with introducing relevant notation. For a signature Σ, we use $\mathcal{I}|^{\mathsf{con}}_\Sigma$ to denote the restriction of the interpretation $\mathcal{I}$ to those elements that can be reached from an ABox individual by traveling along Σ-roles (forwards or backwards). For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, and $a \in \mathsf{ind}(\mathcal{A})$, we use $\mathcal{I}_{\mathcal{T},\mathcal{A}}|_a$ to denote the subtree interpretation in the universal model $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ rooted at a. A *Σ-subtree in $\mathcal{I}_{\mathcal{T},\mathcal{A}}$* is a maximal tree-shaped, Σ-connected sub-interpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ that does not comprise any ABox individuals. The *root* of $\mathcal{I}$ is the (unique) element of $\Delta^\mathcal{I}$ that can be reached from an ABox individual on a shortest path among all element of $\Delta^\mathcal{I}$. The refined characterization uses simulations instead of homomorphisms for the stCQ case because they are insensitive to multi-edges. Given a signature Σ and two interpretations $\mathcal{I}, \mathcal{J}$, a *Σ-simulation of $\mathcal{I}$ in $\mathcal{J}$* is a relation $\sigma \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{J}$ such that: (1) $(a, a) \in \sigma$ for all $a \in \mathsf{N_I}$, (2) if $d \in A^\mathcal{I}$ with $A \in \Sigma$ and $(d, e) \in \sigma$, then $d_2 \in A^\mathcal{J}$, and (3) if $(d, d') \in r^\mathcal{I}$ with r a Σ-role and $(d, e) \in \sigma$, then there is some e' with $(e, e') \in r^\mathcal{J}$ and $(d', e') \in \sigma$. We write $\mathcal{I} \preceq_\Sigma \mathcal{J}$ if there is a Σ-simulation of $\mathcal{I}$ in $\mathcal{J}$.

Theorem 11 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\text{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and for all tree-shaped, finitely branching models $\mathcal{I}_1$ of $\mathcal{A}$ and $\mathcal{T}_1$, the following hold:*

(1) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \to_\Sigma \mathcal{I}_1$;

(2) for all Σ-subtrees $\mathcal{I}$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, one of the following holds:

- *(a) $\mathcal{I} \to_\Sigma \mathcal{I}_1$;*
- *(b) $\mathcal{I} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_1}(a)}$ for some $a \in \mathsf{ind}(\mathcal{A})$.*

Furthermore, $\mathcal{T}_1 \models^{\text{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all $\mathcal{A}$ and $\mathcal{I}_1$ as above, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_1$.

4 Decidability and Complexity

We prove that, in Horn-$\mathcal{ALCHIF}$, CQ entailment can be decided in 2ExpTime. By existing lower bounds, the former is thus 2ExpTime-complete in all fragments of Horn-$\mathcal{ALCHIF}$ that contain $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$. Moreover, stCQ entailment in Horn-$\mathcal{ALCHIF}$ and deductive entailment in $\mathcal{ELHIF}_\bot$ can also be decided in 2ExpTime. We establish a coNExpTime lower bound and leave the precise complexity open.

To obtain the upper bounds, we use a combination of tree automata and mosaics to implement the characterization in Theorem 11. We start with a mosaic-based decision procedure for Condition (2b). Note that a Σ-subtree $\mathcal{I}$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ can be uniquely identified by the type t_2 of its root. It therefore suffices to show the following.

Theorem 12 *Given two Horn-$\mathcal{ALCHIF}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and types t_i for $\mathcal{T}_i$, $i \in \{1, 2\}$, it can be decided in time $2^{2^{p(|\mathcal{T}_2| \log |\mathcal{T}_1|)}}$ whether $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,t_1}$, p a polynomial.*

Although we cannot get rid of bounded homomorphisms in Theorem 11, a central idea for applying a mosaic approach to prove Theorem 12 is to first replace bounded homomorphisms with unbounded ones. To make this possible, we also replace $\mathcal{I}_{\mathcal{T}_1,t_1}$ with a suitable class of interpretations used as targets for the unbounded homomorphisms.

To illustrate, consider Example 10 and let $t_1 = t_2 = \{B\}$. The difference between $\mathcal{I}_{\mathcal{T}_2,t_2} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$ and $\mathcal{I}_{\mathcal{T}_2,t_2} \rightarrow_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$ is that unbounded homomorphisms fail once they "reach the root" of $\mathcal{I}_{\mathcal{T}_1,t_1}$ while bounded homomorphisms can, depending on the bound, map the root of $\mathcal{I}_{\mathcal{T}_2,t_2}$ deeper and deeper into $\mathcal{I}_{\mathcal{T}_1,t_1}$, thus never reaching its root. The latter is possible because $\mathcal{I}_{\mathcal{T}_1,t_1}$ is regular in the sense that any two elements which have the same type root isomorphic subtrees. This is of course not only true in this example, but by construction in any universal model. To transition back from bounded to unbounded homomorphisms, we replace $\mathcal{I}_{\mathcal{T}_1,t_1}$ with a class of (finite and infinite) interpretations that can be seen as a "backwards regularization" of $\mathcal{I}_{\mathcal{T}_1,t_1}$. In our concrete example, we would include an interpretation where a predecessor is added to the root of $\mathcal{I}_{\mathcal{T}_1,t_1}$ because $\mathcal{I}_{\mathcal{T}_1,t_1}$ contains an element of the same type as the root that has such a predecessor, an interpretation where that predecessor has a predecessor, and so on, even ad infinitum. We will now make this precise.

An interpretation $\mathcal{I}$ is *quasi tree-shaped* if:

1. $\Delta^{\mathcal{I}} \subseteq (\{-1\} \cup \mathbb{N})^*$;
2. $(d, e) \in r^{\mathcal{I}}$ implies that $e = d \cdot c$ or $d = e \cdot c$ for some $c \in \{-1\} \cup \mathbb{N}$.

For $d, e \in \Delta^{\mathcal{I}}$, we say that e is a *successor* of d if $e = d \cdot c$ for some $c \in \mathbb{N}$ or $d = e \cdot -1$. By this convention, quasi tree-shaped interpretations can be viewed as directed graphs. The directedness does not correspond to the distinction between roles and inverse roles; in particular, there can be several role edges in both directions between the same d and e. Quasi tree-shaped interpretations can be viewed as a finite or infinite trees that need not have a root as they can extend indefinitely not only downwards but also upwards.

Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox and let $\mathsf{tp}(\mathcal{T})$ be the set of all types for $\mathcal{T}$ consistent with $\mathcal{T}$. For every $t_0 \in \mathsf{tp}(\mathcal{T})$, we use $\mathsf{tp}(\mathcal{T}, t_0)$ to denote the set of all $t \in \mathsf{tp}(\mathcal{T})$ that occur in the universal model $\mathcal{I}_{\mathcal{T},t_0}$ of t_0 and $\mathcal{T}$. Furthermore, given a quasi tree-shaped interpretation $\mathcal{I}$ and an element $d \in \Delta^{\mathcal{I}}$, the *1-neighborhood of d in $\mathcal{I}$* is a tuple $n_1^{\mathcal{I}}(d) = (t^-, \rho, t, S)$ such that (a) $t = \mathsf{tp}_{\mathcal{I}}(d)$; (b) if there is a predecessor $d_0 \in \Delta^{\mathcal{I}}$ of d, then $t^- = \mathsf{tp}_{\mathcal{I}}(d_0)$ and $\rho = \{r \mid (d_0, d) \in r^{\mathcal{I}}\}$, otherwise $\rho = t^- = \bot$; (c) S is the set of all pairs (ρ', t') such that there is a successor d' of d with $(d, d') \in \rho'$ (ρ' maximal) and $t' = \mathsf{tp}_{\mathcal{I}}(d')$. We write $(t_1^-, \rho_1, t_1, S_1) \sqsubseteq (t_2^-, \rho_2, t_2, S_2)$ if $t_1 = t_2$, $S_1 \subseteq S_2$ and, if $\rho_1 \neq \bot$, then $\rho_1 = \rho_2$ and $t_1^- = t_2^-$.

In the following, we define a class $\mathsf{can}_\omega(\mathcal{T}, t_0)$ of quasi tree-shaped models of $\mathcal{T}$. To construct a model from this class, choose a type $t \in \mathsf{tp}(\mathcal{T}, t_0)$ and define $\mathcal{I} = (\{d_0\}, \cdot^{\mathcal{I}})$ such that $\mathsf{tp}_{\mathcal{I}}(d_0) = t$. Then extend $\mathcal{I}$ by applying the following rule exhaustively in a fair way:

(R) Let $d \in \Delta^{\mathcal{I}}$. Choose some $e \in \Delta^{\mathcal{I}_{\mathcal{T},t_0}}$ such that $n_1^{\mathcal{I}}(d) \sqsubseteq n_1^{\mathcal{I}_{\mathcal{T},t_0}}(e)$, and add to d the predecessor and/or successors required to achieve $n_1^{\mathcal{I}}(d) = n_1^{\mathcal{I}_{\mathcal{T},t_0}}(e)$.

The potentially infinite class $\mathsf{can}_\omega(\mathcal{T}, t_0)$ is the set of all interpretations $\mathcal{I}$ obtained as a limit of this construction.

Lemma 13 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox, $t_0 \in \mathsf{tp}(\mathcal{T})$, and $\mathcal{I}$ a tree-shaped interpretation. Then $\mathcal{I} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T},t_0}$ iff there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}, t_0)$ with $\mathcal{I} \rightarrow_{\Sigma} \mathcal{J}$.*

We can now use Lemma 13 to devise the mosaic-based procedure for deciding the existence of a bounded homomorphism. Let $\mathcal{T}_1, \mathcal{T}_2$ be as in Theorem 11. We denote with $\mathsf{rol}(\mathcal{T}_i)$ the set of all roles r, r^- such that the (possibly inverse) role r occurs in $\mathcal{T}_i$. Moreover, for a set of roles ρ, denote with $\rho|_{\Sigma}$ the restriction of ρ to Σ-roles.

Fix now some $t_1 \in \mathsf{tp}(\mathcal{T}_1)$. Intuitively, a mosaic for t_1 represents a possible 1-neighborhood of some element in $\mathcal{I}_{\mathcal{T}_1,t_1}$ together with a decoration with sets of types for $\mathcal{T}_2$ that can be homomorphically embedded into the neighborhood. Formally, a *mosaic for* t_1 is a tuple $M = (t^-, \rho, t, S, \ell)$ such that $(t^-, \rho, t, S) = n_1^{\mathcal{I}_{\mathcal{T}_1,t_1}}(d)$ for some $d \in \Delta^{\mathcal{I}_{\mathcal{T}_1,t_1}}$ and $\ell : \{t^-, t\} \cup S \rightarrow 2^{\mathsf{tp}(\mathcal{T}_2)}$ satisfies the following condition:

(M) For all $\widehat{t} \in \ell(t)$ we have $\widehat{t} \cap \Sigma \subseteq t$ and, for all $\widehat{t'} \in \mathsf{tp}(\mathcal{T}_2)$, $r \in \mathsf{rol}(\mathcal{T}_2)$ with $\widehat{t} \rightsquigarrow^{\mathcal{T}_2}_{r} \widehat{t'}$, one the following holds for $\sigma = \{s \mid \mathcal{T} \models r \sqsubseteq s\}$:

(a) $\sigma|_{\Sigma} = \emptyset$;

(b) $t^- \neq \bot$, $\sigma|_{\Sigma} \subseteq \rho^-$, and $\widehat{t'} \in \ell(t^-)$;

(c) there is $(\rho', t') \in S$ with $\widehat{t'} \in \ell(\rho', t')$ and $\sigma|_{\Sigma} \subseteq \rho'$.

To ease notation, we use t_M^- to denote t^-, ρ_M to denote ρ, and likewise for the other components of a mosaic M. Let $\mathcal{M}$ be the set of all mosaics for t_1 and $\mathcal{M}' \subseteq \mathcal{M}$. An $M \in \mathcal{M}'$ is *good in* $\mathcal{M}'$ if the following conditions are satisfied:

1. for each $(\rho, t) \in S_M$, there is an $N \in \mathcal{M}'$ such that $(t_M, \rho, t) = (t_N^-, \rho_N, t_N)$, $\ell_M(\rho, t) = \ell_N(t_N)$, and $\ell_M(t_M) = \ell_N(t_N^-)$.
2. if $t_M^- \neq \bot$, there is $N \in \mathcal{M}'$ with $(\rho_M, t_M) \in S_N$, $t_M^- = t_N$, $\ell_M(t_M^-) = \ell_N(t_N)$, and $\ell_M(t_M) = \ell_N(\rho_M, t_M)$.

Let $\mathcal{M}_0, \mathcal{M}_1, \ldots$ be the sequence obtained by starting with $\mathcal{M}_0 = \mathcal{M}$ and defining $\mathcal{M}_{i+1}$ to be $\mathcal{M}_i$ when all mosaics that are not good in $\mathcal{M}_i$ have been removed. Assume that $\mathcal{M}_p$ is where the sequence stabilizes.

Lemma 14 *Let $t_i \in \mathsf{tp}(\mathcal{T}_i)$ for $i \in \{1, 2\}$. Then there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}_1, t_1)$ such that $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow_{\Sigma} \mathcal{J}$ iff $\mathcal{M}_p$ contains a mosaic M with $t_2 \in \ell_M(t_M)$.*

Since there are at most $2^{|\mathcal{T}_1| 2^{|\mathcal{T}_2|}}$ mosaics for t_1, we obtain the desired Theorem 12.

We now develop the decision procedure for CQ and stCQ entailment in Horn-$\mathcal{ALCHIF}$, based on Theorems 11 and 12. Our main tool are *alternating two-way tree automata with counting (2ATA$_c$)*, an extension of alternating tree automata over *unranked* trees [Grädel and Walukiewicz, 1999] with the ability to count. A *tree* is a non-empty (potentially infinite) set of words $T \subseteq (\mathbb{N} \setminus 0)^*$ closed under prefixes. We assume that trees are finitely branching, i.e., for every $w \in T$, the set $\{i \mid w \cdot i \in T\}$ is finite. For any $w \in (\mathbb{N} \setminus 0)^*$, we set $w \cdot 0 := w$. If $w = n_0 n_1 \cdots n_k$, $k \geq 0$, we set $w \cdot -1 := n_0 \cdots n_{k-1}$. For an alphabet Θ, a Θ-*labeled tree* is a pair (T, L) with T a tree and $L : T \rightarrow \Theta$ a node labeling function.

A *2ATA*$_c$ is a tuple $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ where Q is a finite set of *states*, Σ is the *input alphabet*, $q_0 \in Q$ is the *initial state*, δ is a *transition function*, and $\Omega : Q \rightarrow \mathbb{N}$ is a *priority function*. The transition function δ maps every state q and input letter $a \in \Theta$ to a positive Boolean formula $\delta(q, a)$ over the truth constants true and false and *transition atoms* of the form q, $\langle - \rangle q$, $[-]q$, $\Diamond_n q$ and $\Box_n q$. Informally, a transition q expresses that a copy of $\mathfrak{A}$ is sent to the current node in state q; $\langle - \rangle q$ means that a copy is sent in state q to the predecessor node, which is required to exist; $[-]q$ means the same except that the predecessor node is not required to exist; $\Diamond_n q$ (resp., $\Box_n q$) means that a copy of q is sent to k (resp., to all but k) successors. The semantics of 2ATA$_c$ is given in terms of runs as usual, please see the appendix. We use $L(\mathfrak{A})$ to denote the set of trees accepted by $\mathfrak{A}$. It is standard to verify closure of 2ATA$_c$ under intersection. The following is obtained via reduction to standard alternating parity tree automata [Vardi, 1998].

Theorem 15 *The emptiness problem for 2ATA$_c$ can be solved in time exponential in the number of states.*

Let $\mathcal{T}_1, \mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes and Γ, Σ signatures. We aim to show that one can construct a 2ATA$_c$ $\mathfrak{A}$ such that $L(\mathfrak{A}) = \emptyset$ iff $\mathcal{T}_1 \not\models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$. In fact, $\mathfrak{A}$ is the intersection of four 2ATA$_c$ $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3, \mathfrak{A}_4$. They run over Θ-labeled trees with $\Theta = 2^{\Theta_0} \times 2^{\Theta_1} \times 2^{\Theta_2}$, where $\Theta_0 = \Gamma \cup \{r^- \mid r \in \Gamma\}$ and $\Theta_i = \mathsf{sig}(\mathcal{T}_i) \cup \{r^- \mid r \in \mathsf{sig}(\mathcal{T}_i)\}$ for $i = 1, 2$. For a Θ-labeled tree (T, L), we use L_i, $i \in \{0, 1, 2\}$ to refer to the i-th component of L, that is, $L(n) = (L_0(n), L_1(n), L_2(n))$, for all $n \in T$. The component L_0 represents a (possibly infinite) ABox $\mathcal{A} = \{A(n) \mid A \in L_0(n)\} \cup \{r(n \cdot -1, n) \mid n \neq \varepsilon, r \in L_0(n)\}$, where $r^-(a, b)$ is identified with $r(b, a)$. The 2ATA$_c$ $\mathfrak{A}_1$ accepts a Θ-labeled tree (T, L) iff $\mathcal{A}$ is finite, connected and includes the root of T (and is thus tree-shaped), and it is straightforward to construct.

Components L_1, L_2 give rise to interpretations $\mathcal{I}_1 = (T, \cdot^{\mathcal{I}_1})$ and $\mathcal{I}_2 = (\mathsf{ind}(\mathcal{A}), \cdot^{\mathcal{I}_2})$, where for $i \in \{1, 2\}$:

$$A^{\mathcal{I}_i} = \{n \mid A \in L_i(n)\}$$

$$r^{\mathcal{I}_i} = \{(n, n \cdot -1) \mid r^- \in L_i(n)\} \cup \{(n \cdot -1, n) \mid r \in L_i(n)\}$$

$\mathfrak{A}_2$ verifies that $\mathcal{I}_1$ is a model of $\mathcal{A}$ and $\mathcal{T}_1$, which is standard, too. $\mathfrak{A}_3$ verifies that $\mathcal{A}$ is consistent with $\mathcal{T}_2$, and $\mathcal{I}_2$ is $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ restricted to $\mathsf{ind}(\mathcal{A})$. This involves computing the type of an ABox element without having access to the anonymous (that is: non-ABox) part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, using a characterization of ABox entailments [Bienvenu *et al.*, 2013] in terms of derivation trees. Finally, $\mathfrak{A}_4$ verifies that either (1) or (2) from Theorem 11 is *not* satisfied. For (1), $\mathfrak{A}_4$ sends a copy of itself to every tree $\mathcal{I}$ starting at an ABox element in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, and attempts to show that $\mathcal{I}$ cannot be homomorphically embedded into a corresponding tree in $\mathcal{I}_1$. This attempt is successful if either incompatible types are found in the root or, recursively, there is some successor of the current type in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ that cannot be mapped to any neighbor in $\mathcal{I}_1$. Since the anonymous part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not explicit in the input, the current type is stored in the states, and the generating relation $t \rightsquigarrow^{\mathcal{T}_2}_r t'$ is "hard-coded" into the transition function. For Condition (2a), $\mathfrak{A}_4$ non-deterministically guesses a Σ-subtree $\mathcal{I}$ and proceeds as in (1); Condition (2b) is verified based on Theorem 12 by pre-computing $\rightarrow^{\mathsf{fin}}_{\Sigma}$. Thus the number of states of $\mathfrak{A}_4$ is exponential in $\mathcal{T}_2$ (because of the types) but only polynomial in $|\mathcal{T}_1|$. Automata $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ have polynomially many states.

In the special case of stCQ entailment, we simply replace $\mathfrak{A}_4$ with a 2ATA$_c$ $\mathfrak{A}'_4$ that refutes the simulation condition of Theorem 11 analogously to how $\mathfrak{A}_4$ refutes Condition (1).

To obtain the desired upper complexity bounds for query and deductive entailment, we observe that, in both cases, $\mathfrak{A}$ can be constructed in time polynomial in $|\mathcal{T}_1|$ and exponential in $|\mathcal{T}_2|$, and the emptiness check adds an exponential blowup (Theorem 15). For deductive entailment, we use the reduction to stCQ entailment (Lemma 7).

Theorem 16 *In Horn-$\mathcal{ALCHIF}$, the following problems can be decided in time $2^{2^{p(|\mathcal{T}_2| \log |\mathcal{T}_1|)}}$, p a polynomial: (Γ, Σ)-CQ entailment, (Γ, Σ)-CQ inseparability, and (Γ, Σ)-CQ conservative extensions. The same holds for Σ-deductive entailment, Σ-deductive inseparability, and Σ-deductive conservative extensions in $\mathcal{ELHIF}_\bot$.*

Matching lower bounds for all problems except deductive entailment are provided by [Botoeva *et al.*, 2016c]. They hold even in the case where $\Gamma = \Sigma$.

Corollary 17 *In any fragment of Horn-$\mathcal{ALCHIF}$ that contains $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$, the following problems are* 2EXPTIME*-complete: (Γ, Σ)-CQ entailment, (Γ, Σ)-CQ inseparability, and (Γ, Σ)-CQ conservative extensions.*

In the description logic $\mathcal{EL}$, which is $\mathcal{ELI}$ without inverse roles, deductive conservative extensions and deductive Σ-entailment are EXPTIME-complete [Lutz and Wolter, 2010]. This raises the question whether the upper bound for deductive entailment reported in Theorem 16 is tight. While we leave this question open, we observe that the transition from $\mathcal{EL}$ to $\mathcal{ELI}$ does increase the complexity of deductive conservative extensions and related problems to at least CONEXPTIME. We consider this a surprising result since in reasoning problems that are not defined in terms of conjunctive queries, adding inverse roles does typically not result in an increase of complexity. The following is established by a non-trivial reduction of a tiling problem.

Theorem 18 *In $\mathcal{ELI}$ and in $\mathcal{ELHIF}_\bot$, deductive conservative extensions, deductive Σ-entailment, and deductive Σ-inseparability are* CONEXPTIME*-hard.*

5 Conclusion

As future work, it would be interesting to close the gap in complexity between CONEXPTIME and 2EXPTIME for deductive entailment in $\mathcal{ELI}$ and $\mathcal{ELHIF}_\bot$. Furthermore, it would be interesting to extend the results to ontology languages from the family of Datalog+/- (aka existential rules), in particular to frontier-guarded TGDs.

Acknowledgments

This work was partially supported by DFG project LU 1417/2 and ERC Consolidator Grant 647289 CODA.

References

[Arenas *et al.*, 2016] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Knowledge base exchange: The case of OWL 2 QL. *Artif. Intell.*, 238:11–62, 2016.

[Baader *et al.*, 2017] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logics*. Cambride University Press, 2017.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Proc. RW*, pages 218–307, 2015.

[Bienvenu and Rosati, 2015] Meghyn Bienvenu and Riccardo Rosati. Query-based comparison of OBDA specifications. In *Proc. DL*, volume 1350. ceur-ws.org, 2015.

[Bienvenu *et al.*, 2013] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. First order-rewritability of atomic queries in Horn description logics. In *Proc. IJCAI*, pages 754–760, 2013.

[Bienvenu *et al.*, 2016] Meghyn Bienvenu, Peter Hansen, Carsten Lutz, and Frank Wolter. First order-rewritability and containment of conjunctive queries in Horn description logics. In *Proc. IJCAI*, pages 965–971, 2016.

[Botoeva *et al.*, 2016a] Elena Botoeva, Boris Konev, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Inseparability and conservative extensions of description logic ontologies: A survey. In *Proc. RW*, 2016.

[Botoeva *et al.*, 2016b] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Games for query inseparability of description logic knowledge bases. *Artif. Intell.*, 234:78–119, 2016.

[Botoeva *et al.*, 2016c] Elena Botoeva, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query-based entailment and inseparability for $\mathcal{ALC}$ ontologies. In *Proc. IJCAI*, pages 1001–1007, 2016.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. Autom. Reas.*, 39(3):385–429, 2007.

[Eiter *et al.*, 2008] Thomas Eiter, Georg Gottlob, Magdalena Ortiz, and Mantas Šimkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. JELIA*, volume 5293 of *LNCS*, pages 166–179. Springer, 2008.

[Eiter *et al.*, 2012] Thomas Eiter, Magdalena Ortiz, Mantas Šimkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *Proc. AAAI*, 2012.

[Ghilardi *et al.*, 2006] Silvio Ghilardi, Carsten Lutz, and Frank Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proc. KR*, pages 187–197, 2006.

[Grädel and Walukiewicz, 1999] Erich Grädel and Igor Walukiewicz. Guarded fixed point logic. In *Proc. LICS*, pages 45–54, 1999.

[Hustadt *et al.*, 2007] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reasoning in description logics by a reduction to disjunctive datalog. *J. Autom. Reasoning*, 39(3), 2007.

[Ibáñez-García *et al.*, 2014] Yazmín Ibáñez-García, Carsten Lutz, and Thomas Schneider. Finite model reasoning in Horn description logics. In *Proc. KR*, 2014.

[Kazakov, 2009] Yevgeny Kazakov. Consequence-driven reasoning for Horn-$\mathcal{SHIQ}$ ontologies. In *Proc. IJCAI*, pages 2040–2045, 2009.

[Kollia and Glimm, 2013] Ilianna Kollia and Birte Glimm. Optimizing SPARQL query answering over OWL ontologies. *J. Artif. Intell. Res.*, 48:253–303, 2013.

[Konev *et al.*, 2009] Boris Konev, Carsten Lutz, Dirk Walther, and Frank Wolter. Formal properties of modularisation. In H. Stuckenschmidt, S. Spacciapietra, and C. Parent, editors, *Modular Ontologies*, volume 5445 of *LNCS*, pages 25–66. Springer, 2009.

[Konev *et al.*, 2012] Boris Konev, Michel Ludwig, Dirk Walther, and Frank Wolter. The logical difference for the lightweight description logic $\mathcal{EL}$. *J. Artif. Intell. Res.*, 44:633–708, 2012.

[Kontchakov *et al.*, 2009] Roman Kontchakov, Luca Pulina, Ulrike Sattler, Thomas Schneider, P. Selmer, Frank Wolter, and Michael Zakharyaschev. Minimal module extraction from DL-Lite ontologies using QBF solvers. In *Proc. IJCAI*, pages 836–840, 2009.

[Kontchakov *et al.*, 2010] Roman Kontchakov, Frank Wolter, and Michaek Zakharyaschev. Logic-based ontology comparison and module extraction, with an application to DL-Lite. *Artif. Intell.*, 174:1093–1141, 2010.

[Krötzsch *et al.*, 2007] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexity boundaries for Horn description logics. In *Proc. AAAI*, pages 452–457, 2007.

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. Symb. Comput.*, 45(2):194–228, 2010.

[Lutz and Wolter, 2012] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. KR*, 2012.

[Lutz *et al.*, 2007] Carsten Lutz, Dirk Walther, and Frank Wolter. Conservative extensions in expressive description logics. In *Proc. IJCAI*, pages 453–458, 2007.

[Matentzoglu and Parsia, 2017] Nico Matentzoglu and Bijan Parsia. BioPortal Snapshot 30 March 2017 (data set), 2017. `http://doi.org/10.5281/zenodo.439510`.

[Poggi *et al.*, 2008] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. 10:133–173, 2008.

[Tobies, 2001] Stephan Tobies. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation*. PhD thesis, RWTH Aachen, 2001.

[Trivela *et al.*, 2015] Despoina Trivela, Giorgos Stoilos, Alexandros Chortaras, and Giorgos B. Stamou. Optimising resolution-based rewriting algorithms for OWL ontologies. *J. Web Sem.*, 33:30–49, 2015.

[Vardi, 1998] Moshe Y. Vardi. Reasoning about the past with two-way automata. In *Proc. ICALP*, pages 628–641, 1998.

[Wang *et al.*, 2014] Kewen Wang, Zhe Wang, Rodney W. Topor, Jeff Z. Pan, and Grigoris Antoniou. Eliminating concepts and roles from ontologies in expressive descriptive logics. *Comput. Intell.*, 30(2):205–232, 2014.

[Zhou *et al.*, 2015] Yujiao Zhou, Bernardo Cuenca Grau, Yavor Nenov, Mark Kaminski, and Ian Horrocks. PAGOdA: Pay-as-you-go ontology query answering using a Datalog reasoner. *J. Artif. Intell. Res.*, 54:309–367, 2015.