Querying Circumscribed Description Logic Knowledge Bases

Carsten Lutz[1], **Quentin Manière**[1,2], **Robin Nolte**[3]

[1]Department of Computer Science, Leipzig University, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany
[3]University of Bremen, Digital Media Lab, Germany
{carsten.lutz, quentin.maniere}@uni-leipzig.de, nolte@uni-bremen.de

Abstract

Circumscription is one of the main approaches for defining non-monotonic description logics (DLs) and the decidability and complexity of traditional reasoning tasks, such as satisfiability of circumscribed DL knowledge bases (KBs), are well understood. For evaluating conjunctive queries (CQs) and unions thereof (UCQs), in contrast, not even decidability has been established. In this paper, we prove decidability of (U)CQ evaluation on circumscribed DL KBs and obtain a rather complete picture of both the combined complexity and the data complexity for DLs ranging from $\mathcal{ALCHIO}$ via $\mathcal{EL}$ to various versions of DL-Lite. We also study the much simpler atomic queries (AQs).

1 Introduction

While standard description logics (DLs), such as those underlying the OWL 2 ontology language, do not support non-monotonic reasoning, it is generally acknowledged that extending DLs with non-monotonic features is very useful. Concrete examples of applications include ontological modeling in the biomedical domain (Rector 2004; Stevens et al. 2007) and the formulation of access control policies (Bonatti and Samarati 2003). Circumscription is one of the traditional AI approaches to non-monotonicity, and it provides an important way to define non-monotonic DLs. In contrast to other approaches, such as default rules, it does not require the adoption of strong syntactic restrictions to preserve decidability. DLs with circumscription are closely related to several other approaches to non-monotonic DLs, in particular to DLs with defeasible inclusions and typicality operators (Bonatti, Faella, and Sauro 2011; Casini and Straccia 2013; Giordano et al. 2013; Bonatti et al. 2015b; Pensel and Turhan 2018).

The main feature of circumscription is that selected predicate symbols can be minimized, that is, the extension of these predicates must be minimal regarding set inclusion. Other predicates may vary freely or be declared fixed. In addition, a preference order can be declared on the minimized predicates. In DLs, minimizing or fixing role names causes undecidability of reasoning, and consequently, only concept names may be minimized or fixed (Bonatti, Lutz, and Wolter 2009). The traditional AI use of circumscription is to introduce and minimize abnormality predicates such as $\mathsf{AbnormalBird}$, which makes it possible to formulate defeasible implications such as 'birds fly, unless they are abnormal, which shouldn't be assumed unless there is a reason to do so.' Circumscription is also closely related to the closure of predicates symbols as studied, for instance, in (Lutz, Seylan, and Wolter 2013; Ngo, Ortiz, and Simkus 2016; Lutz, Seylan, and Wolter 2019); in fact, this observation goes back to (Reiter 1977; Lifschitz 1985). While DLs usually assume open-world semantics and represent incomplete knowledge, such closed predicates are interpreted under a closed-world assumption, reflecting that complete knowledge is available regarding those predicates. Circumscription may then be viewed as a soft form of closing concept names: there are no other instances of a minimized concept name except the explicitly asserted ones unless we are forced to introduce (a minimal set of) additional instances to avoid inconsistency.

A primary application of DLs is ontology-mediated querying, where an ontology is used to enrich data with domain knowledge. Surprisingly, little is known about ontology-mediated querying with DLs that support circumscription. For the important conjunctive queries (CQs) and unions of CQs (UCQs), in fact, not even decidability has been established. This paper aims to close this gap by studying the decidability and precise complexity of ontology-mediated querying for DLs with circumscription, both w.r.t. combined complexity and data complexity. We consider the expressive DL $\mathcal{ALCHIO}$, the tractable (without circumscription) DL $\mathcal{EL}$, and several members of tge DL-Lite family. These may be viewed as logical cores of the profiles OWL 2 DL, OWL 2 EL, and OWL 2 QL of the OWL 2 ontology language (Motik et al. 2009).

One of our main results is that UCQ evaluation is decidable in all these DLs when circumscription is added. It is 2EXP-complete in $\mathcal{ALCHIO}$ w.r.t. combined complexity, and thus not harder than query evaluation without circumscription. W.r.t. data complexity, however, there is a significant increase from CONP- to Π^p_2-completeness. For $\mathcal{EL}$, the combined and data complexity turns out to be identical to that of $\mathcal{ALCHIO}$. This improves lower bounds from (Bonatti, Faella, and Sauro 2011). All these lower bounds already hold for CQs. Remarkably, the Π^p_2 lower bound for data complexity already holds when there is only a single minimized concept name (and thus no preference relation) and without fixed predicates. The complexities for DL-Lite are lower, though still high. A summary can be found in Table 1. Evaluation is 'only' CONP-complete w.r.t. data complexity for all considered versions of DL-Lite, except when role disjointness constraints are added (this case is not in the table). The combined complexity remains at 2EXP when role inclusions are present and drops to CONEXP without them. The lower bounds already apply to very basic versions of DL-Lite that are positive in the sense that they do not provide concept disjointness constraints, and the upper bounds apply to expressive versions that include all Boolean operators.

We also study the evaluation of the basic yet important atomic queries (AQs), conjunctive queries of the form $A(x)$ with A a concept name. Also here, we obtain a rather complete picture of the complexity landscape. It is known from (Bonatti, Lutz, and Wolter 2009) that AQ evaluation in $\mathcal{ALCHIO}$ is CONEXP$^{\mathrm{NP}}$-complete w.r.t. combined complexity. We show that the lower bound holds already for $\mathcal{EL}$. Moreover, our Π_2^{P}-lower bound for the data complexity of (U)CQ-evaluation in $\mathcal{EL}$ mentioned above only requires an AQ, and thus AQ evaluation in both $\mathcal{ALCHIO}$ and $\mathcal{EL}$ are Π_2^{P}-complete w.r.t. data complexity. For DL-Lite, the data complexity drops to PTIME in all considered versions, and the combined complexity ranges from CONEXP-complete to Π_2^{P}-complete, depending on which Boolean operators are admitted. A summary can be found in Table 2.

An appendix with full proofs can be found in the long version of this paper, see (Lutz, Manière, and Nolte 2023).

Related Work. A foundational paper on description logics with circumscription is (Bonatti, Lutz, and Wolter 2009), which studies concept satisfiability and knowledge base consistency in the $\mathcal{ALCHIO}$ family of DLs; these problems are interreducible with AQ evaluation in polynomial time (up to complementation). The same problems have been considered in (Bonatti, Faella, and Sauro 2011) for $\mathcal{EL}$ and DL-Lite and in (Bonatti et al. 2015a) for DLs without the finite model property, including a version of DL-Lite. The recent (Bonatti 2021) is the only work we are aware of that considers ontology-mediated querying in the context of circumscription. It provides lower bounds for $\mathcal{EL}$ and DL-Lite, which are both improved in the current paper, but no decidability results / upper bounds. A relaxed version of circumscription that enjoys lower complexity has recently been studied in (Stefano, Ortiz, and Simkus 2022). We have already mentioned connections to DLs with defeasible inclusions and typicality operators, see above for references. A connection between circumscription and the complexity class Π_2^{P} was first observed in (Eiter and Gottlob 1993), and this complexity shows up in our data complexity results. Our proofs, however, are quite different.

2 Preliminaries

Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be countably infinite sets of *concept names*, *role names*, and *individual names*. An *inverse role* takes the form r^- with r a role name, and a *role* is a role name or an inverse role. If $r = s^-$ is an inverse role, then r^- denotes s. An $\mathcal{ALCIO}$ *concept* C is built according to the rule $C, D ::= \top \mid A \mid \{a\} \mid \neg C \mid C \sqcap D \mid \exists r.D$ where A ranges over concept names, a over individual names, and r over roles. A concept of the form $\{a\}$ is called a *nominal*. We write $\bot$ as abbreviation for $\neg\top$, $C \sqcup D$ for $\neg(\neg C \sqcap \neg D)$, and $\forall r.C$ for $\neg\exists r.\neg C$. An $\mathcal{ALCI}$ *concept* is a nominal-free $\mathcal{ALCIO}$ concept. An $\mathcal{EL}$ *concept* is an $\mathcal{ALCI}$ concept that uses neither negation nor inverse roles.

An $\mathcal{ALCHIO}$ *TBox* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $\mathcal{ALCIO}$ concepts, and *role inclusions (RIs)* $r \sqsubseteq s$, where r, s are roles. In an $\mathcal{EL}$ *TBox*, only $\mathcal{EL}$ concepts may be used in CIs, and RIs are disallowed. An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ where A is a concept name, r a role name, and a, b are individual names. We use $\mathsf{ind}(\mathcal{A})$ to denote the set of individual names used in $\mathcal{A}$. An $\mathcal{ALCHIO}$ *knowledge base (KB)* takes the form $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ an $\mathcal{ALCHIO}$ TBox and $\mathcal{A}$ an ABox. $\mathcal{ALCHI}$ TBoxes and KBs are defined analogously but may not use nominals.

The semantics is defined as usual in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ with $\Delta^\mathcal{I}$ the (non-empty) *domain* and $\cdot^\mathcal{I}$ the *interpretation function*, we refer to (Baader et al. 2017) for full details. An interpretation satisfies a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$ and likewise for RIs. It satisfies an assertion $A(a)$ if $a \in A^\mathcal{I}$ and $r(a, b)$ if $(a, b) \in r^\mathcal{I}$; we thus make the *standard names assumption*. An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$, written $\mathcal{I} \models \mathcal{T}$, if it satisfies all inclusions in it. Models of ABoxes and KBs are defined likewise. For an interpretation $\mathcal{I}$ and $\Delta \subseteq \Delta^\mathcal{I}$, we use $\mathcal{I}|_\Delta$ to denote the restriction of $\mathcal{I}$ to subdomain Δ.

A *signature* is a set of concept and role names referred to as *symbols*. For any syntactic object O such as a TBox or an ABox, we use $\mathsf{sig}(O)$ to denote the symbols used in O and $|O|$ to denote the *size* of O, meaning the encoding of O as a word over a suitable alphabet.

We next introduce several more restricted DLs of the DL-Lite family. A *basic concept* is of the form A or $\exists r.\top$. A DL-Lite$^\mathcal{H}_{\mathsf{core}}$ TBox is a finite set of concept inclusions $C \sqsubseteq D$, *(concept) disjointness assertions* $C \sqsubseteq \neg D$, and role inclusions $r \sqsubseteq s$ where C, D are basic concepts and r, s roles. We drop superscript $\cdot^\mathcal{H}$ if no role inclusions are admitted, use subscript $\cdot_{\mathsf{horn}}$ to indicate that the concepts C, D in concept inclusions may be conjunctions of basic concepts, and subscript $\cdot_{\mathsf{bool}}$ to indicate that C, D may be Boolean combinations of basic concepts, that is, built from basic concepts using $\neg, \sqcap, \sqcup$.

A *circumscription pattern* is a tuple $\mathsf{CP} = (\prec, \mathsf{M}, \mathsf{F}, \mathsf{V})$, where $\prec$ is a strict partial order on M called the *preference relation*, and M, F and V are a partition of $\mathsf{N_C}$. The elements of M, F and V are the *minimized*, *fixed* and *varying* concept names. Role names always vary to avoid undecidability (Bonatti, Lutz, and Wolter 2009). The preference relation $\prec$ on M induces a preference relation $<_\mathsf{CP}$ on interpretations by setting $\mathcal{J} <_\mathsf{CP} \mathcal{I}$ if the following conditions hold:

1. $\Delta^\mathcal{J} = \Delta^\mathcal{I}$,
2. for all $A \in \mathsf{F}$, $A^\mathcal{J} = A^\mathcal{I}$,
3. for all $A \in \mathsf{M}$ with $A^\mathcal{J} \not\subseteq A^\mathcal{I}$, there is a $B \in \mathsf{M}$, $B \prec A$, such that $B^\mathcal{J} \subsetneq B^\mathcal{I}$,
4. there exists an $A \in \mathsf{M}$ such that $A^\mathcal{J} \subsetneq A^\mathcal{I}$ and for all $B \in \mathsf{M}$, $B \prec A$ implies $B^\mathcal{J} = B^\mathcal{I}$.

	$\mathcal{EL}$, $\mathcal{ALCHIO}$	DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$, DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$	DL-Lite$_{\mathsf{bool}}$	DL-Lite$_{\mathsf{core}}$, DL-Lite$_{\mathsf{horn}}$
Combined complexity	2EXP-c.$^{\text{(Thm. 1, 2)}}$	2EXP-c.$^{(\dagger)}$ $^{\text{(Thm. 1, 5)}}$	coNEXP-c.$^{\text{(Thm. 6, 13)}}$	coNEXP-c.$^{(\dagger)}$ $^{\text{(Thm. 6, 7)}}$
Data complexity	Π^{P}_2-c.$^{\text{(Thm. 3, 4)}}$	coNP-c.$^{\text{(Thm. 8, 9)}}$	coNP-c.$^{\text{(Thm. 8, 9)}}$	coNP-c.$^{\text{(Thm. 8, 9)}}$

Table 1: Complexity of (U)CQ evaluation on circumscribed KBs. $\cdot^{(\dagger)}$ indicates that UCQs are needed for lower bound.

A *circumscribed KB (cKB)* takes the form $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ where $\mathcal{K}$ is a KB and CP a circumscription pattern. A model $\mathcal{I}$ of $\mathcal{K}$ is a *model* of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$, denoted $\mathcal{I} \models \mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$, if no $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$ is a model of $\mathcal{K}$. A cKB $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ is *satisfiable* if it has a model.

A *conjunctive query (CQ)* takes the form $q(\bar{x}) = \exists \bar{y}\, \varphi(\bar{x}, \bar{y})$ where $\bar{x}$ and $\bar{y}$ are tuples of variables and φ is a conjunction of *atoms* $A(z)$ and $r(z, z')$, with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and z, z' variables from $\bar{x} \cup \bar{y}$. The variables in $\bar{x}$ are the *answer variables*, and $\mathsf{var}(q)$ denotes $\bar{x} \cup \bar{y}$. We take the liberty to view q as a set of atoms, writing, e.g., $\alpha \in q$ to indicate that α is an atom in q. We may also write $r^-(x, y) \in q$ in place of $r(y, x) \in q$. A CQ q gives rise to an interpretation $\mathcal{I}_q$ with $\Delta^{\mathcal{I}_q} = \mathsf{var}(q)$, $A^{\mathcal{I}_q} = \{x \mid A(x) \in q\}$, and $r^{\mathcal{I}_q} = \{(x, y) \mid r(x, y) \in q\}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. A *union of conjunctive queries (UCQ)* $q(\bar{x})$ is a disjunction of CQs that all have the same answer variables $\bar{x}$. The *arity* of q is the length of $\bar{x}$, and q is *Boolean* if it is of arity zero. An *atomic query (AQ)* is a CQ of the simple form $A(x)$, with A a concept name.

With a homomorphism from a CQ q to an interpretation $\mathcal{I}$, we mean a homomorphism from $\mathcal{I}_q$ to $\mathcal{I}$ (defined as usual). A tuple $\bar{d} \in (\Delta^{\mathcal{I}})^{|\bar{x}|}$ is an *answer* to a UCQ $q(\bar{x})$ on an interpretation $\mathcal{I}$, written $\mathcal{I} \models q(\bar{d})$, if there is a homomorphism h from a CQ p in q to $\mathcal{I}$ with $h(\bar{x}) = \bar{d}$. A tuple $\bar{a} \in \mathsf{ind}(\mathcal{A})$ is an *answer* to q on a cKB $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, written $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$, if $\mathcal{I} \models q(\bar{a})$ for all models $\mathcal{I}$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$.

Example 1. *Consider a database about universities. The TBox contains domain knowledge such as*

$$\begin{aligned} \mathsf{University} &\sqsubseteq \mathsf{Organization} \\ \mathsf{Organization} &\sqsubseteq \mathsf{Public} \sqcup \mathsf{Private} \\ \mathsf{Public} \sqcap \mathsf{Private} &\sqsubseteq \bot. \end{aligned}$$

Circumscription can be used to express defeasible inclusions. For example, from a European perspective, universities are usually public:

$$\mathsf{University} \sqsubseteq \mathsf{Public} \sqcup \mathsf{Ab}_U$$

where Ab_U *is a fresh 'abnormality' concept name that is minimized. If the ABox contains*

$$\mathsf{University}(\mathsf{leipzigu}), \mathsf{University}(\mathsf{mit}), \mathsf{Private}(\mathsf{mit})$$

and we pose the CQ $q(x) = \mathsf{Organization}(x) \wedge \mathsf{Public}(x)$*, then the answer is* leipzigu.

We may also use circumscription to implement a soft closed-world assumption, similar in spirit to soft constraints in constraint satisfaction. Assume that the ABox additionally contains a database of nonprofit corporations:

$$\mathsf{NPC}(\mathsf{greenpeace}) \quad \mathsf{NPC}(\mathsf{wwf})$$

and that this database is essentially complete*, expressed by minimizing* NPC*. If we also know that*

$$\begin{aligned} \mathsf{IvyLeagueU} &\sqsubseteq \exists \mathsf{ownedBy}.(\mathsf{NPC} \sqcap \mathsf{Rich}) \\ \mathsf{DonationBased} &\sqsubseteq \neg \mathsf{Rich} \end{aligned}$$
$$\mathsf{DonationBased}(\mathsf{greenpeace}) \quad \mathsf{DonationBased}(\mathsf{wwf})$$
$$\mathsf{IvyLeagueU}(\mathsf{harvard})$$

then we are forced to infer that our list of NPCs was not actually complete as all explicitly known NPCs are not rich, but a rich NPC must exist. A strict closed-world assumption would instead result in an inconsistency.

Let $\mathcal{L}$ be a description logic such as $\mathcal{ALCHIO}$ or $\mathcal{EL}$ and let $\mathcal{Q}$ be a query language such as UCQ, CQ, or AQ. With $\mathcal{Q}$ *evaluation on circumscribed* $\mathcal{L}$ *KBs*, we mean the problem to decide, given an $\mathcal{L}$ cKB $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a query $q(\bar{x})$ from $\mathcal{Q}$, and a tuple $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$, whether $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$. When studying the combined complexity of this problem, all of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$, q, and $\bar{a}$ are treated as inputs. For data complexity, in contrast, we assume q, $\mathcal{T}$, and CP to be fixed and thus of constant size.

3 Between $\mathcal{ALCHIO}$ and $\mathcal{EL}$

We study the complexity of query evaluation on circumscribed KBs for DLs between $\mathcal{ALCHIO}$ and $\mathcal{EL}$. In fact, we prove 2EXP-completeness in combined complexity and Π^{P}_2-completeness in data complexity for all these DLs.

3.1 Fundamental Observations

We start with some fundamental observations that underlie the subsequent proofs, first observing a reduction from UCQ evaluation on circumscribed $\mathcal{ALCHIO}$ KBs to UCQ evaluation on circumscribed $\mathcal{ALCHI}$ KBs. Note that a nominal may be viewed as a (strictly) closed concept name with a single instance. This reduction is a simple version of a reduction from query evaluation with closed concept names to query evaluation on cKBs in the proof of Theorem 2 below.

Proposition 1. *UCQ evaluation on circumscribed* $\mathcal{ALCHIO}$ *KBs can be reduced in polynomial time to UCQ evaluation on circumscribed* $\mathcal{ALCHI}$ *KBs.*

Proof. Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be an $\mathcal{ALCHIO}$ cKB, with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, and let $q(\bar{x})$ be a UCQ. Let N be the set of individual names a such that the nominal $\{a\}$ is used in $\mathcal{T}$. Introduce fresh concept names A_a, B_a, D_a for every $a \in N$. We obtain $\mathcal{T}'$ from $\mathcal{T}$ by replacing every $a \in N$ with A_a and adding the CI $A_a \sqcap \neg B_a \sqsubseteq D_a$, $\mathcal{A}'$ from $\mathcal{A}$ by adding $A_a(a)$ and $B_a(a)$ for every $a \in N$, and q' from q by adding the disjunct $\exists y\, D_a(y)$ for every $a \in N$.[1] Set $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$. The circumscription pattern CP' is obtained from CP by minimizing the concept names B_a with higher preference than any other concept name (and with no preferences between them). We show in the appendix that $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$ iff $\mathsf{Circ}_{\mathsf{CP}'}(\mathcal{K}') \models q'(\bar{a})$ for all $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$. $\Box$

We are thus left with $\mathcal{ALCHI}$ KBs. We generally assume that TBoxes are in *normal form*, meaning that every concept inclusion in $\mathcal{T}$ has one of the following shapes:

$$\top \sqsubseteq A \qquad A \sqsubseteq \exists r.B \qquad \exists r.B \sqsubseteq A$$
$$A_1 \sqcap A_2 \sqsubseteq A \qquad A \sqsubseteq \neg B \qquad \neg B \sqsubseteq A$$

where A, A_1, A_2, B range over $\mathsf{N_C}$ and r ranges over roles. The set of concept names in $\mathcal{T}$ is denoted $\mathsf{N_C}(\mathcal{T})$.

Lemma 1. *Every $\mathcal{ALCHI}$ TBox $\mathcal{T}$ can be transformed in linear time into an $\mathcal{ALCHI}$ TBox $\mathcal{T}'$ in normal form such that for all cKBs $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{T}, \mathcal{A})$, UCQs $q(\bar{x})$ that do not use symbols from $\mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$, and $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$: $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{T}, \mathcal{A}) \models q(\bar{a})$ iff $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{T}', \mathcal{A}) \models q(\bar{a})$.*

Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be a cKB with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. A *type* is a set of concept names $t \subseteq \mathsf{N_C}(\mathcal{T})$. For an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, we define $\mathsf{tp}_{\mathcal{I}}(d) := \{A \in \mathsf{N_C}(\mathcal{T}) \mid d \in A^{\mathcal{I}}\}$. For a subset $\Delta \subseteq \Delta^{\mathcal{I}}$, we set $\mathsf{tp}_{\mathcal{I}}(\Delta) = \{\mathsf{tp}_{\mathcal{I}}(d) \mid d \in \Delta\}$. We further write $\mathsf{TP}(\mathcal{I})$ for $\mathsf{tp}_{\mathcal{I}}(\Delta^{\mathcal{I}})$. Finally, we set

$$\mathsf{TP}(\mathcal{T}) := \bigcup_{\mathcal{I} \text{ model of } \mathcal{T}} \mathsf{TP}(\mathcal{I}).$$

For a role r, we write $t \rightsquigarrow_r t'$ if for all $A, B \in \mathsf{N_C}$:

- $B \in t'$ and $\mathcal{T} \models \exists r.B \sqsubseteq A$ implies $A \in t$ and
- $B \in t$ and $\mathcal{T} \models \exists r^-.B \sqsubseteq A$ implies $A \in t'$.

We next show how to identify a 'core' part of a model $\mathcal{I}$ of $\mathcal{K}$. These cores will play an important role in dealing with circumscription in our upper bound proofs.

Definition 1. *Let $\mathcal{I}$ be a model of $\mathcal{K}$. We use $\mathsf{TP}_{\mathsf{core}}(\mathcal{I})$ to denote the set of all types $t \in \mathsf{TP}(\mathcal{I})$ such that*

$$|\{d \in \Delta^{\mathcal{I}} \setminus \mathsf{ind}(\mathcal{A}) \mid \mathsf{tp}_{\mathcal{I}}(d) = t\}| < |\mathsf{TP}(\mathcal{T})|$$

and set $\mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I}) = \mathsf{TP}(\mathcal{I}) \setminus \mathsf{TP}_{\mathsf{core}}(\mathcal{I})$ and

$$\Delta^{\mathcal{I}}_{\mathsf{core}} = \{d \in \Delta^{\mathcal{I}} \mid \mathsf{tp}_{\mathcal{I}}(d) \in \mathsf{TP}_{\mathsf{core}}(\mathcal{I})\}.$$

So the core consists of all elements whose types are realized not too often, except possibly in the ABox. A good way of thinking about cores is that if a model $\mathcal{I}$ of $\mathcal{K}$ is minimal w.r.t. $<_{\mathsf{CP}}$, then all instances of minimized concept names are in the core. This is, however not strictly true since we may have $A \sqsubseteq B$ where A is $\top$ or fixed, and B is minimized.

The following crucial lemma provides a sufficient condition for a model $\mathcal{J}$ of $\mathcal{K}$ to be minimal w.r.t. $<_{\mathsf{CP}}$, relative to a model $\mathcal{I}$ of $\mathcal{K}$ that is known to be minimal w.r.t. $<_{\mathsf{CP}}$.

[1]Strictly speaking, we need to adjust $\exists y\, D_a(y)$ so that it has the same answer variables as the other CQs in q. This is easy by adding to $\mathcal{T}'$ a CI $\top \sqsubseteq T$ for a fresh concept name T and extending $\exists y\, D_a(y)$ with atom $T(x)$ for every answer variable x.

Lemma 2 (Core Lemma)**.** *Let $\mathcal{I}$ be a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ and let $\mathcal{J}$ be a model of $\mathcal{K}$ with $\Delta^{\mathcal{I}}_{\mathsf{core}} \subseteq \Delta^{\mathcal{J}}$. If*

1. *$\mathsf{tp}_{\mathcal{I}}(d) = \mathsf{tp}_{\mathcal{J}}(d)$ for all $d \in \Delta^{\mathcal{I}}_{\mathsf{core}}$ and*
2. *$\mathsf{tp}_{\mathcal{J}}(\Delta^{\mathcal{J}} \setminus \Delta^{\mathcal{I}}_{\mathsf{core}}) = \mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I})$,*

then $\mathcal{J}$ is a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$.

We give a sketch of the proof of (the contrapositive of) Lemma 2. Assume that $\mathcal{J}$ is not a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$. Then there must be a model $\mathcal{J}'$ of $\mathcal{K}$ with $\mathcal{J}' <_{\mathsf{CP}} \mathcal{J}$ and to obtain a contradiction it suffices to construct from $\mathcal{J}'$ a model $\mathcal{I}'$ of $\mathcal{K}$ with $\mathcal{I}' <_{\mathsf{CP}} \mathcal{I}$. Note that $\mathcal{I}$ and $\mathcal{J}$ may have domains of different sizes. The elements of $\Delta^{\mathcal{I}}_{\mathsf{core}}$ receive the same type in $\mathcal{I}'$ as in $\mathcal{J}'$. For each non-core type t in $\mathcal{I}$, we consider the set S_t of types in $\mathcal{J}'$ of those elements that have type t in $\mathcal{J}$. Since t is realized in $\mathcal{I}$ at least $|\mathsf{TP}(\mathcal{T})|$ many times, we have enough room to realize in $\mathcal{I}'$ exactly the types from S_t among those elements that had type t in $\mathcal{I}$, that is, within $(\mathsf{tp}_{\mathcal{I}})^{-1}(t)$. It is easy to see that $\mathcal{I}'$ is a model of $\mathcal{K}$: it satisfies $\mathcal{T}$ as it realizes the same types as $\mathcal{J}'$ and it satisfies $\mathcal{A}$ since $\mathcal{I}'|_{\mathsf{ind}(\mathcal{A})} = \mathcal{J}'|_{\mathsf{ind}(\mathcal{A})}$. By construction and since $\mathcal{J}' <_{\mathsf{CP}} \mathcal{J}$, it further satisfies $\mathcal{I}' <_{\mathsf{CP}} \mathcal{I}$.

We next use the core lemma to show that if $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \not\models q(\bar{a})$ for any CQ q and $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$, then this is witnessed by a countermodel $\mathcal{I}$ that has a regular shape. Here and in what follows, a *countermodel* is a model $\mathcal{I}$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ with $\mathcal{I} \not\models q(\bar{a})$. By regular shape, we mean that there is a 'base part' that contains the ABox, the core of $\mathcal{I}$, as well as some additional elements; all other parts of $\mathcal{I}$ are tree-shaped with their root in the base part, and potentially with edges that go back to the core (but not to other parts of the base!). We next make this precise.

Let $\mathcal{I}$ be a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$. Set $\Omega = \{rA \mid B \sqsubseteq \exists r.A \in \mathcal{T}\}$ and fix a function f that chooses, for every $d \in \Delta^{\mathcal{I}}$ and $rA \in \Omega$ with $d \in (\exists r.A)^{\mathcal{I}}$, an element $f(d, rA) = e \in A^{\mathcal{I}}$ such that $(d, e) \in r^{\mathcal{I}}$. Further choose, for every $t \in \mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I})$, a representative $e_t \in \Delta^{\mathcal{I}}$ with $\mathsf{tp}_{\mathcal{I}}(e_t) = t$. We inductively define the set $\mathcal{P}$ of *paths through* $\mathcal{I}$ along with a mapping h assigning to each $p \in \mathcal{P}$ an element of $\Delta^{\mathcal{I}}$:

- each element d of the set
$$\Delta^{\mathcal{I}}_{\mathsf{base}} := \mathsf{ind}(\mathcal{A}) \cup \Delta^{\mathcal{I}}_{\mathsf{core}} \cup \{e_t \mid t \in \mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I})\}$$
is a path in $\mathcal{P}$ and $h(d) = d$;
- if $p \in \mathcal{P}$ with $h(p) = d$ and $rA \in \Omega$ with $f(d, rA)$ defined and not from $\Delta^{\mathcal{I}}_{\mathsf{core}}$, then $p' = prA$ is a path in $\mathcal{P}$ and $h(p') = f(d, rA)$.

For every role r, define

$$\begin{aligned} R_r = \; & \{(a, b) \mid a, b \in \mathsf{ind}(\mathcal{A}), r(a, b) \in \mathcal{A}\} \cup \\ & \{(d, e) \mid d, e \in \Delta^{\mathcal{I}}_{\mathsf{core}}, (d, e) \in r^{\mathcal{I}}\} \cup \\ & \{(p, p') \mid p' = prA \in \mathcal{P}\} \cup \\ & \{(p, e) \mid e = f(h(p), rA) \in \Delta^{\mathcal{I}}_{\mathsf{core}}\}. \end{aligned}$$

Now the *unraveling* of $\mathcal{I}$ is defined by setting

$$\Delta^{\mathcal{I}'} = \mathcal{P} \qquad A^{\mathcal{I}'} = \{p \in \mathcal{P} \mid h(p) \in A^{\mathcal{I}}\}$$
$$r^{\mathcal{I}'} = \bigcup_{\mathcal{T} \models s \sqsubseteq r} (R_s \cup \{(e, d) \mid (d, e) \in R_{s^-}\})$$

for all concept names A and role names r. It is easy to verify that h is a homomorphism from $\mathcal{I}'$ to $\mathcal{I}$. This version of unraveling is inspired by constructions from (Manière 2022) where the resulting model is called an *interlacing*.

Lemma 3. *Let $q(\bar{x})$ be a UCQ and $\bar{a} \in \mathsf{ind}(\mathcal{A})$. If $\mathcal{I}$ is a countermodel against $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$, then so is its unraveling $\mathcal{I}'$.*

For some of our upper bound proofs, it will be important that the reference model $\mathcal{I}$ in the core lemma is finite and sufficiently small. The following lemma shows that we can always achieve this.

Lemma 4. *Let $\mathcal{I}$ be a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$. There exists a model $\mathcal{J}$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ such that $|\Delta^{\mathcal{J}}| \leq |\mathcal{A}| + 2^{2|\mathcal{T}|}$, $\mathcal{I}|_{\Delta^{\mathcal{I}}_{\mathsf{core}}} = \mathcal{J}|_{\Delta^{\mathcal{J}}_{\mathsf{core}}}$, and $\mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I}) = \mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{J})$.*

In the proof of Lemma 4, we construct the desired model $\mathcal{J}$ by starting from $\mathcal{I}|_{\mathsf{ind}(\mathcal{A}) \cup \Delta^{\mathcal{I}}_{\mathsf{core}}}$ and adding exactly $m = |\mathsf{TP}(\mathcal{T})|$ instances of each type from $\mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{I})$.

3.2 Combined Complexity

We show that in any DL between $\mathcal{EL}$ and $\mathcal{ALCHIO}$, the evaluation of CQs and UCQs on cKBs is 2EXP-complete w.r.t. combined complexity, starting with the upper bound.

Theorem 1. *UCQ evaluation on circumscribed $\mathcal{ALCHIO}$ KBs is in* 2EXP *w.r.t. combined complexity.*

By Proposition 1, it suffices to consider $\mathcal{ALCHI}$. Assume that we are given as an input an $\mathcal{ALCHI}$ cKB $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a UCQ $q(\bar{x})$, and a tuple $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$. We have to decide whether or not there is a countermodel $\mathcal{I}$ against $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$.

Fix a set Δ of size $|\mathsf{ind}(\mathcal{A})| + 2^{2|\mathcal{T}|+1}$ such that $\mathsf{ind}(\mathcal{A}) \subseteq \Delta$. Note that Δ is sufficiently large so that we may assume the base domain $\Delta^{\mathcal{I}}_{\mathsf{base}}$ of unraveled interpretations to be a subset of Δ. In an outer loop, our algorithm iterates over all triples $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$ such that the following conditions are satisfied:

- $\mathcal{I}_{\mathsf{base}}$ is a model of $\mathcal{A}$ with $\Delta_{\mathsf{core}} \subseteq \Delta^{\mathcal{I}_{\mathsf{base}}} \subseteq \Delta$;
- $\mathsf{tp}_{\mathcal{I}_{\mathsf{base}}}(\Delta^{\mathcal{I}_{\mathsf{base}}} \setminus \Delta_{\mathsf{core}}) = T_{\overline{\mathsf{core}}}$;
- $\mathsf{tp}_{\mathcal{I}_{\mathsf{base}}}(\Delta_{\mathsf{core}}) \cap T_{\overline{\mathsf{core}}} = \emptyset$.

For each triple $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$, we then check whether the following additional conditions are satisfied:

(I) $\mathcal{I}_{\mathsf{base}}$ can be extended to a model $\mathcal{I}$ of $\mathcal{T}$ such that
 (a) $\mathcal{I}|_{\Delta^{\mathcal{I}_{\mathsf{base}}}} = \mathcal{I}_{\mathsf{base}}$,
 (b) $\mathsf{tp}_{\mathcal{I}}(\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{I}_{\mathsf{base}}}) \subseteq T_{\overline{\mathsf{core}}}$, and
 (c) $\mathcal{I} \not\models q(\bar{a})$;

(II) there exists a model $\mathcal{J}$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ such that $\mathcal{J}|_{\Delta^{\mathcal{J}}_{\mathsf{core}}} = \mathcal{I}_{\mathsf{base}}|_{\Delta_{\mathsf{core}}}$ and $\mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{J}) = T_{\overline{\mathsf{core}}}$.

We return 'yes' if all triples fail the check and 'no' otherwise.

If the checks succeed, then the model $\mathcal{I}$ of $\mathcal{K}$ from Condition (I) is a countermodel against $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$. In particular, we may apply Lemma 2, using the model $\mathcal{J}$ from Condition (II) as the reference model, to show that $\mathcal{I}$ is minimal w.r.t. $<_{\mathsf{CP}}$. Conversely, any countermodel $\mathcal{I}_0$ against $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$ can be unraveled into a countermodel $\mathcal{I}$ from which we can read off a triple $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$ in the obvious way, and then $\mathcal{I}$ witnesses Condition (I) and choosing $\mathcal{J} = \mathcal{I}$ witnesses Condition (II).

Of course, we have to prove that Conditions (I) and (II) can be verified in 2EXP. This is easy for Condition (II): by Lemma 4, it suffices to consider models $\mathcal{J}$ of size at most $|\mathcal{A}| + 2^{2|\mathcal{T}|}$ and thus we can iterate over all candidate interpretations $\mathcal{J}$ up to this size, check whether $\mathcal{J}$ is a model of $\mathcal{K}$ with $\mathcal{J}|_{\Delta^{\mathcal{J}}_{\mathsf{core}}} = \mathcal{I}_{\mathsf{base}}|_{\Delta_{\mathsf{core}}}$ and $\mathsf{TP}_{\overline{\mathsf{core}}}(\mathcal{J}) = T_{\overline{\mathsf{core}}}$, and then iterate over all models $\mathcal{J}'$ of $\mathcal{K}$ with $\Delta^{\mathcal{J}'} = \Delta^{\mathcal{J}}$ to check that $\mathcal{J}$ is minimal w.r.t. $<_{\mathsf{CP}}$.

Condition (I) requires more work. We use a mosaic approach, that is, we attempt to assemble the interpretation $\mathcal{I}$ from Condition (I) by combining small pieces called *mosaics*. Each mosaic contains the base part of $\mathcal{I}$ and at most two additional elements e_1^*, e_2^*. We trace partial homomorphisms from CQs in q through the mosaics, as follows.

Fix a triple $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$. A *match triple* for an interpretation $\mathcal{J}$ takes the form $(p, \widehat{p}, h)$ such that p is a CQ in q, $\widehat{p} \subseteq p$, and h is a partial map from $\mathsf{var}(\widehat{p})$ to $\Delta^{\mathcal{J}}$ that is a homomorphism from $\widehat{p}|_{\mathsf{dom}(h)}$ to $\mathcal{J}$. Intuitively, $\mathcal{J}$ is a mosaic and the triple $(p, \widehat{p}, h)$ expresses that a homomorphism from $\widehat{p}$ to $\mathcal{I}$ exists, with the variables in $\mathsf{dom}(h)$ being mapped to the current piece $\mathcal{J}$ and the variables in $\mathsf{var}(\widehat{p}) \setminus \mathsf{dom}(h)$ mapped to other mosaics. A match triple is *complete* if $\widehat{p} = p$ and *incomplete* otherwise. To make $\mathcal{I}$ a countermodel, we must avoid complete match triples. A *specification* for $\mathcal{J}$ is a set S of match triples for $\mathcal{J}$ and we call S *saturated* if the following conditions are satisfied:

- if p is a CQ in q, $\widehat{p} \subseteq p$, and h is a homomorphism from $\widehat{p}$ to $\mathcal{J}$, then $(p, \widehat{p}, h) \in S$;
- if $(p, \widehat{p}, h), (p, \widehat{p}', h') \in S$ and $h(x) = h'(x)$ is defined for all $x \in \mathsf{var}(\widehat{p}) \cap \mathsf{var}(\widehat{p}')$, then $(p, \widehat{p} \cup \widehat{p}', h \cup h') \in S$.

Definition 2. *A* mosaic *for $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$ is a tuple $M = (\mathcal{J}, S)$ where*

- *$\mathcal{J}$ is an interpretation such that*
  1. *$\Delta^{\mathcal{I}_{\mathsf{base}}} \subseteq \Delta^{\mathcal{J}} \subseteq \Delta^{\mathcal{I}_{\mathsf{base}}} \uplus \{e_1^*, e_2^*\}$;*
  2. *$\mathcal{J}|_{\Delta^{\mathcal{I}_{\mathsf{base}}}} = \mathcal{I}_{\mathsf{base}}$;*
  3. *$\mathsf{tp}_{\mathcal{J}}(e_i^*) \in T_{\overline{\mathsf{core}}}$ if $e_i^* \in \Delta^{\mathcal{J}}$, for $i \in \{1, 2\}$;*
  4. *$\mathcal{J}$ satisfies all $\exists r.A \sqsubseteq B \in \mathcal{T}$ and all $r \sqsubseteq s \in \mathcal{T}$.*
- *S is a saturated specification for $\mathcal{J}$ that contains only incomplete match triples.*

We use $\mathcal{J}_M$ to refer to $\mathcal{J}$ and S_M to refer to S.

Let $\mathcal{M}$ be a set of mosaics for $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$. We say that $M \in \mathcal{M}$ is *good in* $\mathcal{M}$ if for every $e \in \Delta^{\mathcal{J}_M}$ and every $A \sqsubseteq \exists r.B \in \mathcal{T}$ with $e \in (A \sqcap \neg\exists r.B)^{\mathcal{J}_M}$, we find a mosaic $M' \in \mathcal{M}$ such that the following conditions are satisfied:

1. $\mathsf{tp}_{\mathcal{J}_M}(e) = \mathsf{tp}_{\mathcal{J}_{M'}}(e)$;
2. $e \in (\exists r.B)^{\mathcal{J}_{M'}}$;
3. if $(p, \widehat{p}, h') \in S_{M'}$, then $(p, \widehat{p}, h) \in S_M$ where h is the restriction of h' to range $\Delta^{\mathcal{I}_{\mathsf{base}}} \cup \{e\}$.

If M is not good in $\mathcal{M}$, then it is *bad*. To verify Cond. (I), we start with the set of all mosaics for $(\mathcal{I}_{\mathsf{base}}, \Delta_{\mathsf{core}}, T_{\overline{\mathsf{core}}})$ and repeatedly and exhaustively eliminate bad mosaics.

Lemma 5. *$\mathcal{I}_{\mathsf{base}}$ can be extended to a model $\mathcal{I}$ of $\mathcal{T}$ that satisfies Conditions (a) to (c) iff at least one mosaic survives the elimination process.*

We provide a matching lower bound for Theorem 1. It is rather strong as it already applies to CQs, to $\mathcal{EL}$ KBs, and uses a single minimized concept name (and consequently no preferences) and no fixed concept names. It is proved by a subtle reduction from CQ evaluation on $\mathcal{EL}$ KBs with closed concept names, that is, with KBs $(\mathcal{T}, \mathcal{A})$ enriched with a set Σ of closed concept names A that have to be interpreted as $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$ in all models $\mathcal{I}$. This problem was proved to be 2EXP-hard in (Ngo, Ortiz, and Simkus 2016). The reduction also sheds some light on the connection between circumscription and closing concept names.

Theorem 2. *CQ evaluation on circumscribed $\mathcal{EL}$ KBs is* 2EXP*-hard w.r.t. combined complexity. This holds even with a single minimized concept name and no fixed concept names.*

3.3 Data Complexity

We show that in any DL between $\mathcal{EL}$ and $\mathcal{ALCHIO}$, the evaluation of CQs and UCQs on cKBs is Π_2^{P}-complete w.r.t. data complexity. We start with the upper bound.

Theorem 3. *UCQ evaluation on circumscribed $\mathcal{ALCHIO}$ KBs is in Π_2^{P} w.r.t. data complexity.*

We may again limit our attention to $\mathcal{ALCHI}$. To prove Theorem 3, we show that if a countermodel exists, then there is one of polynomial size (with TBox and query of constant size). Note that this is not a consequence of Lemma 4 since there is no guarantee that, if the model $\mathcal{I}$ from that lemma is a countermodel, then so is $\mathcal{J}$. Once the size bound is in place, we obtain the Π_2^{P} upper bound by a straightforward guess-and-check procedure.

Lemma 6. *Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be an $\mathcal{ALCHI}$ cKB, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $q(\bar{x})$ a UCQ, and $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$. If $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \not\models q(\bar{a})$, then there exists a countermodel $\mathcal{I}$ with $|\Delta^{\mathcal{I}}| \leq (|\mathcal{A}|+2^{|\mathcal{T}|})^{|\mathcal{T}|^{|q|}}$.*

To prove Lemma 6, let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be an $\mathcal{ALCHI}$ cKB, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, and let $\mathcal{I}$ be a model of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ with $\mathcal{I} \not\models q(\bar{a})$. We construct a small countermodel by starting from the unraveling $\mathcal{I}'$ of $\mathcal{I}$ and applying a quotient construction. The latter is based on a suitable equivalence relation on $\Delta^{\mathcal{I}'}$ which we define next. Recall that $\mathcal{I}'$ consists of a base part $\Delta^{\mathcal{I}}_{\mathsf{base}}$ and a tree-part with backedges to $\Delta^{\mathcal{I}}_{\mathsf{base}}$.

For $n \geq 0$ and $d \in \Delta^{\mathcal{I}'} \setminus \Delta^{\mathcal{I}}_{\mathsf{base}}$, we use $\mathcal{N}_n(d)$ to denote the *n-neighborhood* of d in $\mathcal{I}'$ up to $\Delta^{\mathcal{I}}_{\mathsf{base}}$, that is, the set of all elements $e \in \Delta^{\mathcal{I}'} \setminus \Delta^{\mathcal{I}}_{\mathsf{base}}$ such that the undirected graph

$$G_{\mathcal{I}'} = (\Delta^{\mathcal{I}'}, \{\{d, e\} \mid (d, e) \in r^{\mathcal{I}'} \text{ for some } r \in \mathsf{N_R}\})$$

contains a path $d_0, \ldots, d_k$ with $d_0 = d$, $d_0, \ldots, d_{k-1} \notin \Delta^{\mathcal{I}}_{\mathsf{base}}$, and $d_k = e$, $0 \leq k \leq n$.

Recall that the elements of $\Delta^{\mathcal{I}'}$ are paths through $\mathcal{I}$, sequences $p = d_0 r_1 A_1 \cdots r_n A_n$ with $d_0 \in \Delta^{\mathcal{I}}_{\mathsf{base}}$ and $r_i A_i \in \Omega$; the *length* of p, denoted by $|p|$, is n. By definition of $\mathcal{I}'$ and of neighborhoods, for every $n \geq 0$ and $d \in \Delta^{\mathcal{I}'} \setminus \Delta^{\mathcal{I}}_{\mathsf{base}}$, there is a unique path $p_{d,n} \in \mathcal{N}_n(d)$ that is a prefix of all paths in $\mathcal{N}_n(d)$, that is, all $e \in \mathcal{N}_n(d) \setminus \Delta^{\mathcal{I}}_{\mathsf{base}}$ take the form $p_{d,n} r_1 A_1 \cdots r_k A_k$.

For $n \geq 0$, the equivalence relation $\sim_n$ on $\Delta^{\mathcal{I}'}$ is defined by setting $d_1 \sim_n d_2$ if $d_1 = d_2 \in \Delta^{\mathcal{I}}_{\mathsf{base}}$ or $d_1, d_2 \notin \Delta^{\mathcal{I}}_{\mathsf{base}}$ and the following conditions are satisfied:

1. $d_1 = p_{d_1,n} w$ and $d_2 = p_{d_2,n} w$ for some $w \in \Omega^*$;
2. for every $w = r_1 A_1 \cdots r_k A_k \in \Omega^*$:
$$p_{d_1,n} w \in \mathcal{N}_n(d_1) \text{ iff } p_{d_2,n} w \in \mathcal{N}_n(d_2)$$
   and if $p_{d_1,n} w \in \mathcal{N}_n(d_1)$, then
   (a) $\mathsf{tp}_{\mathcal{I}'}(p_{d_1,n} w) = \mathsf{tp}_{\mathcal{I}'}(p_{d_2,n} w)$;
   (b) $(p_{d_1,n} w, e) \in r^{\mathcal{I}'}$ iff $(p_{d_2,n} w, e) \in r^{\mathcal{I}'}$ for all roles r and $e \in \Delta^{\mathcal{I}}_{\mathsf{base}}$.
3. $|d_1| \equiv |d_2| \mod 2|q| + 3$.

For an element $d \in \Delta^{\mathcal{I}'}$, we use $\overline{d}$ to denote the equivalence class of d w.r.t. $\sim_{|q|+1}$. The *quotient* $\mathcal{I}'/{\sim_{|q|+1}}$ of $\mathcal{I}'$ is the interpretation whose domain is the set of all equivalence classes of $\sim_{|q|+1}$ and where

$$\begin{aligned} A^{\mathcal{I}'/\sim_{|q|+1}} &= \{\overline{d} \mid d \in A^{\mathcal{I}'}\} \\ r^{\mathcal{I}'/\sim_{|q|+1}} &= \{(\overline{d}, \overline{e}) \mid (d, e) \in r^{\mathcal{I}'}\} \end{aligned}$$

for all concept names A and role names r. It can be verified that $|\Delta^{\mathcal{I}'/\sim_{|q|+1}}| \leq (|\mathcal{A}| + 2^{|\mathcal{T}|})^{|\mathcal{T}|^{|q|}}$, as desired.

Lemma 7. *Let $q(\bar{x})$ be a UCQ and $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$. If $\mathcal{I}$ is a countermodel against $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models q(\bar{a})$, then so is $\mathcal{I}'/{\sim_{|q|+1}}$.*

It is in fact straightforward to show that $\mathcal{I}'/{\sim_{|q|+1}}$ is a model of $\mathcal{K}$. Minimality w.r.t. $<_{\mathsf{CP}}$ is proved using Lemma 2. The most subtle part of the proof of Lemma 7 is showing that $\mathcal{I}'/{\sim_{|q|+1}} \not\models q(\bar{a})$. This is done by exhibiting suitable 'local' homomorphisms from $\mathcal{I}'/{\sim_{|q|+1}}$ back to $\mathcal{I}'$ so that from any homomorphism from a CQ p in q to $\mathcal{I}'/{\sim_{|q|+1}}$ with $h(\bar{x}) = \bar{a}$, we obtain a homomorphism from p to $\mathcal{I}$ with the same property by composition. This finishes the proof of Lemmas 7 and 6 and of Theorem 3.

We next provide a matching lower bound for Theorem 3. It already applies to AQs and $\mathcal{EL}$ KBs, and when there is a single minimized concept name and no fixed concept name. It is proved by a subtle reduction from the validity of $\forall\exists$3SAT sentences. Several non-obvious technical tricks are needed to make the reduction work with a single minimized concept name. Our result improves upon a known coNP lower bound from (Bonatti 2021).

Theorem 4. *AQ evaluation on circumscribed $\mathcal{EL}$ KBs is Π_2^{P}-hard. This holds even with a single minimized concept name and no fixed concept names.*

4 DL-Lite

We consider the DL-Lite family of DLs. Without circumscription, these DLs enjoy low complexity of query evaluation, typically NP-complete in combined complexity and in AC^0 in data complexity (depending on the dialect). With circumscription, the complexity tends to still be very high, though in some relevant cases it is lower than in $\mathcal{ALCHIO}$.

4.1 Combined Complexity

Our first result shows that, when role inclusions are admitted, nothing is gained from transitioning from $\mathcal{ALCHIO}$ to DL-Lite. The proof is a variation of that of Theorem 2, but technically much simpler.

Theorem 5. *UCQ evaluation on circumscribed DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ KBs is* 2EXP*-hard w.r.t. combined complexity. This holds even with a single minimized concept name, no fixed concept names, and no disjointness constraints.*

We now move to DL-Lite$_{\mathsf{bool}}$ as a very expressive DL-Lite dialect without role inclusions and observe that the combined complexity decreases.

Theorem 6. *UCQ evaluation on circumscribed DL-Lite$_{\mathsf{bool}}$ KBs is in* CONEXP *w.r.t. combined complexity.*

To prove Theorem 6, we first observe that we can refine the unraveling of countermodels $\mathcal{I}$ from Section 3.1 such that each element outside of the base part $\Delta^{\mathcal{I}}_{\mathsf{base}}$ has at most one successor per role. This property allows us to simplify the notion of a neighborhood in the quotient construction in Section 3.3. This, in turn, yields the following result. Note that in contrast to Lemma 6, there is no double exponential dependence on $|q|$.

Lemma 8. *Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be a DL-Lite$_{\mathsf{bool}}$ cKB with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $q(\bar{x})$ a UCQ, and $\bar{a} \in \mathsf{ind}(\mathcal{A})^{|\bar{x}|}$. If $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \not\models q(\bar{a})$, then there exists a countermodel $\mathcal{I}$ with $|\Delta^{\mathcal{I}}| \leq (|\mathcal{A}| + 2^{|\mathcal{T}|})^{|q|^2(|\mathcal{T}|+1)}$.*

Lemma 8 yields a CONEXP$^{\mathrm{NP}}$ upper bound by a straightforward guess-and-check procedure: guess a model $\mathcal{I}$ of $\mathcal{K}$ with $\mathcal{I} \not\models q(\bar{a})$ and $|\Delta^{\mathcal{I}}|$ bounded as in Lemma 8, and use an oracle to check that $\mathcal{I}$ is minimal w.r.t. $<_{\mathsf{CP}}$ where the oracle decides, given a cKB $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ and a model $\mathcal{I}$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$, whether $\mathcal{I}$ is non-minimal w.r.t. $<_{\mathsf{CP}}$ by guessing a model $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$ of $\mathcal{K}$; such an approach was used also in (Bonatti, Lutz, and Wolter 2009).

To obtain a CONEXP upper bound as desired, we replace the oracle with a more efficient method to check whether a given model $\mathcal{I}$ of $\mathcal{K}$ is minimal w.r.t. $<_{\mathsf{CP}}$. The crucial observation is that instead of guessing a model $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$ of $\mathcal{K}$, it suffices to consider certain interpretations $\mathcal{I}'$ of polynomial size, derived from sub-interpretations of $\mathcal{I}$, and guess models $\mathcal{J}' <_{\mathsf{CP}} \mathcal{I}'$ of $\mathcal{K}$. Intuitively, we decompose an expensive 'global' test into exponentially many inexpensive 'local' tests. This even works in the presence of role inclusions, that is, in DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$, which shall be useful in Section 5.

Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be a DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ cKB with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ and $\mathcal{I}$ a model of $\mathcal{K}$. For each role r used in $\mathcal{T}$ such that $r^{\mathcal{I}} \neq \emptyset$, we choose a witness $w_r \in (\exists r^-)^{\mathcal{I}}$. Every $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$ gives rise to an interpretation $\mathcal{I}_{\mathcal{P}}$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{P}}} &= \mathcal{P} \cup \mathsf{ind}(\mathcal{A}) \cup \{w_r \mid r \text{ used in } \mathcal{T}, r^{\mathcal{I}} \neq \emptyset\} \\
A^{\mathcal{I}_{\mathcal{P}}} &= A^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}} \\
r^{\mathcal{I}_{\mathcal{P}}} &= r^{\mathcal{I}} \cap (\mathsf{ind}(\mathcal{A}) \times \mathsf{ind}(\mathcal{A})) \\
&\quad \cup \{(e, w_s) \mid e \in (\exists s)^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}}, \mathcal{T} \models s \sqsubseteq r\} \\
&\quad \cup \{(w_s, e) \mid e \in (\exists s)^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}}, \mathcal{T} \models s \sqsubseteq r^-\}.
\end{aligned}$$

Note that $\mathcal{I}_{\mathcal{P}}$ is derived from the subinterpretation $\mathcal{I}|_{\mathcal{P} \cup \mathsf{ind}(\mathcal{A})}$ by 'rerouting' some role edges to elements w_r. We have $|\mathcal{I}_{\mathcal{P}}| \leq |\mathcal{A}| + |\mathcal{T}| + |\mathcal{P}|$ and will only consider sets $\mathcal{P}$ with $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$. It is not difficult to show that $\mathcal{I}_{\mathcal{P}}$ is a model of $\mathcal{K}$, for all $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$. The next lemma characterizes the (non)-minimality of $\mathcal{I}$ in terms of the (non)-minimality of the interpretations $\mathcal{I}_{\mathcal{P}}$.

Lemma 9. *The following are equivalent:*

1. *There exists a model $\mathcal{I}$ of $\mathcal{K}$ with $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$;*
2. *There exist a $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$ with $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$ and a family $(\mathcal{J}_e)_{e \in \Delta^{\mathcal{I}}}$ of models of $\mathcal{K}$ such that $\mathcal{J}_e <_{\mathsf{CP}} \mathcal{I}_{\mathcal{P} \cup \{e\}}$ and $\mathcal{J}_e|_{\Delta^{\mathcal{I}_{\mathcal{P}}}} = \mathcal{J}_{e'}|_{\Delta^{\mathcal{I}_{\mathcal{P}}}}$ for all $e, e' \in \Delta^{\mathcal{I}}$.*

It should now be clear that we have established Theorem 5. After guessing the model $\mathcal{I}$ of $\mathcal{K}$ with $\mathcal{I} \not\models q(\bar{a})$, we check the complement of Point 2 of Lemma 9 in a brute-force way. More precisely, we first iterate over all $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$ with $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$, then over all interpretations $\mathcal{J}_0$ with $\Delta^{\mathcal{J}_0} = \Delta^{\mathcal{I}_{\mathcal{P}}}$ (as candidates for the common restriction of the models $(\mathcal{J}_e)_{e \in \Delta^{\mathcal{I}}}$ to $\Delta^{\mathcal{I}_{\mathcal{P}}}$), then over all $e \in \Delta^{\mathcal{I}}$, and finally over all models $\mathcal{J}_e$ of $\mathcal{K}$ with $\mathcal{J}_e|_{\Delta^{\mathcal{I}_{\mathcal{P}}}} = \mathcal{J}_0$, and test whether $\mathcal{J}_e <_{\mathsf{CP}} \mathcal{I}_{\mathcal{P} \cup \{e\}}$. We accept if for every $\mathcal{P}$ and $\mathcal{J}_0$, there is an e such that for all $\mathcal{J}_e$ the final check fails. Overall, we obtain a CONEXP algorithm.

We provide a matching lower bound that holds even for DL-Lite$_{\mathsf{core}}$ cKBs. It is proved by reduction from the complement of Succinct3COL, which is known to be NEXP-complete (Papadimitriou and Yannakakis 1986).

Theorem 7. *UCQ answering on circumscribed DL-Lite$_{\mathsf{core}}$ KBs is* CONEXP*-hard w.r.t. combined complexity. This holds even with a single minimized concept name, no fixed concept names, and no disjointness constraints.*

4.2 Data Complexity

We now consider data complexity, where the landscape is less diverse. Indeed, we obtain CONP-completeness for all DLs between DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ and DL-Lite$_{\mathsf{core}}$, and both for UCQs and CQs.

Theorem 8. *UCQ evaluation on circumscribed DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ KBs is in* CONP *w.r.t. data complexity.*

To prove Theorem 8, we again guess a model $\mathcal{I}$ of $\mathcal{K}$ with $\mathcal{I} \not\models q(\bar{a})$ and $|\Delta^{\mathcal{I}}|$ bounded as in Theorem 8, and then verify that $\mathcal{I}$ is minimal w.r.t. $<_{\mathsf{CP}}$. For the latter, we introduce a variation of Lemma 9. The original version of Lemma 9 is not helpful because its Point 2 involves deciding whether, given an interpretation $\mathcal{I}_{\mathcal{P} \cup \{e\}}$, there is a model $\mathcal{J}_e <_{\mathsf{CP}} \mathcal{I}_{\mathcal{P} \cup \{e\}}$ of $\mathcal{K}$, and given that $\mathsf{ind}(\mathcal{A}) \subseteq \mathcal{I}_{\mathcal{P} \cup \{e\}}$ there is no reason to believe that this can be done in polynomial time in data complexity. We actually conjecture this problem to be CONP-complete. In the following, we vary the definition of the interpretations $\mathcal{I}_{\mathcal{P}}$ so that their size is independent of that of $\mathcal{A}$.

Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be a DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ cKB with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ and $\mathcal{I}$ a model of $\mathcal{K}$. We assume $\mathcal{T}$ to be in normal form. For an element $e \in \Delta^{\mathcal{I}}$, we define its *ABox type* to be

$$\mathsf{tp}_{\mathcal{A}}(e) = \{A \mid A \in \mathsf{N_C}, \mathcal{K} \models A(e)\}.$$

Note that $\mathsf{tp}_{\mathcal{A}}(e)$ actually needs not be a proper type as defined in Section 3 due to the presence of disjunction in DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$. If $e \notin \mathsf{ind}(\mathcal{A})$, which is permitted, then $\mathsf{tp}_{\mathcal{A}}(e)$ is simply the set of all concepts names A with $\mathcal{K} \models \top \sqsubseteq A$.

For each pair (t_1, t_2) such that $\mathsf{tp}_{\mathcal{A}}(e) = t_1$ and $\mathsf{tp}_{\mathcal{I}}(e) = t_2$ for some $e \in \Delta^{\mathcal{I}}$, we choose such an element e_{t_1,t_2}; we use E to denote the set of all elements chosen in this way. For each role r used in $\mathcal{T}$ such that $r^{\mathcal{I}} \neq \emptyset$, we choose an element $w_r \in (\exists r^-)^{\mathcal{I}}$. Now define, for every $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$, the interpretation $\mathcal{I}_{\mathcal{P}}$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{P}}} &= \mathcal{P} \cup E \cup \{w_r \mid r \text{ used in } \mathcal{T}, r^{\mathcal{I}} \neq \emptyset\} \\
A^{\mathcal{I}_{\mathcal{P}}} &= A^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}} \\
r^{\mathcal{I}_{\mathcal{P}}} &= \{(e, w_s) \mid e \in (\exists s)^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}}, \mathcal{T} \models s \sqsubseteq r\} \\
&\quad \cup \{(w_s, e) \mid e \in (\exists s)^{\mathcal{I}} \cap \Delta^{\mathcal{I}_{\mathcal{P}}}, \mathcal{T} \models s \sqsubseteq r^-\}
\end{aligned}$$

Note that $|\mathcal{I}_{\mathcal{P}}| \leq 4^{|\mathcal{T}|} + |\mathcal{T}| + |\mathcal{P}|$. We also define an ABox

$$\mathcal{A}_{\mathcal{P}} = \{A(a) \mid a \in \mathsf{ind}(\mathcal{A}) \cap \Delta^{\mathcal{I}_{\mathcal{P}}}, A \in \mathsf{tp}_{\mathcal{A}}(a)\}$$

and set $\mathcal{K}_{\mathcal{P}} = (\mathcal{T}, \mathcal{A}_{\mathcal{P}})$. The ABoxes $\mathcal{A}_{\mathcal{P}}$ act as a decomposition of the ABox $\mathcal{A}$, similarly to how the interpretations $\mathcal{I}_{\mathcal{P}}$ act as a decomposition of the interpretation $\mathcal{I}$. Note that $\bigcup_{\mathcal{P}} \mathcal{A}_{\mathcal{P}}$ does not contain role assertions. This is compensated by the use of $\mathsf{tp}_{\mathcal{A}}(a)$ in the definition of $\mathcal{A}_{\mathcal{P}}$ and the fact that, since $\mathcal{T}$ is in normal form, all relevant endpoints of role assertions are 'visible' in the ABox types. Note that this fails in the case of $\mathcal{EL}$ where dropping role assertions could result in CIs $\exists r.B \sqsubseteq A$ to be left unsatisfied, and such CIs are crucial for proving Π^{P}_2-hardness in that case, see the proof of Theorem 4.

The following is the announced variation of Lemma 9.

Lemma 10. *The following are equivalent:*

1. $\mathcal{I} \models \mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$*;*
2. $\mathcal{I}_{\mathcal{P}} \models \mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}_{\mathcal{P}})$ *for all* $\mathcal{P} \subseteq \Delta^{\mathcal{I}}$ *with* $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$.

In the "$1 \Rightarrow 2$" direction we use the witnesses e_{t_1,t_2} to extend a potential model $\mathcal{J}' <_{\mathsf{CP}} \mathcal{I}_{\mathcal{P}}$ of some $\mathcal{K}_{\mathcal{P}}$ to a model $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$ of $\mathcal{K}$, obtaining a contradiction. In the "$2 \Rightarrow 1$" direction, for a potential model $\mathcal{J} <_{\mathsf{CP}} \mathcal{I}$ of $\mathcal{K}$, we can find a $\mathcal{P}$ with $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$ so that we can construct from $\mathcal{J}$ a model $\mathcal{J}' <_{\mathsf{CP}} \mathcal{I}_{\mathcal{P}}$ of $\mathcal{K}_{\mathcal{P}}$.

To establish Theorem 8, it thus suffices to argue that checking Point 2 of Lemma 10 can be implemented in time polynomial in $|\mathcal{A}|$. We iterate over all polynomially many sets $\mathcal{P}$ with $|\mathcal{P}| \leq 2|\mathcal{T}| + 1$ (recall that $\mathcal{T}$ is fixed) and check whether $\mathcal{I}_{\mathcal{P}} \models \mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}_{\mathcal{P}})$ in a brute force way. To verify the minimality of $\mathcal{I}_{\mathcal{P}}$, we iterate over all models $\mathcal{J}$ of $\mathcal{K}_{\mathcal{P}}$ with $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}_{\mathcal{P}}}$ (of which there are only polynomially many, thanks to the modified definition of $\Delta^{\mathcal{I}_{\mathcal{P}}}$), and make sure that $\mathcal{J} \not<_{\mathsf{CP}} \mathcal{I}_{\mathcal{P}}$ for any such $\mathcal{J}$. Overall, we obtain a CONP algorithm.

The CONP upper bound turns out to be tight, even for DL-Lite$_{\mathsf{core}}$ cKBs and CQs, and with very restricted circumscription patterns. We reduce from 3-colorability.

Theorem 9. *CQ evaluation on circumscribed DL-Lite*$_{\mathsf{core}}$ *KBs is* CONP*-hard w.r.t. data complexity. This holds even with a single minimized concept name, no fixed concept names, and no disjointness constraints.*

5 Atomic Queries

We study the evaluation of atomic queries on circumscribed KBs, which is closely related to concept satisfiability w.r.t. such KBs. In fact, the two problems are mutually reducible in polynomial time. Our results from this section, summarized in Table 2, can thus also be viewed as completing the complexity landscape for concept satisfiability, first studied in (Bonatti, Lutz, and Wolter 2006; Bonatti, Faella, and Sauro 2011).

5.1 Between $\mathcal{ALCHIO}$ and $\mathcal{EL}$

Concept satisfiability w.r.t. circumscribed $\mathcal{ALCIO}$ KBs was proved to be CONEXP$^{\mathsf{NP}}$-complete in (Bonatti, Lutz, and Wolter 2009). The proof of the upper bound can easily be extended to cover also role inclusions. We thus obtain:

Theorem 10. (Bonatti, Lutz, and Wolter 2009) *AQ evaluation on circumscribed $\mathcal{ALCHIO}$ KBs is in* CONEXP$^{\mathsf{NP}}$ *w.r.t. combined complexity.*

An alternative way to obtain Theorem 10 for $\mathcal{ALCHI}$ is to use Lemma 6, which yields the existence of single exponentially large countermodels in the special case where the query q is of constant size (here $|q| = 1$), and a straightforward guess-and-check procedure as sketched in Section 4.1. This can be lifted to $\mathcal{ALCHIO}$ by a variation of (the proof of) Proposition 1 tailored to AQs.

We next prove a matching lower bound for $\mathcal{EL}$, improving on an EXP lower bound from (Bonatti, Faella, and Sauro 2011).

Theorem 11. *AQ evaluation on circumscribed $\mathcal{EL}$ KBs is* CONEXP$^{\mathsf{NP}}$*-hard w.r.t. combined complexity. This holds even without fixed concept names and with an empty preference order.*

This is proved by a reduction from AQ evaluation on $\mathcal{ALC}$ cKBs, which is known to be CONEXP$^{\mathsf{NP}}$-hard (Bonatti, Lutz, and Wolter 2009).

Regarding data complexity, it suffices to recall that Theorem 4 applies even to AQs and $\mathcal{EL}$ cKBs.

5.2 DL-Lite

Recall that in Section 4.1, we have proved that UCQ evaluation over DL-Lite$_{\mathsf{bool}}$ cKBs is in CONEXP w.r.t. combined complexity. We started with a guess-and-check procedure that gives a CONEXP$^{\mathsf{NP}}$ upper bound, relying on countermodels of single exponential size as per Lemma 8, and then improved to CONEXP using Lemma 9. Here, we use the same algorithm. The only difference in the correctness proof is that Lemma 8 is replaced with Lemma 6 as the former does not support role inclusions and the latter delivers a single exponential upper bound for queries of constant size.

Theorem 12. *AQ evaluation on circumscribed DL-Lite*$^{\mathcal{H}}_{\mathsf{bool}}$ *KBs is in* CONEXP *w.r.t. combined complexity.*

We match this upper bound even in the absence of role inclusions, demonstrating that evaluating AQs and UCQs over DL-Lite$_{\mathsf{bool}}$ cKBs is equally difficult.

Theorem 13. *AQ evaluation on circumscribed DL-Lite*$_{\mathsf{bool}}$ *KBs is* CONEXP*-hard w.r.t. combined complexity.*

	$\mathcal{EL}$, $\mathcal{ALCHIO}$	DL-Lite$_{\mathsf{bool}}$, DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$	DL-Lite$_{\mathsf{core}}$, DL-Lite$^{\mathcal{H}}_{\mathsf{horn}}$
Combined complexity	coNExp$^{\text{NP}}$-c.$^{(\dagger)}$ (Thm. 10, 11)	coNExp-c. (Thm. 12, 13)	Π^{P}_2-c.$^{(\ddagger)}$ (Thm. 14)
Data complexity	Π^{P}_2-c. (Thm. 3, 4)	in PTime (Thm. 15)	in PTime (Thm. 15)

Table 2: Complexity of AQ evaluation on circumscribed KBs. $^{(\dagger)}$: completeness already known for $\mathcal{ALC}(\mathcal{IO})$. $^{(\ddagger)}$: hardness already known.

The proof is by reduction from the complement of the NExp-complete problem Succinct-3COL and relies on fixed concept names.

The situation is more favorable if we restrict our attention to DL-Lite$^{\mathcal{H}}_{\mathsf{horn}}$ cKBs, which is still a very expressive dialect of the DL-Lite family. For DL-Lite$^{\mathcal{H}}_{\mathsf{horn}}$, we prove that if there is a countermodel, then there is one of linear size, in drastic contrast to the DL-Lite$_{\mathsf{bool}}$ case where the proof of Theorem 13 shows that exponentially large countermodels cannot be avoided.

Theorem 14. *AQ evaluation on circumscribed DL-Lite$^{\mathcal{H}}_{\mathsf{horn}}$ KBs is in Π^{P}_2 w.r.t. combined complexity.*

A matching lower bound for DL-Lite$_{\mathsf{core}}$ cKBs can be found in (Bonatti, Faella, and Sauro 2011).

We now move to data complexity, where we obtain tractability even for the maximally expressive DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ dialect of DL-Lite.

Theorem 15. *AQ evaluation on circumscribed DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ KBs is in* PTime *w.r.t. data complexity.*

We sketch the proof of Theorem 15. Let $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K})$ be a DL-Lite$^{\mathcal{H}}_{\mathsf{bool}}$ cKB with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $A_0(x)$ an AQ, and $a_0 \in \mathsf{Ind}(\mathcal{A})$. We construct an ABox $\mathcal{A}'$ whose size is independent of that of $\mathcal{A}$ and which can replace $\mathcal{A}$ when deciding whether a_0 is an answer to A_0. The construction of $\mathcal{A}'$ may be viewed as a variation of the constructions that we have used in the proofs of Theorems 6 and 8 to avoid an NP oracle. In particular, it reuses the notion of ABox types from the latter. However, the construction employed here works directly with ABoxes rather than with countermodels.

Let $\mathsf{TP}(\mathcal{A})$ denote the set of all ABox types t realized in $\mathcal{A}$, that is, all t such that $\mathsf{tp}_{\mathcal{A}}(a) = t$ for some $a \in \mathsf{Ind}(\mathcal{A})$. For every $t \in \mathsf{TP}(\mathcal{A})$, set

$$m_t = \min(|\{a \in \mathsf{ind}(\mathcal{A}) \mid \mathsf{tp}_{\mathcal{A}}(a) = t\}|,\ 4^{|\mathcal{T}|})$$

and choose m_t individuals $a_{t,1}, \ldots, a_{t,m_t} \in \mathsf{ind}(\mathcal{A})$ such that $\mathsf{tp}_{\mathcal{A}}(a_{t,i}) = t$ for $1 \leq i \leq m_t$. We assume that the individual a_0 of interest is among the chosen ones. Let W be the set of all chosen individuals, that is, $W = \{a_{t,i} \mid t \in \mathsf{TP}(\mathcal{A}),\ 1 \leq i \leq m_t\}$. Now define an ABox

$$\mathcal{A}' \quad = \quad \{A(a) \mid a \in W, A \in \mathsf{tp}_{\mathcal{A}}(a)\}$$

and set $\mathcal{K}' = (\mathcal{T}, \mathcal{A}')$. Note that $|\mathsf{ind}(\mathcal{A}')| \leq 8^{|\mathcal{T}|}$, and thus the size of $\mathcal{A}'$ depends only on the TBox $\mathcal{T}$, but not on the original ABox $\mathcal{A}$. Also, note that, just like the ABoxes $\mathcal{A}_{\mathcal{P}}$ from the proof of Theorem 8, $\mathcal{A}'$ no longer contains role assertions.

Lemma 11. $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}) \models A_0(a_0)$ *iff* $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}') \models A_0(a_0)$.

It is important for the 'only if' direction of Lemma 11 that we keep at least $4^{|\mathcal{T}|}$ individuals of each ABox type t (if existent). In fact, this allows us to convert a model $\mathcal{I}'$ of $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}')$ into a model of $\mathcal{K}$ that is minimal w.r.t. $<_{\mathsf{CP}}$, using arguments similar to those in the proof of Lemma 2. A crucial point is that if $a \in \mathsf{ind}(\mathcal{A}) \setminus \mathsf{ind}(\mathcal{A}')$ and $t = \mathsf{tp}_{\mathcal{A}}(a)$, then $m_t = 4^{|\mathcal{T}|}$ which implies that there is a regular type t' such that the combination (t, t') is realized at least $2^{|\mathcal{T}|}$ many times in $\mathcal{I}'$ among the individuals from $\mathcal{A}'$.

Lemma 11 gives PTime membership as the size of $\mathcal{A}'$ is bounded by a constant. We compute $\mathcal{K}'$ in polynomial time and check whether $\mathsf{Circ}_{\mathsf{CP}}(\mathcal{K}') \models A_0(a_0)$, which can be decided in 2Exp by Theorem 1, that is, in constant time w.r.t. data complexity. The correctness of this procedure immediately follows from Lemma 11.

5.3 Negative Role Inclusions

DL-Lite is often defined to additionally include negative role inclusions of the form $r \sqsubseteq \neg s$, with the obvious semantics. It is known that these sometimes lead to increased computational complexity; see, for example, (Manière 2022). We close by observing that this is also the case for circumscription. While querying circumscribed DL-Lite KBs (in all considered dialects) is coNP-complete w.r.t. data complexity for (U)CQs and in PTime for AQs, adding negative role inclusions results in a jump back to Π^{P}_2. We prove this by reduction from $\forall\exists$3SAT. Some ideas are shared with the proof of Theorem 4, but the general strategy of the reduction is different.

Theorem 16. *AQ evaluation on circumscribed DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ KBs with negative role inclusions is Π^{P}_2-hard. This holds even without fixed concept names and with a single negative role inclusion.*

6 Conclusion

We have provided a rather complete picture of the complexity of query evaluation on circumscribed KBs. Some cases, however, remain open. For example, the lower bounds in combined complexity for UCQ evaluation on DL-Lite cKBs given in Theorems 5 and 7 cannot be improved in an obvious way to CQs, for which the complexity remains open. Also, the lower bounds provided in Theorems 13, 16 and the Π^{P}_2 one from (Bonatti, Faella, and Sauro 2011) rely on the preference relation in circumscription patterns, and it remains open whether the complexity decreases when the preference relation is forced to be empty. Finally, it would be interesting to study query evaluation under ontologies that are sets of existential rules or formulated in the guarded (negation) fragment of first-order logic, extended with circumscription. We believe that these problems are still decidable.

Acknowledgments

The research reported in this paper has been supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 Project-ID 329551904 "EASE - Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/). The research was conducted in subprojects "P02 – Ontologies with Abstraction" and "P05-N – Principles of Metareasoning for Everyday Activities".

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI

This work is partly supported by BMBF (Federal Ministry of Education and Research) in DAAD project 57616814 (SECAI, School of Embedded Composite AI) as part of the program Konrad Zuse Schools of Excellence in Artificial Intelligence.

References

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Bonatti, P. A., and Samarati, P. 2003. Logics for authorization and security. In *Logics for Emerging Applications of Databases*, 277–323. Springer.

Bonatti, P. A.; Faella, M.; Lutz, C.; Sauro, L.; and Wolter, F. 2015a. Decidability of circumscribed description logics revisited. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation - Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, volume 9060 of *LNCS*, 112–124. Springer.

Bonatti, P. A.; Faella, M.; Petrova, I. M.; and Sauro, L. 2015b. A new semantics for overriding in description logics. *Artif. Intell.* 222:1–48.

Bonatti, P. A.; Faella, M.; and Sauro, L. 2011. Defeasible inclusions in low-complexity DLs. *J. Artif. Intell. Res.* 42:719–764.

Bonatti, P.; Lutz, C.; and Wolter, F. 2006. Description logics with circumscription. In *Proc. of KR2006*, 400–410. AAAI Press.

Bonatti, P. A.; Lutz, C.; and Wolter, F. 2009. The complexity of circumscription in description logic. *J. Artif. Intell. Res.* 35:717–773.

Bonatti, P. A. 2021. Query answering in circumscribed OWL2 profiles. *Annals of Mathematics and Artificial Intelligence* 89(12):1155–1173.

Casini, G., and Straccia, U. 2013. Defeasible inheritance-based description logics. *J. Artif. Intell. Res.* 48:415–473.

Eiter, T., and Gottlob, G. 1993. Propositional circumscription and extended closed-world reasoning are Π_2^{P}-complete. *Theor. Comput. Sci.* 114(2):231–245.

Giordano, L.; Gliozzi, V.; Olivetti, N.; and Pozzato, G. L. 2013. A non-monotonic description logic for reasoning about typicality. *Artif. Intell.* 195:165–202.

Lifschitz, V. 1985. Closed-world databases and circumscription. *Artif. Intell.* 27(2):229–235.

Lutz, C.; Manière, Q.; and Nolte, R. 2023. Querying circumscribed description logic knowledge bases. arXiv:2306.04546 [cs.AI].

Lutz, C.; Seylan, I.; and Wolter, F. 2013. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *Proc. of IJCAI*, 1024–1030. IJCAI/AAAI.

Lutz, C.; Seylan, I.; and Wolter, F. 2019. The data complexity of ontology-mediated queries with closed predicates. *Log. Methods Comput. Sci.* 15(3).

Manière, Q. 2022. *Counting queries in ontology-based data access*. Ph.D. Dissertation, Université de Bordeaux.

Motik, B.; Grau, B. C.; Horrocks, I.; Wu, Z.; Fokoue, A.; and Lutz, C., eds. 2009. *OWL 2 Web Ontology Language Profiles (Second Edition)*. W3C Recommendation. Available at https://www.w3.org/TR/owl2-profiles/.

Ngo, N.; Ortiz, M.; and Simkus, M. 2016. Closed predicates in description logics: Results on combined complexity. In *Proc. of KR*, 237–246. AAAI Press.

Papadimitriou, C. H., and Yannakakis, M. 1986. A note on succinct representations of graphs. *Information and Control* 71(3):181–185.

Pensel, M., and Turhan, A. 2018. Reasoning in the defeasible description logic $\mathcal{EL}$ - computing standard inferences under rational and relevant semantics. *Int. J. Approx. Reason.* 103:28–70.

Rector, A. L. 2004. Defaults, context, and knowledge: Alternatives for OWL-indexed knowledge bases. In *Proc. of the Pac. Symp. in Biocomp.*, 226–237. World Scientific.

Reiter, R. 1977. On closed world data bases. In *Symposium on Logic and Databases*, Advances in Database Theory, 55–76. New York: Plemum Press.

Stefano, F. D.; Ortiz, M.; and Simkus, M. 2022. Pointwise circumscription in description logics. In *Proc. of DL2022*, volume 3263 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Stevens, R.; Aranguren, M. E.; Wolstencroft, K.; Sattler, U.; Drummond, N.; Horridge, M.; and Rector, A. L. 2007. Using OWL to model biological knowledge. *Int. J. Hum. Comput. Stud.* 65(7):583–594.