Finding New Diamonds: Temporal Minimal-World Query Answering over Sparse ABoxes*

Stefan Borgwardt, Walter Forkel, and Alisa Kovtunova

Chair for Automata Theory, Technische Universität Dresden, Germany
firstname.lastname@tu-dresden.de

Abstract. Lightweight temporal ontology languages have become a very active field of research in recent years. Many real-world applications, like processing electronic health records (EHRs), inherently contain a temporal dimension, and require efficient reasoning algorithms. Moreover, since medical data is not recorded on a regular basis, reasoners must deal with sparse data with potentially large temporal gaps. In this paper, we introduce a temporal extension of the tractable language $\mathcal{ELH}_\bot$, which features a new class of *convex diamond* operators that can be used to bridge temporal gaps. We develop a completion algorithm for our logic, which shows that entailment remains tractable. Based on this, we develop a *minimal-world* semantics for answering metric temporal conjunctive queries with negation. We show that query answering is combined first-order rewritable, and hence in polynomial time in data complexity.

1 Introduction

Temporal description logics (DLs) combine terminological and temporal knowledge representation capabilities and have been investigated in detail in the last decades [3, 28, 32]. To obtain tractable reasoning procedures, lightweight temporal DLs have been developed [4, 20]. The idea is to use temporal operators, often from the linear temporal logic LTL, inside DL axioms. For example, $\diamondsuit\exists$diagnosis.BrokenLeg $\sqsubseteq \exists$treatment.LegCast states that after breaking a leg one has to wear a cast. However, this basic approach cannot represent the distance of events, e.g., that the cast only has to be worn for a fixed amount of time. Recently, metric temporal ontology languages have been investigated [7, 14, 21], which allow to replace $\diamondsuit$ in the above axiom with $\diamondsuit_{[-8,0]}$, i.e., wearing the cast is required only if the leg was broken ≤ 8 time points (e.g., weeks) ago.

Such knowledge representation capabilities are important for biomedical applications. For example, many clinical trials contain temporal eligibility criteria [16] such as: "type 1 diabetes with duration at least 12 months"[1]; "known history of heart disease or heart rhythm abnormalities"[2]; "CD4+ lymphocytes count > 250/mm3, for at least 6 months"[3]; or "symptomatic recurrent paroxysmal

* This work was partially supported by DFG grant 389792660 as part of TRR 248 and the DFG project BA 1122/19-1 (GOASQ).

[1] https://clinicaltrials.gov/ct2/show/NCT02280564

[2] https://clinicaltrials.gov/ct2/show/NCT02873052

[3] https://clinicaltrials.gov/ct2/show/NCT02157311

atrial fibrillation (PAF) (>2 episodes in the last 6 months)"[4]. Moreover, measurements, diagnoses, and treatments in a patients' EHR are clearly valid only for a certain amount of time. To automatically screen patients according to the temporal criteria above, one needs a sufficiently powerful formalism that can reason about biomedical and temporal knowledge. This is an active area of current research [11,16,22]. For the atemporal part, one can use existing large biomedical ontologies that are based on lightweight (atemporal) DLs, e.g., SNOMED CT[5], which is formulated using the DL $\mathcal{ELH}$.

Since EHRs only contain information for specific points in time, it is especially important to be able to infer what happened to the patient in the meantime. For example, if a patient is diagnosed with a (currently) incurable disease like Diabetes, they will still have the disease at any future point in time. Similarly, if the EHR contains two entries of CD4Above250 four weeks apart, one may reasonably infer that this was true for the whole four weeks. Qualitative temporal DLs such as $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ [20] can express the former statement by declaring Diabetes as *expanding* via the axiom $\Diamond\text{Diabetes} \sqsubseteq \text{Diabetes}$. We propose to extend this logic by adding a special kind of metric temporal operators to write $\diamondsuit_4\text{CD4Above250} \sqsubseteq \text{CD4Above250}$, making the measurement *convex* for a specified length of time n (e.g., 4 weeks). This means that information is interpolated between time points of distance less than n, thereby computing a convex closure of the available information. The threshold n allows us to distinguish the case where two mentions of CD4Above250 are years apart, and are therefore unrelated.

The distinguishing feature of $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ is that $\Diamond$-operators are only allowed on the left-hand side of concept inclusions [20], which is also common for temporal DLs based on *DL-Lite* [2,5]. Apart from adding convex metric temporal operators to this logic, we allow temporal roles like $\diamondsuit_2\text{hasTreatment} \sqsubseteq \text{hasTreatment}$, and deal with the problem of having large gaps in the data, e.g., in patient records. We show that reasoning in the extended logic $\mathcal{TELH}^{\diamondsuit,\mathsf{lhs}}_{\bot}$ remains tractable.

Additionally, we consider the problem of answering temporal queries over $\mathcal{TELH}^{\diamondsuit,\mathsf{lhs}}_{\bot}$ knowledge bases. As argued in [6,12], evaluating clinical trial criteria over patient records requires both negated and temporal queries, but standard certain answer semantics is not suitable to deal with negation over patient records, which is why we adopt the *minimal-world* semantics from [12] for our purposes. Our query language extends the temporal conjunctive queries from [8] by metric temporal operators [7,21] and negation. For example, we can use queries like $\Box_{[-12,0]}(\exists y.\text{diagnosedWith}(x,y) \wedge \text{Diabetes}(y))$ to detect whether the first criterion from above is satisfied.

Using a combined rewriting approach, we show that the data complexity of query answering is not higher than for positive atemporal queries in $\mathcal{ELH}_{\bot}$, and also provide a tight combined complexity result of ExpSpace. Unlike most research on temporal query answering [2,8], we do not assume that input data is given for all time points in a certain interval, but rather at sporadic time points with arbitrarily large gaps. The main technical difficulty is to determine which

[4] `https://clinicaltrials.gov/ct2/show/NCT00969735`
[5] `https://www.snomed.org/`

additional time points are relevant for answering a query, and how to access these time points without having to fill all the gaps.

Full proofs can be found in the extended version at `https://tu-dresden.de/inf/lat/papers`.

2 The Lightweight Temporal Logic $\mathcal{TELH}_{\bot}^{\Diamond^{c},\mathsf{lhs}}$

We first introduce the metric LTL operators that we will use and analyze their properties. LTL formulas are formulated over a finite set P of *propositional variables*. In this section, we consider only formulas built according to the syntax rule $\varphi ::= p \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Diamond_I \varphi$, where $p \in P$ and I is an interval in $\mathbb{Z}$. The semantics is given by *LTL-structures* $\mathfrak{W} = (w_i)_{i \in \mathbb{Z}}$, where $w_i \subseteq P$. We write

$$\mathfrak{W}, i \models p \text{ iff } p \in w_i \text{ if } p \in P, \qquad \mathfrak{W}, i \models \varphi \wedge \psi \text{ iff } \mathfrak{W}, i \models \varphi \text{ and } \mathfrak{W}, i \models \psi,$$
$$\mathfrak{W}, i \models \Diamond_I \varphi \text{ iff } \exists k \in I\colon \mathfrak{W}, i+k \models \varphi, \qquad \mathfrak{W}, i \models \varphi \vee \psi \text{ iff } \mathfrak{W}, i \models \varphi \text{ or } \mathfrak{W}, i \models \psi.$$

More specifically, we only consider the following derived operators, where $n \geq 1$:

$$\Diamond^{\pm}\varphi := \Diamond_{(-\infty,\infty)}\varphi \qquad \Diamond^{+}\varphi := \Diamond_{[0,\infty)}\varphi \qquad \Diamond^{-}\varphi := \Diamond_{(-\infty,0]}\varphi$$
$$\Diamond^{c}\varphi := \Diamond_{(-\infty,0]}\varphi \wedge \Diamond_{[0,\infty)}\varphi \qquad \Diamond^{c}_n\varphi := \bigvee_{\substack{k,m \geq 0 \\ k+m=n-1}} \left(\Diamond_{[-k,0]}\varphi \wedge \Diamond_{[0,m]}\varphi\right) \tag{1}$$

The operator $\Diamond^{+}$ is the "eventually" operator of classical LTL, and $\Diamond^{-}, \Diamond^{\pm}$ are two variants that refer to the past, or to both past and future, respectively. The operator $\Diamond^{c}$ requires that φ holds *both* in the past and in the future, thereby distinguishing time points that lie within an interval enclosed by time points at which φ holds. This can be used to express the convex closure of time points, as described in the introduction. Finally, the operators $\Diamond^{c}_n$ represent a metric variant of $\Diamond^{c}$, requiring that different occurrences of φ are at most $n-1$ time points apart, i.e., enclose an interval of length n. To study the behavior of these operators, we consider their semantics in a more abstract way: given a set of time points where a certain information is available (e.g., a diagnosis), described by a propositional variable p, we consider the resulting set of time points at which $\Diamond^{\star}p$ holds, where $\Diamond^{\star}$ is a placeholder for one of the operators defined above (we will similarly use $\Diamond^{\bullet}, \Diamond^{\dagger}, \Diamond^{\ddagger}$ as placeholders for different $\Diamond$-operators in the following).

Definition 1. *We consider the sets* $\mathfrak{D}^{c} := \{\Diamond^{c}\} \cup \{\Diamond^{c}_i \mid i \geq 1\}$, $\mathfrak{D}^{\pm} = \{\Diamond^{-}, \Diamond^{+}, \Diamond^{\pm}\}$, *and* $\mathfrak{D} := \mathfrak{D}^{\pm} \cup \mathfrak{D}^{c}$ *of* diamond operators. *Each* $\Diamond^{\star} \in \mathfrak{D}$ *induces a function* $\Diamond^{\star}\colon 2^{\mathbb{Z}} \to 2^{\mathbb{Z}}$ *with* $\Diamond^{\star}(M) := \{i \mid \mathfrak{W}_M, i \models \Diamond^{\star}p\}$ *for all* $M \subseteq \mathbb{Z}$, *with the LTL-structure* $\mathfrak{W}_M := (w_i)_{i \in \mathbb{Z}}$ *such that* $w_i := \{p\}$ *if* $i \in M$, *and* $w_i := \emptyset$ *otherwise.*

We will omit the parentheses in $\Diamond^{\star}(M)$ for a cleaner presentation. If M is empty, then $\Diamond^{\star}M$ is also empty, for any $\Diamond^{\star} \in \mathfrak{D}$. For any non-empty $M \subseteq \mathbb{Z}$, we obtain the following expressions, where $\max M$ may be ∞ and $\min M$ may be $-\infty$.

$$\Diamond^{\pm}M = \mathbb{Z} \qquad \Diamond^{-}M = (-\infty, \max M] \qquad \Diamond^{+}M = [\min M, \infty) \qquad \Diamond^{c}M = [\min M, \max M]$$
$$\Diamond^{c}_1 M = M \qquad \Diamond^{c}_n M = \{i \in \mathbb{Z} \mid \exists j, k \in M \text{ with } j \leq i \leq k \text{ and } k - j < n\}$$

In this representation, the convex closure operation behind $\diamondc$ becomes apparent. We now list several useful properties of these functions.

Lemma 2. *Using the pointwise inclusion order* $\subseteq$ *on the induced functions, we obtain the following ordered set* $(\mathfrak{D}, \subseteq)$*, where* $\mathrm{id}_{2^{\mathbb{Z}}}$ *is the identity function on* $2^{\mathbb{Z}}$*:*

$$\mathrm{id}_{2^{\mathbb{Z}}} = \diamondc_1 \subseteq \cdots \subseteq \diamondc_n \subseteq \diamondc_{n+1} \subseteq \cdots \subseteq \diamondc \;\begin{matrix} & \diamondminus & \\ \subseteq & & \subseteq \\ & & & \diamondpm \\ \subseteq & & \subseteq \\ & \diamondplus & \end{matrix}$$

The most important property is the following, which allows us to combine diamond operators without leaving the set $\mathfrak{D}$.

Lemma 3. *The set* $\mathfrak{D}$ *is closed under composition* $\circ$*, pointwise intersection* $\cap$*, and pointwise union* $\cup$*, and for any* $\diamondstar, \diamonddot \in \mathfrak{D}$ *these operators can be computed as:*

$$\diamondstar \cap \diamonddot = \inf_{(\mathfrak{D},\subseteq)}\{\diamondstar, \diamonddot\} \qquad \textit{and} \qquad \diamonddot \circ \diamondstar = \diamondstar \cup \diamonddot = \sup_{(\mathfrak{D},\subseteq)}\{\diamondstar, \diamonddot\},$$

where $\inf_{(\mathfrak{D},\subseteq)}$ *denotes the infimum in* $(\mathfrak{D}, \subseteq)$*, and* $\sup_{(\mathfrak{D},\subseteq)}$ *the supremum.*

2.1 A New Temporal Description Logic

We define a new temporal description logic based on the operators in $\mathfrak{D}$. The main differences to $\mathcal{TEL}^{\Diamond}_{\mathsf{infl}}$ from [20] are that $\diamondc_n$-operators may occur in concept and role inclusions, and ABoxes may have gaps, which require special consideration during reasoning.

Syntax. Let $\mathbf{C}, \mathbf{R}, \mathbf{I}$ be disjoint sets of *concept*, *role*, and *individual names*, respectively. A *temporal role* is of the form $\diamondstar r$ with $\diamondstar \in \mathfrak{D}$ and $r \in \mathbf{R}$. A $\mathcal{TELH}^{\diamondstar,\mathsf{lhs}}_{\bot}$ *concept* is built using the rule $C ::= A \mid \top \mid \bot \mid C \sqcap C \mid \exists r.C \mid \diamondstar C$, where $A \in \mathbf{C}$, $\diamondstar \in \mathfrak{D}$, and r is a temporal role. Such a C is an $\mathcal{ELH}_{\bot}$ *concept* (or *atemporal concept*) if it does not contain any diamond operators.

A $\mathcal{TELH}^{\diamondstar,\mathsf{lhs}}_{\bot}$ *TBox* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$ and *role inclusions (RIs)* $r \sqsubseteq s$, where C is a $\mathcal{TELH}^{\diamondstar,\mathsf{lhs}}_{\bot}$ concept, D is an atemporal concept, r is a temporal role, and $s \in \mathbf{R}$. We write $C \equiv D$ to abbreviate the two inclusions $C \sqsubseteq D$, $D \sqsubseteq C$, and similarly for role inclusions. An *ABox* is a finite set of *concept assertions* $A(a, i)$ and *role assertions* $r(a, b, i)$, where $A \in \mathbf{C}$, $r \in \mathbf{R}$, $a, b \in \mathbf{I}$, and $i \in \mathbb{Z}$. We denote the set of time points $i \in \mathbb{Z}$ occurring in $\mathcal{A}$ by $\mathrm{tem}(\mathcal{A})$. Additionally, we assume that each time point is encoded in binary with at most n digits. A *knowledge base (KB)* (or *ontology*) $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. In the following, we always assume a KB $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ to be given.

Semantics. An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ has a *domain* $\Delta^{\mathcal{I}} \supseteq \mathbf{I}$ and assigns to each $A \in \mathbf{C}$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to each $r \in \mathbf{R}$ a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. A *temporal interpretation* $\mathfrak{I} = (\Delta^{\mathfrak{I}}, (\mathcal{I}_i)_{i \in \mathbb{Z}})$, is a collection of interpretations

$\mathcal{I}_i = (\Delta^{\mathfrak{I}}, \cdot^{\mathcal{I}_i})$, $i \in \mathbb{Z}$, over $\Delta^{\mathfrak{I}}$. The functions $\cdot^{\mathcal{I}_i}$ are extended as follows.

$$(\Diamond^{*} r)^{\mathcal{I}_i} := \{(d,e) \in \Delta^{\mathfrak{I}} \times \Delta^{\mathfrak{I}} \mid i \in \Diamond^{*}\{j \mid (d,e) \in r^{\mathcal{I}_j}\}\} \qquad \top^{\mathcal{I}_i} := \Delta^{\mathfrak{I}} \qquad \bot^{\mathcal{I}_i} := \emptyset$$

$$(C \sqcap D)^{\mathcal{I}_i} := C^{\mathcal{I}_i} \cap D^{\mathcal{I}_i} \qquad (\exists r.C)^{\mathcal{I}_i} := \{d \in \Delta^{\mathfrak{I}} \mid \exists e \in C^{\mathcal{I}_i}\colon (d,e) \in r^{\mathcal{I}_i}\}$$

$$(\Diamond^{*} C)^{\mathcal{I}_i} := \{d \in \Delta^{\mathfrak{I}} \mid i \in \Diamond^{*}\{j \mid d \in C^{\mathcal{I}_j}\}\}$$

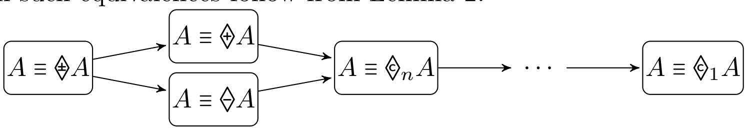
$\mathfrak{I}$ is a *model* of (or *satisfies*) a concept inclusion $C \sqsubseteq D$ if $C^{\mathcal{I}_i} \subseteq D^{\mathcal{I}_i}$ holds for all $i \in \mathbb{Z}$, a role inclusion $r \sqsubseteq s$ if $r^{\mathcal{I}_i} \subseteq s^{\mathcal{I}_i}$ holds for all $i \in \mathbb{Z}$, a concept assertion $A(a,i)$ if $a \in A^{\mathcal{I}_i}$, a role assertion $r(a,b,i)$ if $(a,b) \in r^{\mathcal{I}_i}$, and the KB $\mathcal{K}$ if it satisfies all axioms in $\mathcal{K}$. This fact is denoted by $\mathfrak{I} \models \alpha$, where α is an *axiom* (i.e., inclusion or assertion) or a KB. An ontology $\mathcal{K}$ is *consistent* if it has a model, and it *entails* α, written $\mathcal{K} \models \alpha$, if all models of $\mathcal{K}$ satisfy α. $\mathcal{K}$ is inconsistent iff $\mathcal{K} \models \top \sqsubseteq \bot$, and thus we focus on deciding entailment. In $\mathcal{ELH}_\bot$, this is possible in polynomial time [9].

We do not allow diamonds to occur on the right-hand side of CIs, because that would make the logic undecidable [4]. As usual, we can simulate CIs involving complex concepts by introducing fresh concept and role names as abbreviations. For example, $\exists \Diamond^{\pm} r.\Diamond^{+} A \sqsubseteq B$ can be split into $\Diamond^{\pm} r \sqsubseteq r'$, $\Diamond^{+} A \sqsubseteq A'$, and $\exists r'.A' \sqsubseteq B$. Hence, we can restrict ourselves w.l.o.g. to CIs in the following *normal form*:

$$\Diamond^{*} A \sqsubseteq B,\ A_1 \sqcap A_2 \sqsubseteq B,\ \Diamond^{*} r \sqsubseteq s,\ \Diamond^{*} A \sqsubseteq \exists r.B,\ \exists r.A \sqsubseteq B, \tag{2}$$

where $\Diamond^{*} \in \mathfrak{D}$, $A, A_1, A_2, B \in \mathbf{C} \cup \{\bot, \top\}$, and $r, s \in \mathbf{R}$.

Convex Names. When considering axioms of the form $\Diamond^{*} A \sqsubseteq A$ for $A \in \mathbf{C}$, we can first observe that the converse direction $A \sqsubseteq \Diamond^{*} A$, although syntactically not allowed, trivially holds in all interpretations. Moreover, the following implications between such equivalences follow from Lemma 2:



Since $\{A \equiv \Diamond^{+} A, A \equiv \Diamond^{-} A\}$ entails $A \equiv \Diamond^{\pm} A$, it thus makes sense to consider the *unique strongest* such axiom that is entailed by $\mathcal{K}$ (for a given A). We call A *rigid* if $A \equiv \Diamond^{\pm} A$ is the strongest such axiom, *shrinking* in case of $A \equiv \Diamond^{+} A$, *expanding* for $A \equiv \Diamond^{-} A$, and *(n-)convex* for $A \equiv \Diamond^{c}_{(n)} A$, i.e., whenever A is satisfied at two time points (with distance $< n$), then it is also satisfied at all time points in between. 1-convex concept names do not satisfy any special property, and are also called *flexible*. We use the same terms for role names.

2.2 A Completion Algorithm

We use the completion rules in Figure 1 to derive new axioms from $\mathcal{K}$. For simplicity, we treat $\top$ and $\bot$ like concept names, and thus allow assertions of the form $\top(a,i)$ and $\bot(a,i)$ here. It is clear that we cannot derive all possible entailments of the forms $\Diamond^{*} A \sqsubseteq B$ or $A(a,i)$, because (1) $\mathfrak{D}$ is infinite, and (2) $\mathbb{Z}$

$$\textsf{T1}\ \frac{}{\Diamond^{c}_1 A \sqsubseteq A} \qquad \textsf{T2}\ \frac{}{\Diamond^{\pm} A \sqsubseteq \top} \qquad \textsf{T3}\ \frac{}{\Diamond^{c}_1 r \sqsubseteq r} \qquad \textsf{T4}\ \frac{\Diamond^{\star} A_1 \sqsubseteq A_2 \quad \Diamond^{\bullet} A_2 \sqsubseteq A_3}{(\Diamond^{\bullet} \circ \Diamond^{\star}) A_1 \sqsubseteq A_3}$$

$$\textsf{T5}\ \frac{\Diamond^{\star} r_1 \sqsubseteq r_2 \quad \Diamond^{\bullet} r_2 \sqsubseteq r_3}{(\Diamond^{\bullet} \circ \Diamond^{\star}) r_1 \sqsubseteq r_3} \qquad \textsf{T6}\ \frac{\Diamond^{\star} A \sqsubseteq A_1 \quad \Diamond^{\bullet} A \sqsubseteq A_2 \quad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond^{\star} \cap \Diamond^{\bullet}) A \sqsubseteq B}$$

$$\textsf{T7}\ \frac{}{\exists r.\bot \sqsubseteq \bot} \qquad \textsf{T8}\ \frac{\Diamond^{\star} A \sqsubseteq \exists r.A_1 \quad \Diamond^{\bullet} r \sqsubseteq s \quad \Diamond^{\dagger} A_1 \sqsubseteq B_1 \quad \exists s.B_1 \sqsubseteq B}{\Diamond^{\star} A \sqsubseteq B}$$

$$\textsf{T8}'\ \frac{\Diamond^{\star} A \sqsubseteq \exists r.A_1 \quad \Diamond^{\bullet} r \sqsubseteq s \quad \Diamond^{\dagger} A_1 \sqsubseteq B_1 \quad \exists s.B_1 \sqsubseteq B \quad (\Diamond^{\bullet} \cap \Diamond^{\dagger}) \in \mathfrak{D}^{\pm}}{((\Diamond^{\bullet} \cap \Diamond^{\dagger}) \circ \Diamond^{\star}) A \sqsubseteq B}$$

$$\textsf{A1}\ \frac{}{\top(a,i)} \qquad \textsf{A2}\ \frac{i \in \Diamond^{\star} A(a) \quad \Diamond^{\star} A \sqsubseteq B}{B(a,i)} \qquad \textsf{A3}\ \frac{i \in \Diamond^{\star} r(a,b) \quad \Diamond^{\star} r \sqsubseteq s}{s(a,b,i)}$$

$$\textsf{A4}\ \frac{A_1(a,i) \quad A_2(a,i) \quad A_1 \sqcap A_2 \sqsubseteq B}{B(a,i)} \qquad \textsf{A5}\ \frac{r(a,b,i) \quad A(b,i) \quad \exists r.A \sqsubseteq B}{B(a,i)}$$

Fig. 1. Completion rules for $\mathcal{TELH}_{\bot}^{\Diamond^{\pm},\mathsf{lhs}}$ knowledge bases

is infinite. Moreover, there may be arbitrarily many time points between two assertions in $\mathcal{A}$ (exponentially many in the size of $\mathcal{A}$ if we assume time points to be encoded in binary). To deal with (1), we restrict the rule applications to the operators that occur in $\mathcal{K}$, in addition to $\Diamond^{c}$ and $\Diamond^{\pm}$, which are the only elements of $\mathfrak{D}$ that can be obtained via $\cup$, $\cap$, or $\circ$ from other $\Diamond$-operators, namely from $\Diamond^{+}$ and $\Diamond^{-}$. For (2), we consider the set of time points $\mathrm{tem}(\mathcal{A})$ (of linear size). Additionally, consider a maximal interval $[i,j]$ in $\mathbb{Z} \setminus \mathrm{tem}(\mathcal{A})$ (where i may be $-\infty$ and j may be ∞). To represent this interval, we choose a single representative time point $k \in [i,j]$, which is denoted by $|\ell| := k$ for all $\ell \in [i,j]$. For consistency, the representative $|i|$ for any $i \in \mathrm{tem}(\mathcal{A})$ is defined as i itself. Moreover, for any $k \in \mathbb{Z}$, we denote by $\lfloor k \rfloor := \max\{i \in \mathrm{tem}(\mathcal{A}) \mid i \leq k\}$ the maximal element of $\mathrm{tem}(\mathcal{A})$ below (or equal to) k, which we consider to be $-\infty$ in the case that there is no such element, and similarly define $\lceil k \rceil$. Note that $\lfloor i \rfloor = i = \lceil i \rceil$ whenever $i \in \mathrm{tem}(\mathcal{A})$, and otherwise $\lfloor i \rfloor < i < \lceil i \rceil$. By restricting all assertions to the finite set of representative time points

$$\mathrm{rep}(\mathcal{A}) := \{|i| \mid i \in \mathbb{Z}\} \supset \mathrm{tem}(\mathcal{A}),$$

we can encode infinitely many entailments in a finite set. We also define the following abbreviations, for all $A \in \mathbf{C}$, $r \in \mathbf{R}$, and $a, b \in \mathbf{I}$ ($\mathcal{K}$ refers to the KB after possibly already applying some completion rules):

$$A(a) := \{i \in \mathrm{rep}(\mathcal{A}) \mid A(a,i) \in \mathcal{K}\}$$
$$r(a,b) := \{i \in \mathrm{rep}(\mathcal{A}) \mid r(a,b,i) \in \mathcal{K}\}$$

Hence, we can write $\Diamond^{\star} A(a)$ in the completion rules to refer to the set of time points at which $\Diamond^{\star} A$ is inferred to be satisfied by a, given only the assertions in $\mathcal{A}$.

In the rules of Figure 1, we allow to instantiate A, B, A_1, A_2, A_3, B_1 by $\top$, $\bot$ or (normalized) $\mathcal{ELH}_\bot$ concepts from $\mathcal{K}$, r, s, r_1, r_2, r_3 by role names from $\mathcal{K}$, $\Diamond^{\star}, \Diamond^{\bullet}, \Diamond^{\dagger}$ by $\Diamond^{c}$, $\Diamond^{\pm}$ or elements of $\mathfrak{D}$ ocurring in $\mathcal{K}$, a, b by individual names from $\mathcal{K}$, and i by values from $\mathrm{rep}(\mathcal{A})$, such that the resulting axioms are in normal form. The side conditions $(\Diamond^{\bullet} \cap \Diamond^{\dagger}) \in \mathfrak{D}^{\pm}$, $i \in \Diamond^{\star}A(a)$, $i \in \Diamond^{\star}r(a,b)$ can be checked in polynomial time. All rules also apply to axioms without diamonds since we can treat A as $\Diamond^{c}_1 A$.

If $\mathcal{K}$ contains all axioms in the precondition of an instantiated rule, we consider the axiom in its conclusion. If it is a new assertion, we add it to $\mathcal{K}$. If it is a concept inclusion $\Diamond^{\star}A \sqsubseteq B$, we check whether $\mathcal{K}$ already contains a CI of the form $\Diamond^{\bullet}A \sqsubseteq B$. If not, then we simply add $\Diamond^{\star}A \sqsubseteq B$ to $\mathcal{K}$; otherwise, and if $\Diamond^{\star} \cup \Diamond^{\bullet} \neq \Diamond^{\bullet}$, we replace $\Diamond^{\bullet}A \sqsubseteq B$ by the new axiom $(\Diamond^{\star} \cup \Diamond^{\bullet})A \sqsubseteq B$, in order to reflect the validity of both axioms at once. RIs are handled in the same way. For example, if we know that $\Diamond^{+}A \sqsubseteq B$ holds, and have just inferred that $\Diamond^{-}A \sqsubseteq B$ holds as well, then $\Diamond^{\pm}A \sqsubseteq B$ is a valid entailment, because $\Diamond^{\pm} \subseteq \Diamond^{+} \cup \Diamond^{-}$, and thus whenever an element satisfies $\Diamond^{\pm}A$, it must satisfy either $\Diamond^{+}A$ or $\Diamond^{-}A$. In this way, for any two concepts A, B, the KB always contains at most one axiom $\Diamond^{\star}A \sqsubseteq B$, and similarly for roles.

Let $\mathcal{K}^*$ be the KB obtained by exhaustive application of the completion rules in Figure 1 to $\mathcal{K}$, where we assume (for technical reasons explained in the extended version) that A2 and A3 are always applied at the same time for all $i \in \Diamond^{\star}A(a)$ and $i \in \Diamond^{\star}r(a,b)$, respectively. This process terminates since we only produce axioms of the form $\Diamond^{\star}A \sqsubseteq B$, $\Diamond^{\star}r \sqsubseteq s$, $A(a,i)$, or $r(a,b,i)$, where $\Diamond^{\star}$ was already present in the initial $\mathcal{K}$ or it belongs to $\{\Diamond^{c}_1, \Diamond^{c}, \Diamond^{\pm}\}$, $i \in \mathrm{rep}(\mathcal{A})$, and A, B, r, s, a, b are from $\mathcal{K}$; there are only polynomially many such axioms.

To decide whether a concept assertion $D(a,i)$ follows from $\mathcal{K}$, we then simply look up whether $D(a,|i|)$ belongs to $\mathcal{K}^*$. For a concept inclusion $\Diamond^{\star}C \sqsubseteq D$ with $C, D \in \mathbf{C}$, we check whether $\mathcal{K}^*$ contains an inclusion of the form $\Diamond^{\bullet}C \sqsubseteq D$ with $\Diamond^{\star} \subseteq \Diamond^{\bullet}$, which can be done in polynomial time (see Lemma 2). One can also check entailment of role axioms in a similar way, but we omit them here for brevity.

Lemma 4. *$\mathcal{K}$ is inconsistent iff $\bot(a,i) \in \mathcal{K}^*$ for some $a \in \mathbf{I}$ and $i \in \mathrm{rep}(\mathcal{A})$.*

Let now $\mathcal{K}$ be consistent, C be a $\mathcal{TELH}_\bot^{\Diamond^{c},\mathsf{lhs}}$ concept, D be an $\mathcal{ELH}_\bot$ concept, and $\Diamond^{\star} \in \mathfrak{D}$. Then $\mathcal{K} \models \Diamond^{\star}C \sqsubseteq D$ iff either there is $\Diamond^{\dagger} \in \mathfrak{D}$ with $\Diamond^{\dagger}C \sqsubseteq \bot \in \mathcal{K}^*$, or there is $\Diamond^{\bullet} \sqsupseteq \Diamond^{\star}$ with $\Diamond^{\bullet}C \sqsubseteq D \in \mathcal{K}^*$. Moreover, $\mathcal{K} \models D(a,i)$ iff $D(a,|i|) \in \mathcal{K}^*$.

We obtain the following result, where the lower bound follows from propositional Horn logic [23].

Theorem 5. *Entailment in $\mathcal{TELH}_\bot^{\Diamond^{c},\mathsf{lhs}}$ is* P*-complete.*

Example 6. Consider *rheumatoid arthritis*, an autoimmune disorder that cannot be healed. In irregular intervals, it produces so-called *flare ups*, that cause pain in the joints. We formalize this knowledge as follows:

$$\mathrm{RheumatoidArthritisPatient} \equiv \exists \mathrm{diagnosedWith}.\mathrm{RheumatoidArthritis} \tag{3}$$

$$\mathrm{FlareUpPatient} \sqsubseteq \mathrm{RheumatoidArthritisPatient} \tag{4}$$

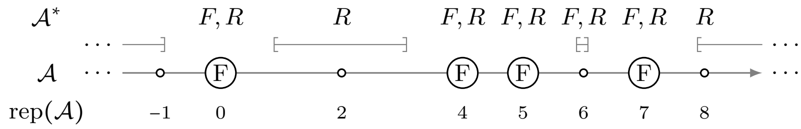
$$\Diamond^{-}\mathrm{RheumatoidArthritisPatient} \sqsubseteq \mathrm{RheumatoidArthritisPatient} \tag{5}$$

$$\Diamond^{c}_2\mathrm{FlareUpPatient} \sqsubseteq \mathrm{FlareUpPatient} \tag{6}$$

We make the assumption that a flare up is 2-month convex, hence if two flare ups are reported at most 2 months apart, we assume that they refer to the same flare up and hence the flare up also present in between the two reports. By applying Rule T4 from the completion algorithm to axioms (4) and (5), we can add

$$\Diamond\!\!\!\!\;|\,\text{FlareUpPatient} \sqsubseteq \text{RheumatoidArthritisPatient}$$

to the KB. Suppose the ABox consists of the assertions FlareUpPatient(p_1, i), $i \in \{0, 4, 5, 7\}$, for a patient p_1. The completed ABox, denoted by $\mathcal{A}^*$, is illustrated below, where for simplicity we omit the individual name p_1.



Here, RheumatoidArthritisPatient and FlareUpPatient are abbreviated by their first letters, respectively. Representatives $-1, 2, 6$ and 8 have been introduced and the intervals they represent are illustrated in gray.

3 Minimal-World Semantics for Metric Temporal Conjunctive Queries with Negation

We now consider the reasoning problem of query answering, which generalizes entailment of assertions. We develop a new temporal query language and follow an approach from [12] to find an appropriate closed-world semantics for negation.

Let $\mathbf{V}$ be a set of *variables*, and $\mathbf{T} \coloneqq \mathbf{I} \cup \mathbf{V}$ be the set of terms. An *atom* is either a *concept atom* of the form $A(\tau)$ or a *role atom* of the form $r(\tau, \rho)$, where $A \in \mathbf{C}$, $r \in \mathbf{R}$ and $\tau, \rho \subseteq \mathbf{T}$. A *conjunctive query (CQ)* $\phi(\mathbf{x})$ is a first-order formula of the form $\exists \mathbf{y}.\psi(\mathbf{x}, \mathbf{y})$, where ψ is a finite conjunction of atoms over the free variables $\mathbf{x}$ (also called the *answer variables*) and the quantified variables $\mathbf{y}$. *Conjunctive queries with (guarded) negation (NCQs)* are constructed by extending CQs with negated concept atoms $\neg A(\tau)$ and negated role atoms $\neg r(\tau, \rho)$ in such a way that, for any negated atom over terms τ (and ρ), the query contains at least one positive atom over τ (and ρ) containing all the variables of the negated atom. An NCQ is *rooted* if its variables are all connected via role atoms to an answer variable (from $\mathbf{x}$) or an individual name. An NCQ is *Boolean* if it does not have answer variables. To determine whether $\mathcal{I} \vDash \phi$ holds for an NCQ ϕ and an atemporal interpretation $\mathcal{I}$, we use standard first-order semantics.

We now extend the temporal CQs from [8] by metric operators [1, 7, 21] and negation.

Definition 7. Metric temporal conjunctive queries with negation (MTNCQs) *are built by the grammar rule*

$$\phi ::= \psi \mid \top \mid \bot \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \mathcal{U}_I \phi \mid \phi \mathcal{S}_I \phi, \tag{7}$$

where ψ is an NCQ, and I is an interval over $\mathbb{N}$. An MTNCQ ϕ is rooted/Boolean *if all NCQs in it are rooted/Boolean.*

ϕ	$\mathfrak{I}, i \models \phi$ iff
CQ ψ	$\mathcal{I}_i \models \psi$
$\top$	true
$\bot$	false
$\neg\phi$	$\mathfrak{I}, i \not\models \phi$
$\phi \land \psi$	$\mathfrak{I}, i \models \phi$ and $\mathfrak{I}, i \models \psi$
$\phi \lor \psi$	$\mathfrak{I}, i \models \phi$ or $\mathfrak{I}, i \models \psi$
$\phi \mathcal{U}_I \psi$	$\exists k \in I$ such that $\mathfrak{I}, i+k \models \psi$ and $\forall j : 0 \le j < k : \mathfrak{I}, i+j \models \phi$
$\phi \mathcal{S}_I \psi$	$\exists k \in I$ such that $\mathfrak{I}, i-k \models \psi$ and $\forall j : 0 \le j < k : \mathfrak{I}, i-j \models \phi$

Fig. 2. Semantics of (Boolean) MTNCQs for $\mathfrak{I} = (\Delta^{\mathfrak{I}}, (\mathcal{I}_i)_{i\in\mathbb{Z}})$ and $i \in \mathbb{Z}$.

We employ the standard semantics shown in Figure 2. One can define the *next* operator as $\bigcirc\phi := \top\mathcal{U}_{[1,1]}\phi$, and similarly $\bigcirc^-\phi := \top\mathcal{S}_{[1,1]}\phi$. We can also express $\Diamond_I\phi := (\top\mathcal{S}_{-(I\cap(-\infty,0])}\phi) \lor (\top\mathcal{U}_{I\cap[0,\infty)}\phi)$ and $\Box_I\phi := \neg\Diamond_I\neg\phi$, and hence, by (1), the $\circledDiamond_n$-operators from Section 2. An *MTCQ* (or *positive MTNCQ*) is an MTNCQ without negation, where we assume that the operator $\Box_I$ is nevertheless included as part of the syntax of MTCQs.

Example 8. Consider the criterion "Diagnosis of Rheumatoid Arthritis (RA) of more than 6 months and less than 15 years."[6] This can be expressed as an MTNCQ as follows:

$$\begin{aligned}\phi(x) := {} & \Box_{[-6,0]}\big(\exists y.\text{diagnosedWith}(x,y) \land \text{RheumatoidArthritis}(y)\big) \\ & \land \neg\Box_{[-180,0]}\big(\exists y.\text{diagnosedWith}(x,y) \land \text{RheumatoidArthritis}(y)\big)\end{aligned}$$

The semantics are defined model-theoretically as usual: Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a $\mathcal{TELH}_\bot^{\circledDiamond,\mathsf{lhs}}$ KB, $\phi(\mathbf{x})$ an MTNCQ, $\mathbf{a}$ a tuple of individual names from $\mathcal{A}$, $i \in \text{tem}(\mathcal{A})$, and $\mathfrak{I}$ a temporal interpretation. The pair $(\mathbf{a}, i)$ is an *answer* to $\phi(\mathbf{x})$ w.r.t. $\mathfrak{I}$ if $\mathfrak{I}, i \models \phi(\mathbf{a})$. The set of all answers for ϕ w.r.t. $\mathfrak{I}$ is denoted $\text{ans}(\phi, \mathfrak{I})$. The tuple $(\mathbf{a}, i)$ is a *certain answer* to ϕ w.r.t. $\mathcal{K}$ if it is an answer in every model of $\mathcal{K}$; all these tuples are collected in the set $\text{cert}(\phi, \mathcal{K})$.

Query answering is the decision problem of checking $(\mathbf{a}, i) \in \text{cert}(\phi, \mathcal{K})$ when given $\mathbf{a}$, i, ϕ, and $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. CQ answering over $\mathcal{ELH}_\bot$ KBs is NP-complete in general, and P-complete in *data complexity*, where the query ϕ and the TBox $\mathcal{T}$ are not considered as part of the input [24, 25, 29]. However, certain answer semantics for NCQ answering over $\mathcal{ELH}_\bot$ is coNP-hard [19]. To achieve tractable reasoning in data-oriented applications, we extend the *minimal-world semantics* from [12], which allows for NCQ answering in polynomial time, and gives intuitive semantics to negated query atoms.

3.1 Minimal-World Semantics for MTNCQs

Our goal is to extend the approach from [12] to find a *minimal canonical model* of a $\mathcal{TELH}_\bot^{\circledDiamond,\mathsf{lhs}}$ KB . Similarly to the *core* chase [17], the main idea is that this model

[6] `https://clinicaltrials.gov/ct2/show/NCT01198002`

should not contain redundant elements. Particularly, the minimum necessary number of anonymous objects together with the closed-world semantics adequately represents negative knowledge about the objects; for a detailed discussion, see [12]. We consider here the sublogic $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$ of $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs}}$ without temporal roles $\diamondsuit r$, because temporal roles interfere with the *minimality*: by propagating through time, a temporal role can easily violate the "local" minimality of interpretations at other time points, which could lead to unintuitive answers. In the definition of the model, we make use of entailment in $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$, which can be checked in polynomial time. Thus, we can exclude w.l.o.g. equivalent concept and role names. Also, for simplicity, in the following we assume w.l.o.g. that all CIs are in the following stronger normal form (cf. (2)):

$$\diamondsuit A \sqsubseteq B,\ A_1 \sqcap A_2 \sqsubseteq B,\ r \sqsubseteq s,\ A \sqsubseteq \exists r.B,\ \exists r.A \sqsubseteq B,$$

i.e., $\diamondsuit$-operators are allowed only in CIs of the form $\diamondsuit A \sqsubseteq B$. In particular, disallowing CIs of the form $\diamondsuit A \sqsubseteq \exists r.B$ allows us to draw a stronger connection to the original construction in [12]; see in particular Step 3(a) in Def. 9 below.

We need one more auxiliary definition from [12] to define the minimal temporal canonical model. Given a set V of existential restrictions, we say that $\exists r.A \in V$ is *minimal* in V if there is no other $\exists s.B \in V$ such that $\mathcal{K} \models s \sqsubseteq r$ and $\mathcal{K} \models B \sqsubseteq A$.

Definition 9. *The* minimal temporal canonical model $\mathfrak{I}_\mathcal{K} = (\Delta^{\mathfrak{I}_\mathcal{K}}, (\mathcal{I}_i)_{i\in\mathbb{Z}})$ *of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is obtained by the following steps.*

1. *Set $\Delta^{\mathfrak{I}_\mathcal{K}} := \mathbf{I}$ and $a^{\mathcal{I}_i} := a$ for all $a \in N_I$ and $i \in \mathbb{Z}$.*
2. *For each time point $i \in \mathbb{Z}$, define $A^{\mathcal{I}_i} := \{a \mid \mathcal{K} \models A(a,i)\}$ for all $A \in \mathbf{C}$ and $r^{\mathcal{I}_i} := \{(a,b) \mid \mathcal{K} \models r(a,b,i)\}$ for all $r \in \mathbf{R}$.*
3. *Repeat the following steps:*
   - *(a) Select an element $d \in \Delta^{\mathfrak{I}_\mathcal{K}}$ that has not been selected before and, for each $i \in \mathbb{Z}$, let $V_i := \{\exists r.B \mid d \in A^{\mathcal{I}_i},\ d \notin (\exists r.B)^{\mathcal{I}_i},\ \mathcal{K} \models A \sqsubseteq \exists r.B,\ A, B \in \mathbf{C}\}$.*
   - *(b) For each $\exists r.B$ that is minimal in some V_i, add a fresh element e_{rB} to $\Delta^{\mathfrak{I}_\mathcal{K}}$. For all $i \in \mathbb{Z}$ and $\mathcal{K} \models B \sqsubseteq A$, add e_{rB} to $A^{\mathcal{I}_i}$.*
   - *(c) For all $i \in \mathbb{Z}$, minimal $\exists r.B$ in V_i, and $\mathcal{K} \models r \sqsubseteq s$, add (d, e_{rB}) to $s^{\mathcal{I}_i}$.*

We denote by $\mathfrak{I}_\mathcal{A}$ the result of executing only Steps 1 and 2 of this definition, i.e., restricting $\mathfrak{I}_\mathcal{K}$ to the named individuals. Since there are only finitely many elements of $\mathbf{I}$, $\mathbf{C}$, and $\mathbf{R}$ that are relevant for this definition (i.e., those that occur in $\mathcal{K}$), for simplicity we often treat $\mathfrak{I}_\mathcal{A}$ as if it had a finite object (but still infinite time) domain.

In $\mathfrak{I}_\mathcal{K}$, there may exist anonymous objects that are not connected to any named individuals in $\mathcal{I}_i$ and are not relevant for the satisfaction of the KB. For this reason, in the following we consider only rooted MTNCQs, which can be evaluated only over the parts of $\mathfrak{I}_\mathcal{K}$ that are connected to the named individuals. We show that $\mathfrak{I}_\mathcal{K}$ is actually a model of $\mathcal{K}$ and is canonical in the usual sense that it can be used to answer *positive* queries over $\mathcal{K}$ under certain answer semantics.

Lemma 10. *Let $\mathcal{K}$ be a consistent $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$ KB. Then $\mathfrak{I}_\mathcal{K}$ is a model of $\mathcal{K}$ and, for every rooted MTCQ ϕ, we have* $\mathrm{cert}(\phi, \mathcal{K}) = \mathrm{ans}(\phi, \mathfrak{I}_\mathcal{K})$.

Thus, the following *minimal-world* semantics is compatible with certain answer semantics for positive (rooted) queries, while keeping a tractable data complexity.

Definition 11. *The set of* minimal-world answers *to an MTNCQ q over a consistent $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ KB $\mathcal{K}$ is* $\mathrm{mwa}(\phi,\mathcal{K}) \coloneqq \mathrm{ans}(\phi,\mathfrak{I}_{\mathcal{K}})$.

3.2 A Combined Rewriting for MTNCQs

Since the minimal canonical model $\mathfrak{I}_{\mathcal{K}}$ may still be infinite, we now show that rooted MTNCQ answering under minimal-world semantics is combined first-order rewritable [27], i.e., to compute $\mathrm{mwa}(\phi,\mathcal{K})$ we can equivalently evaluate a rewritten query over a finite interpretation (of polynomial size). Since the rewriting depends only on the query and the TBox, its size is irrelevant for data complexity, and it can be evaluated in polynomial time. We proceed in two steps.

1. We rewrite ϕ into a *metric first-order temporal logic (MFOTL)* formula $\phi_{\mathcal{T}}$, which combines first-order formulas via metric temporal operators; for details, see [10]. This query can be evaluated over $\mathfrak{I}_{\mathcal{A}}$ instead of $\mathfrak{I}_{\mathcal{K}}$. Hence, we reduce the infinite object domain to the finite set $\mathbf{I}(\mathcal{K})$.
2. We then further rewrite $\phi_{\mathcal{T}}$ into a three-sorted first-order formula (with explicit variables for time points), which is then evaluated over a restriction $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ of $\mathfrak{I}_{\mathcal{A}}$ that contains only finitely many time points (essentially those in $\mathrm{rep}(\mathcal{A})$, although we modify them slightly).

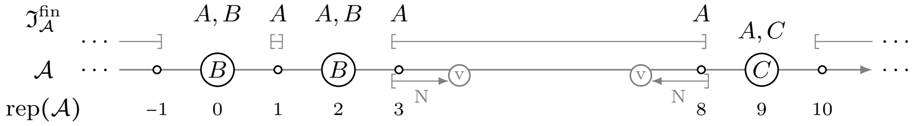
For the first step, we rewrite a rooted MTNCQ ϕ by replacing each (rooted) NCQ ψ with the first-order rewriting $\psi_{\mathcal{T}}$ from [12].[7] The result is a special kind of MFOTL formula $\phi_{\mathcal{T}}$ [10], in which atemporal first-order formulas can be nested inside MTL-operators, similarly as in MTNCQs. The semantics is based on a satisfaction relation $\mathfrak{I}, i \vDash \phi_{\mathcal{T}}$ that is defined in much the same way as in Fig. 2, the only exception being that $\mathfrak{I}, i \vDash \psi_{\mathcal{T}}$ for a first-order formula $\psi_{\mathcal{T}}$ is defined by $\mathcal{I}_i \vDash \psi_{\mathcal{T}}$, using the standard first-order semantics. We can lift the atemporal rewritability result from [12] in a straightforward way to our temporal setting.

Lemma 12. *Let $\mathcal{K} = (\mathcal{T},\mathcal{A})$ be a consistent $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ KB and ϕ be a rooted MTNCQ. Then* $\mathrm{mwa}(\phi,\mathcal{K}) = \mathrm{ans}(\phi_{\mathcal{T}},\mathfrak{I}_{\mathcal{A}})$.

For the second rewriting step, we restrict ourselves to finitely many time points. More precisely, we consider the finite structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$, which is obtained from $\mathfrak{I}_{\mathcal{A}}$ by restricting the set of time points to $\mathrm{rep}(\mathcal{A})$. By Lemma 4, the information contained in this structure is already sufficient to answer atomic queries. We extend this structure a little, by considering the *two* representatives i, j for each maximal interval $[i, j]$ in $\mathbb{Z} \setminus \mathrm{tem}(\mathcal{A})$. In this way, we ensure that the "border" elements are always representatives for their respective intervals. The size of the resulting structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ is polynomial in the size of $\mathcal{K}$.

[7] Strictly speaking, $\psi_{\mathcal{T}}$ in [12] is a *set* of first-order formulas, which is however equivalent to the disjunction of all these formulas.

Example 13. Let $\mathcal{A} = \{B(a,0), B(a,2), C(a,9)\}$ and $\mathcal{T} = \{\diamondsuit\diamondsuit_3 B \sqcap \diamondsuit C \sqsubseteq A\}$. Below one can see the finite structure $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$ over the representative time points $\{-1,0,1,2,3,8,9,10\}$, where for simplicity we omit the individual name.



The rewriting from Lemma 12 can refer to time instants outside of rep$(\mathcal{A})$. However, when we want to evaluate a pure FO formula over the finite structure $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$, this is not possible anymore, because the first-order quantifiers must quantify over the domain of $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$. Moreover, since the query $\phi_{\mathcal{T}}$ can contain metric temporal operators, we need to keep track of the distance between the time points in tem$(\mathcal{A})$. Hence, in the following we assume that $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$ is given as a first-order structure with the domain $\mathbf{I} \cup \{b_1, \dots, b_n\} \cup \text{rep}(\mathcal{A})$ and additional predicates `bit` and `sign` such that $\texttt{bit}(i,j)$, $1 \leq j \leq n$, is true iff the jth bit of the binary representation of the time stamp i is 1, and $\texttt{sign}(i)$ is true iff i is non-negative.

Thus, we now consider three-sorted first-order formulas with the three sorts $\mathbf{I}$ (for objects), $\{b_1, \dots, b_n\}$ (for bits) and rep$(\mathcal{A})$ (for time stamps). We denote variables of sort rep$(\mathcal{A})$ by t, t', t''. To simplify the presentation, we do not explicitly denote the sort of all variables, but this is always clear from the context. Every concept name is now accessed as a binary predicate of sort $\mathbf{I} \times \text{rep}(\mathcal{A})$, e.g., $A(a,i)$ refers to the fact that individual a satisfies A at time point i. Similarly, role names correspond to ternary predicates of sort $\mathbf{I} \times \mathbf{I} \times \text{rep}(\mathcal{A})$. It is clear that the expressions $t' \bowtie t$ and even $t' - t \bowtie m$ for some constant m and $\bowtie\, \in \{\geq, >, =, <, \leq\}$ are definable as first-order formulas using the natural order $<$ on $\{1, \dots, m\}$.

Lemma 14. *For $\phi_{\mathcal{T}}$ there is a constant $N \in \mathbb{N}$ such that, for every subformula ψ of $\phi_{\mathcal{T}}$, every maximal interval J in $\mathbb{Z} \setminus \bigcup\{[i-N, i+N] \mid i \in \text{tem}(\mathcal{A})\}$, all $k, \ell \in J$, and all relevant tuples $\mathbf{a}$ over $\mathbf{I}$, we have $\mathfrak{I}_{\mathcal{A}}, k \models \psi(\mathbf{a})$ iff $\mathfrak{I}_{\mathcal{A}}, \ell \models \psi(\mathbf{a})$.*

Hence, for evaluating subformulas of $\phi_{\mathcal{T}}$, it suffices to keep track of time points up to N steps away from the elements of rep$(\mathcal{A})$; this includes at least one element from each of the intervals J mentioned in Lemma 14, since every element of tem$(\mathcal{A})$ is immediately surrounded by two elements of rep$(\mathcal{A})$.

We exploit Lemma 14 in the following definition of the three-sorted first-order formula $[\psi]^n(\mathbf{x}, t)$ that simulates the behavior of $\psi(\mathbf{x})$ at the "virtual" time point $t+n$, where $n \in [-N, N]$. Whenever we use a formula $[\psi]^n(\mathbf{x}, t)$, we require that t denotes a representative for $t+n$. Due to our assumption that each maximal interval from $\mathbb{Z} \setminus \text{tem}(\mathcal{A})$ is represented by its endpoints (see Example 13), we know that t is a representative for $t+n$ iff there is no element of rep$(\mathcal{A})$ between t and $t+n$. We can encode this check in an auxiliary formula:

$$\texttt{rep}^n(t) \coloneqq \neg\exists t'.\, (t+n \leq t' < t) \vee (t < t' \leq t+n).$$

Example 15. In Example 13, 3 and 8 are representatives for the missing time points 4–7, and we have $\mathfrak{I}_{\mathcal{A}}^{\text{fin}} \models \texttt{rep}^1(3)$ (with $N = 1$). However, for $\phi_{\mathcal{T}} = \bigcirc\neg C(x)$, we have $\mathfrak{I}_\mathcal{A}, 3 \models \phi_\mathcal{T}(a)$, but $\mathfrak{I}_\mathcal{A}, 8 \not\models \phi_\mathcal{T}(a)$, i.e., the behavior at 3 and 8 differs. To distinguish this, we need to refer to the "virtual" time point 4 (gray circled "v") that is not included in $\mathfrak{I}_\mathcal{A}^{\text{fin}}$, via the formula $[\neg C(x)]^1(a, 3)$. By Lemma 14, it is sufficient to consider 4, because this determines the behavior at 5–7 .

We now define $[\psi]^n(\mathbf{x}, t)$ recursively, for each subformula ψ of $\phi_\mathcal{T}$. If ψ is a single rewritten NCQ, then $[\psi]^n(\mathbf{x}, t)$ is obtained by replacing each atemporal atom $A(x)$ by $A(x, t)$, and similarly for role atoms. The parameter n can be ignored here, because we assumed that t is a representative for $t + n$, and hence the time points t and $t + n$ are interpreted in $\mathfrak{I}_\mathcal{A}$ equally. For conjunctions, we set $[\psi_1 \wedge \psi_2]^n(\mathbf{x}, t) := [\psi_1]^n(\mathbf{x}, t) \wedge [\psi_2]^n(\mathbf{x}, t)$ and similarly for the other Boolean constructors. Finally, we demonstrate the translation for $\mathcal{U}$-formulas (the case of $\mathcal{S}$-formulas is analogous). We define $[\psi_1 \mathcal{U}_{[c_1,c_2]} \psi_2]^n(\mathbf{x}, t)$ as

$$\exists t'. \bigvee_{n' \in [-N,N]} \Big((t + n + c_1 \leq t' + n' \leq t + n + c_2) \wedge \mathtt{rep}^{n'}(t') \wedge [\psi_2]^{n'}(\mathbf{x}, t') \wedge$$
$$\forall t''. \bigwedge_{n'' \in [-N,N]} \Big(\big((t + n \leq t'' + n'' < t' + n') \wedge \mathtt{rep}^{n''}(t'') \big) \rightarrow [\psi_1]^{n''}(\mathbf{x}, t'') \Big) \Big),$$

where c_2 may be ∞, in which case the upper bound of $t + n + c_2$ can be removed.

Lemma 16. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$ KB and ϕ be an MTNCQ. Then* $\mathrm{ans}([\phi_\mathcal{T}]^0(\mathbf{x}, t), \mathfrak{I}_\mathcal{A}^{\text{fin}}) = \mathrm{ans}(\phi_\mathcal{T}, \mathfrak{I}_\mathcal{A})$.

This lemma allows us to compute in polynomial time that patient p_1 from Example 6 is an answer to $\phi(x)$ from Example 8 exactly at time point 7. Below we summarize our tight complexity results, which by Lemma 10 also hold for rooted MTCQs under certain answer semantics.

Theorem 17. *Answering rooted MTNCQs under minimal-world semantics over $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$ KBs is* ExpSpace*-complete, and* P*-complete in data complexity.*

Proof. ExpSpace-hardness is inherited from propositional MTL [1,18]. Moreover, first-order formulas over finite structures can be evaluated in PSpace [31]. Finally, the size of $[\phi_\mathcal{T}]^0(\mathbf{x}, t)$ is bounded exponentially in the size of ϕ and $\mathcal{T}$: each rewritten NCQ $\psi_\mathcal{T}$ may be exponentially larger than ψ, and each $[\psi_1 \mathcal{U}_I \psi_2]^n(\mathbf{x}, t)$ introduces exponentially many disjuncts and conjuncts (but the nesting depth of constructors in this formula is linear in the nesting depth of $\psi_1 \mathcal{U}_I \psi_2$).

For data complexity, hardness is inherited from atemporal $\mathcal{EL}$ [15]. Evaluating $\mathrm{FO}(<, \mathtt{bit})$-formulas is in DLogTime-uniform AC^0 in data complexity [26], and the size of our rewriting only depends on the query and the TBox. By Lemmas 12 and 16 and since $\mathfrak{I}_\mathcal{A}^{\text{fin}}$ is of size polynomial in the size of $\mathcal{A}$, deciding whether a tuple $\mathbf{a}$ is a minimal-world answer of an MTNCQ w.r.t. a $\mathcal{TELH}_\bot^{\diamondsuit,\mathsf{lhs},-}$ KB is possible in P. ◻

4 Related Work and Discussion

For a general overview of temporal ontology and query languages, see [3, 28]. In the presence of a single rigid role, allowing the operator $\Diamond_{\pm}$ on both sides of $\mathcal{EL}$ CIs makes subsumption undecidable [4]. In [20], a variety of restrictions are investigated to regain decidability. In particular, allowing the qualitative operators $\Diamond_{\pm}, \Diamond_{-}, \Diamond_{+}, \Diamond_{c}$ only on the left-hand side of CIs makes the logic tractable. Adding LTL operators to concepts was also investigated in other DLs, like $\mathcal{ALC}$ (without temporal roles) [28,32] and *DL-Lite* [4]. Only recently, also metric variants of such logics were considered [7,21,30]. There is a multitude of proposals for (non-metric) temporal query answering for lightweight DLs [2, 5, 8, 13, 14].

We extend previous results by introducing a tractable temporal extension of $\mathcal{ELH}_\bot$ that allows metric temporal operators, and a metric temporal query language. For MTNCQs under minimal-world semantics, we show that the complexity of query answering does not increase from the classical case. Future work includes representing numeric information, such as measurements and dosages of medications, which are important for evaluating eligibility criteria of clinical trials [11, 16] and extending the set $\mathfrak{D}$. It seems possible to allow other diamond operators in $\mathcal{TELH}_\bot^{\Diamond_{\pm},\mathsf{lhs}}$ axioms if they satisfy the relevant properties (see Lemmas 2 and 3). Currently, we are working on an optimized implementation of this method for temporal queries over large medical ontologies such as SNOMED CT.

References

1. Alur, R., Henzinger, T.A.: A really temporal logic. J. ACM **41**(1), 181–204 (1994)
2. Artale, A., Kontchakov, R., Kovtunova, A., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: First-order rewritability of ontology-mediated temporal queries. In: Proc. IJCAI. pp. 2706–2712. AAAI Press (2015)
3. Artale, A., Kontchakov, R., Kovtunova, A., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Ontology-mediated query answering over temporal data: A survey (invited talk). In: Proc. TIME. pp. 1:1–1:37. Schloss Dagstuhl (2017)
4. Artale, A., Kontchakov, R., Lutz, C., Wolter, F., Zakharyaschev, M.: Temporalising tractable description logics. In: Proc. TIME, pp. 11–22. IEEE Press (2007)
5. Artale, A., Kontchakov, R., Wolter, F., Zakharyaschev, M.: Temporal description logic for ontology-based data access. In: Proc. IJCAI. pp. 711–717. AAAI Press (2013)
6. Baader, F., Borgwardt, S., Forkel, W.: Patient selection for clinical trials using temporalized ontology-mediated query answering. In: Proc. HQA. pp. 1069–1074. ACM (2018)
7. Baader, F., Borgwardt, S., Koopmann, P., Ozaki, A., Thost, V.: Metric temporal description logics with interval-rigid names. In: Proc. FroCoS. pp. 60–76. Springer (2017)
8. Baader, F., Borgwardt, S., Lippmann, M.: Temporal query entailment in the description logic $\mathcal{SHQ}$. J. Web Sem. **33**, 71–93 (2015)
9. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. IJCAI. pp. 364–369. Professional Book Center (2005)
10. Basin, D., Klaedtke, F., Müller, S., Zălinescu, E.: Monitoring metric first-order temporal properties. J. ACM **62**(2), 15:1–15:45 (2015)

11. Bonomi, L., Jiang, X.: Patient ranking with temporally annotated data. J. Biomed. Inf. **78**, 43–53 (2018)
12. Borgwardt, S., Forkel, W.: Closed-world semantics for conjunctive queries with negation over $\mathcal{ELH}_{\bot}$ ontologies. In: Proc. JELIA. pp. 371–386. Springer (2019)
13. Borgwardt, S., Lippmann, M., Thost, V.: Temporalizing rewritable query languages over knowledge bases. J. Web Sem. **33**, 50–70 (2015)
14. Brandt, S., Kalaycı, E.G., Ryzhikov, V., Xiao, G., Zakharyaschev, M.: Querying log data with metric temporal logic. J. Artif. Intell. Res. **62**, 829–877 (2018)
15. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Data complexity of query answering in description logics. Art. Intell. **195**, 335–360 (2013)
16. Crowe, C.L., Tao, C.: Designing ontology-based patterns for the representation of the time-relevant eligibility criteria of clinical protocols. AMIA Jt. Summits Transl. Sci. Proc. **2015**, 173–177 (2015)
17. Deutsch, A., Nash, A., Remmel, J.B.: The chase revisited. In: Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2008, June 9-11, 2008, Vancouver, BC, Canada. pp. 149–158 (2008)
18. Furia, C.A., Spoletini, P.: Tomorrow and all our yesterdays: MTL satisfiability over the integers. In: Proc. ICTAC. pp. 126–140. Springer (2008)
19. Gutiérrez-Basulto, V., Ibáñez-García, Y., Kontchakov, R., Kostylev, E.V.: Queries with negation and inequalities over lightweight ontologies. J. Web Sem. **35**, 184–202 (2015)
20. Gutiérrez-Basulto, V., Jung, J.C., Kontchakov, R.: Temporalized $\mathcal{EL}$ ontologies for accessing temporal data: Complexity of atomic queries. In: Proc. IJCAI. pp. 1102–1108. AAAI Press (2016)
21. Gutiérrez-Basulto, V., Jung, J.C., Ozaki, A.: On metric temporal description logics. In: Proc. ECAI. pp. 837–845. IOS Press (2016)
22. Hripcsak, G., Zhou, L., Parsons, S., Das, A.K., Johnson, S.B.: Modeling electronic discharge summaries as a simple temporal constraint satisfaction problem. J. Am. Med. Inform. Assn. **12**(1), 55–63 (2005)
23. Jones, N.D., Laaser, W.T.: Complete problems for deterministic polynomial time. Theoretical Computer Science **3**(1), 105–117 (1976)
24. Krisnadhi, A., Lutz, C.: Data complexity in the $\mathcal{EL}$ family of description logics. In: Proc. LPAR. pp. 333–347. Springer (2007)
25. Krötzsch, M., Rudolph, S.: Conjunctive queries for $\mathcal{EL}$ with role composition. In: Proc. DL. pp. 355–362 (2007)
26. Lindell, S.: A purely logical characterization of circuit uniformity. In: Proc. of the 7th Annual Structure in Complexity Theory Conf. pp. 185–192 (1992)
27. Lutz, C., Toman, D., Wolter, F.: Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In: Proc. IJCAI. pp. 2070–2075. AAAI Press (2009)
28. Lutz, C., Wolter, F., Zakharyaschev, M.: Temporal description logics: A survey. In: Proc. of the 15th Int. Symp. on Temporal Representation and Reasoning (TIME'08). pp. 3–14. IEEE Press (2008)
29. Rosati, R.: On conjunctive query answering in $\mathcal{EL}$. In: Proc. DL. pp. 451–458 (2007)
30. Thost, V.: Metric temporal extensions of *DL-Lite* and interval-rigid names. In: Proc. KR. pp. 665–666. AAAI Press (2018)
31. Vardi, M.Y.: The complexity of relational query languages (extended abstract). In: Proc. STOC. pp. 137–146. ACM (1982)
32. Wolter, F., Zakharyaschev, M.: Temporalizing description logics. In: Frontiers of Combining Systems 2. pp. 379–402. Research Studies Press/Wiley (2000)