Translating Multi-Agent Modal Logics of Knowledge and Belief into Decidable First-Order Fragments

Qihui Feng*
RWTH Aachen University
Aachen, Germany
feng@kbsg.rwth-aachen.de

Hannah Wilk†
University of Amsterdam
Amsterdam, Netherlands
louise.wilk@student.uva.nl

Shakil M Khan
University of Regina
Regina, Canada
Shakil.Khan@uregina.ca

Gerhard Lakemeyer
RWTH Aachen University
Aachen, Germany
gerhard@kbsg.rwth-aachen.de

ABSTRACT

Translation-based modal theorem proving has been studied for decades. By reducing modal formulae to fragments of first-order logic, methods developed for first-order reasoning can be applied to modal inference problems. However, the existing translation approaches are insufficient for modal systems with specific frame properties, such as transitivity or Euclideanity, since they result in formulae not in a decidable first-order fragment.

With a revisit of the set-based possible-worlds semantics, we propose a new translation for multi-agent modal systems of knowledge and belief, such as $K(D)45_n$ and $S5_n$. We prove that the resulting formulae of the translation are in the two-variable guarded fragment. Therefore the decidability of the general satisfiability problem is preserved and it paves the way for translation-based reasoning in these modal systems. We also extend our approach to first-order modal logic and consider a decidable fragment.

KEYWORDS

knowledge representation and reasoning; modal theorem proving; multi-agent systems.

ACM Reference Format:

Qihui Feng, Hannah Wilk, Shakil M Khan, and Gerhard Lakemeyer. 2025. Translating Multi-Agent Modal Logics of Knowledge and Belief into Decidable First-Order Fragments. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

Modal logics, particularly epistemic modal logics, extend ordinary logical systems by introducing modalities for knowledge and belief. These enable the expression and reasoning about the subjective, epistemic attitude of agents. Over the past several decades, epistemic modal logics have attracted significant attention and have been widely applied in various domains, including protocol analysis in distributed systems [16], knowledge-based systems [24] and epistemic planning [4].

The conventional semantics of modal logic is defined via a set of possible-worlds and their relations. Such an approach, commonly known as Kripke semantics, was initially proposed by Hintikka and Kripke [18, 20]. It was later extended by numerous scholars (for a detailed overview, see [36]). One of the advantages of Kripke semantics is that it provides a graphical interpretation for many characteristics of agent reasoning. It has been shown that normal modal systems such as K45 for belief and S5 for knowledge can be precisely captured by Kripke frames satisfying certain relational properties [7]. Based on Kripke semantics, Ladner [21], Halpern and Moses [17] proved that the satisfiability problem for modal logics of knowledge and belief is decidable, yet of a high complexity especially when multiple agents are involved. Therefore, it is essential to explore modal inference approaches, which are practically efficient.

Traditionally, two paradigms in modal reasoning can be distinguished. The first is to develop proving techniques directly for modal logic, such as modal resolution [8, 28, 31] or tableaux calculi [6]. The second is translating the modal formulae into an ordinary system, such as first-order logic. With the second approach, efficient theorem provers and SAT-solvers developed for classical logic, such as VAMPIRE[19], Prover9 and MACE4[25], can be applied for modal reasoning.

The idea of translating modal logic into first-order logic was first introduced by Fine, van Benthem, and Morgan [11, 26, 35], known as the standard translation. As shown in [37], this approach can encode the simplest normal modal logic K into the first-order fragment GF^2, which is the intersection of the guarded fragment GF and the two-variables fragment FO^2 [12, 14]. Since GF^2 has an EXPTIME complexity upper bound, the translation for K is decidability-preserving.[1] However, for agent-based modal logics with introspection, i.e. accounts where agents can reflect on their own knowledge and lack of knowledge, the translation is out of the decidable fragment, mainly because frame properties such as transitivity or Euclideanity cannot be expressed in GF or FO^2.

To translate modal systems with introspection, many variants of the standard translation have been studied: Auffray and Enjalbert

*Contact author.
†Also affiliates with RWTH Aachen University

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

[1] In fact, it translates into an even more restricted fragment of GF^2 [32, 33].

proposed a translation into path logic, where a sub-domain of action is introduced to handle modal operators and their nesting [1]. Nonnengart combined relational translation with functional translation [27]. Schmidt and Hustadt associate (sub-)formulae with separate predicate symbols and propose an axiomatic translation [29, 30], Demri and de Nivelle studied the translation for regular grammar logic with converse[9], which can be used to embed some modal systems. While some of the mentioned approaches including [9, 30] can preserve decidability, these accounts mostly focus on single-agent languages. A clear formalism for multi-agent logics of knowledge and belief, such as $K45_n$, $S5_n$, cannot be found. Additionally, it is unclear how these methods can be extended to first-order modal logics. We notice that all of the existing translation approaches are based on simulation and axiomatization of Kripke structures, and the undecidability usually comes from expressing the frame conditions. It raises the question of whether axiomatizing Kripke models is the optimal choice to represent modal properties in a first-order language.

The rest of the paper is organized as follows: In Section 2, a set-based semantics for propositional modal logic is introduced. We analyze its properties and compare it with the ordinary Kripke semantics. Based on explicitly specifying the inclusion relation of the models in this semantics, we introduce a new translation approach in Section 3. We prove that this approach is decidability-preserving, and the translation has linear complexity with a small linear scaling factor. Furthermore, we extend the method to translating first-order modal formulae and consider a decidable first-order modal fragment.

2 THE LOGIC $\mathcal{ML}_n$

We introduce a dialect of modal logic called $\mathcal{ML}_n$. The language to be expressed is identical to the ordinary modal logic [7]. The semantics can be considered as a propositional fragment of the logic proposed in [2, 3].

2.1 Syntax

Let AP be the set of atomic propositions, which could potentially be countably infinite. The logic includes the usual Boolean connectives $\neg$ and $\wedge$. Others like $\vee, \rightarrow, \leftrightarrow$ are interpreted as abbreviations.

Let $\mathrm{Ag} = \{1, \ldots, n\}$ denote a finite set of agents. For each $i \in \mathrm{Ag}$, an epistemic operator $\boldsymbol{K}_i$ is used to denote knowledge or belief of agent i. The language of $\mathcal{ML}_n$ is defined as follows:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \boldsymbol{K}_i\varphi$$

Where $p \in AP$ and $i \in \mathrm{Ag}$. We say a formula is *objective* if it mentions no $\boldsymbol{K}$-operators and a formula is *subjective* if every proposition occurs in the scope of a $\boldsymbol{K}$-operator. A formula is *i-objective* if every occurrence of $\boldsymbol{K}_i$ is in the scope of a $\boldsymbol{K}_j$ where $j \neq i$, and a formula is *i-subjective* if every occurrence of propositions or $\boldsymbol{K}_j$ for $j \neq i$ is in the scope of a $\boldsymbol{K}_i$. In addition, we use $\textsc{True} := p \vee \neg p$ and $\textsc{False} := p \wedge \neg p$ to represent tautologies and falsity.

As Halpern pointed out in [15], the complexity of the modal logic is affected by the depth of nesting of modal operators. To evaluate the depth of a formula, we define:

Definition 1 (i-depth). *For $i \in \mathrm{Ag}$, the i-depth of a formula α, written as $dep[\alpha, i]$, is recursively defined as*

- $dep[p, i] = 0$ *for* $p \in AP$
- $dep[\neg\alpha, i] = dep[\alpha, i]$
- $dep[\alpha \wedge \beta, i] = max(dep[\alpha, i], dep[\beta, i])$
- $dep[\boldsymbol{K}_i\alpha, i] = \max\{\max\{dep[\alpha, j] \mid j \neq i\} + 1, dep[\alpha, i]\}$
- $dep[\boldsymbol{K}_j\alpha, i] = 0$ *for* $j \neq i$

Intuitively, $dep[\alpha, i]$ stands for the depth of agent i's knowledge in formula α. The depth of a formula α is defined as the maximal i-depth, written as $dep[\alpha] = \max_{i \in \mathrm{Ag}} dep[\alpha, i]$.

Example 1. *Let $\varphi := \neg p \wedge \neg\boldsymbol{K}_2(\neg p)$. Suppose that p means the coin is fair. Let $\alpha := \boldsymbol{K}_1\varphi$, then α means that agent 1 knows that the coin is not fair, and he knows that agent 2 doesn't know the unfairness.*

$$\begin{aligned} dep[\alpha, 1] &= \max\{dep[\varphi, 2] + 1, dep[\varphi, 1]\} \\ &= \max\{1 + 1, 0\} = 2 \\ dep[\alpha, 2] &= dep[\boldsymbol{K}_1\varphi, 2] = 0 \end{aligned}$$

2.2 Semantics

The logic $\mathcal{ML}_n$ has a set-based possible-worlds semantics, where introspection of agents have been embedded. It was first used by Levesque as the semantics for $K45$ [22]. In Levesque's design, a model consists of a distinguished state that describes what is "true" in the real world and a set of states that the agent thinks possible. Levesque's model was further developed as the semantics to formalize the notion of "only-knowing" [23, 24]. Belle and Lakemeyer extended the semantics to multi-agent cases by defining a nested set structure [2, 3] and we adopt similar structures in our semantics. By a *world* we mean a set of atomic propositions considered as true, i.e. $w \subseteq AP$. Let $\mathcal{W}$ be the set of all possible-worlds. For $k \geq 0$, we inductively define the epistemic states (also written as states for short) and structures as follows:

- $\mathcal{E}^0 = \{\{\varnothing\}\}$;
- $\mathcal{S}^{k+1} = \{(w, e_1, \ldots, e_{n-1}) \mid w \in \mathcal{W},\ e_j \in \mathcal{E}^k \text{ for all } j\}$
- $\mathcal{E}^{k+1} = \{e^{k+1} \mid e^{k+1} \subseteq \mathcal{S}^{k+1}\}$

We call $s^k \in \mathcal{S}^k$ a k-structure and $e^k \in \mathcal{E}^k$ a k-state. We call k the depth of s^k and e^k. Intuitively, a structure consists of what an agent considers as possible about the world and about other agents' beliefs. When the context is clear, we omit the superscript k. By a model, we mean a tuple $(w, e_1, \ldots e_n)$ (also written as $(w, \vec{e})$ for simplicity), where w is a possible-world and e_i is the epistemic state of agent i. We use $\mathbb{M}$ to denote the set of all models. Note the difference between a structure and a model: a structure consists of $n-1$ states of the same depth, while a model consists of n states whose depths can be different. For state $\hat{e}$ and structure $s = (w, e_1, \ldots, e_{n-1})$, we use $s \cup_i \hat{e}$ to denote the model obtained by inserting $\hat{e}$ at the i-th position, i.e. $(w, e_1, \ldots e_{i-1}, \hat{e}, e_i, \ldots e_{n-1})$. The satisfaction relation is defined as follows:

- $(w, \vec{e}) \models p$ iff $p \in w$;
- $(w, \vec{e}) \models \neg\alpha$ iff it is not the case that $(w, \vec{e}) \models \alpha$;
- $(w, \vec{e}) \models \alpha \wedge \beta$ iff $(w, \vec{e}) \models \alpha$ and $(w, \vec{e}) \models \beta$;
- $(w, \vec{e}) \models \boldsymbol{K}_i\varphi$ iff for all $s \in e_i$, $s \cup_i e_i \models \varphi$.

We say a model and a formula α is *compatible* if, for any $i \in \mathrm{Ag}$, the i-depth of α is not deeper than the epistemic state of agent i. Given a finite set $\Sigma \subseteq \mathcal{ML}_n$,[2] $\alpha \in \mathcal{ML}_n$, we say Σ entails α (written as $\Sigma \models \alpha$) iff for any model $(w, \vec{e})$ compatible with each formula in

[2] Σ can also be an infinite set, provided that the depth of formulae is bounded, i.e. it exists $k \in \mathbb{N}$ s.t. $dep[\varphi] \leq k$ for all $\varphi \in \Sigma$.

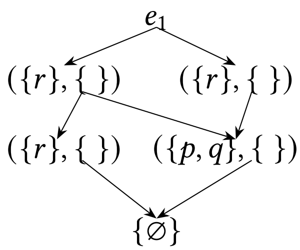

Figure 1: state e_1 in Example 2, arrows stand for inclusion.

$\Sigma \cup \{\alpha\}$, if $(w,\vec{e}) \models \varphi$ for all $\varphi \in \Sigma$, then $(w,\vec{e}) \models \alpha$. We say α is *valid* (written as $\models \alpha$) iff $\{\} \models \alpha$. We say a formula α is *satisfiable* if there is a compatible model $(w,\vec{e})$ s.t. $(w,\vec{e}) \models \alpha$.
When α is objective, we write $w \models \alpha$ instead of $(w,\vec{e}) \models \alpha$. When α is i-subjective, we write $e_i \models \alpha$. Note that this account cannot handle unbounded depth of formulae. When Σ is an infinite set of formulae with unbounded depth, i.e. for any k it exists $\psi \in \Sigma$ s.t. $dep[\psi] > k$, then $\Sigma \models \alpha$ is not well-defined.

Example 2. *Suppose that* $\mathrm{Ag} = \{1,2\}$*, let* $p,q,r \in AP$*,* $w_0 = \{r\}$*,* $w_1 = \{p,q\}$*. We define two states of depth* 1 *for agent 2:*

$$e_2 = \{(w_0,\{\varnothing\}),(w_1,\{\varnothing\})\} \qquad e_2' = \{(w_1,\{\varnothing\})\}$$

and a state of depth 2 for agent 1: $e_1 = \{(w_0,e_2),(w_0,e_2')\}$*. Consider the formula in Example 1, we have* $e_1 \models \boldsymbol{K}_1(\neg p \wedge \neg \boldsymbol{K}_2(\neg p))$*. A diagram of* e_1 *is given in the Figure 1.*

A model in $\mathcal{ML}_n$ cannot assign truth values to all formulae, mainly because the depth of the model is bounded. For instance, in the above example the state $e_1' = \{(w_0,\{\varnothing\})\}$ can not determine the truth of formula $\boldsymbol{K}_1(\neg p \wedge \neg \boldsymbol{K}_2(\neg p))$. We argue that this is not a problem because for any formula α, any model $(w,\vec{e})$ can be extended to a deeper one that is compatible with α while preserving the truth value of all compatible formulae. To understand the relation among models of different depths, we introduce the notion of regression:

Definition 2 (Regression). *The regression of structures and epistemic states is inductively defined as follows:*

- *For* $s \in \mathcal{S}^1$*, i.e.* $s = (w,\{\varnothing\},\ldots,\{\varnothing\})$*, the regression of* s*, written as* $s\downarrow$*, is* $\varnothing$*;*
- *For* $e \in \mathcal{E}^{k+1}$*,* $e' \in \mathcal{E}^k$*, we say* e' *is the regression of* e*, written as* $e' = e\downarrow$*, if and only if* $e' = \{s\downarrow \mid s \in e\}$*;*
- *For* $s \in \mathcal{S}^{k+1}$*,* $s' \in \mathcal{S}^k$*,* s' *is the regression of* s*, written as* $s' = s\downarrow$*, if* $s = (w,\vec{e})$ *and* $s' = (w,\vec{e'})$ *where* $e_j' = e_j\downarrow$*.*

As shown in the following lemma, regression preserves the truth values of compatible formulae.

Lemma 1. *For* $i \in \mathrm{Ag}$*,* $\vec{e}$ *and* $\vec{e'}$ *s.t.* $e_i' = e_i\downarrow$ *and* $e_j = e_j'$ *for any* $j \neq i$*. For any* $w \in \mathcal{W}$ *and* α *compatible with* $\vec{e'}$ *and* $\vec{e}$*,*

$$(w,\vec{e}) \models \alpha \text{ iff } (w,\vec{e'}) \models \alpha$$

A proof is given in Lem. 8 of [3].

2.3 Comparison with Kripke Semantics

We briefly recap Kripke semantics to compare with ours. See [16, 17, 21] for more details on Kripke semantics.

Definition 3 (Kripke structure). *A Kripke structure for a set of agents* Ag *is a tuple* $M = (\Omega, R_1,\ldots,R_n,\pi)$*, where* Ω *is a set of Kripke worlds.*[3] *For* $i \in \mathrm{Ag}$*,* R_i *is a binary relation on the worlds in* Ω*.* π *assigns each* $\omega \in \Omega$ *a set of propositions considered to be true under* ω*, i.e.* $\pi(\omega) \subseteq AP$*.*

Given a Kripke model M and a world $\omega \in \Omega$,

- $(M,\omega) \models p$ iff $p \in \pi(\omega)$;
- $(M,\omega) \models \neg\alpha$ iff it is not the case that $(M,\omega) \models \alpha$;
- $(M,\omega) \models \alpha \wedge \beta$ iff $(M,\omega) \models \alpha$ and $(M,\omega) \models \beta$;
- $(M,\omega) \models \boldsymbol{K}_i\alpha$ iff $(M,\omega') \models \alpha$ for all ω' s.t. $\omega R_i \omega'$.

The relation between frame conditions and the axiom systems of different modal logics are investigated in [7, 20], for example. We only introduce some conditions related to our account:

- *Transitivity*: If $\omega R_i \omega'$ and $\omega' R_i \omega''$, then $\omega R_i \omega''$.
- *Euclideanity*: If $\omega R_i \omega'$ and $\omega R_i \omega''$, then $\omega' R_i \omega''$.

Theorem 1 (Halpern [17]). *A Kripke structure* M *is a model of* $K45_n$ *iff all relations* R_i *are Euclidean and transitive.*

Intuitively, $K45_n$ axiom system requires both *positive* and *negative introspection*. Namely, for any α and agent $i \in \mathrm{Ag}$,

- $\models \boldsymbol{K}_i\alpha \to \boldsymbol{K}_i\boldsymbol{K}_i\alpha$
- $\models \neg\boldsymbol{K}_i\alpha \to \boldsymbol{K}_i\neg\boldsymbol{K}_i\alpha$

Given a $\mathcal{ML}_n$ model $(w,\vec{e})$, we show that it amounts to a tree-like exploration of a Kripke model up to a finite depth. Therefore it does not utilize the global transitivity or Euclideanity of all worlds. This is part of the reason why the translation can avoid axioms on frame conditions. For Kripke model $M = (\Omega, R_1,\ldots,R_n,\pi)$, we define a set of mappings from Ω to structures and models in $\mathcal{ML}_n$:

Definition 4. *Given* $M = (\Omega, R_1,\ldots,R_n,\pi)$*,* $\sigma \in \mathrm{Ag}^*$*,* $\omega \in \Omega$*,*

- *For* $|\sigma| = k$*,* $\phi_{M,k}^{(\sigma\cdot i)}(\omega) = \varnothing$*;*
- *For* $|\sigma| < k$*,* $\phi_{M,k}^{(\sigma\cdot i)}(\omega) = (w, e_1,\ldots,e_{i-1},e_{i+1},\ldots,e_n)$*, where* $w = \pi(\omega)$*,* $e_j = \{\phi_{M,k}^{(\sigma\cdot i\cdot j)}(\omega') \mid$ *for any* ω' *s.t.* $\omega R_j \omega'\}$*;*
- $\Phi_{M,k}(\omega) = (\pi(\omega),\vec{e})$*, where* $e_i = \{\phi_{M,k}^{(i)}(\omega') \mid \omega R_i \omega'\}$*.*

When the context is clear, we fix the Kripke model and omit M for short. Here σ is a finite word which uses Ag as the alphabet.[4] We use $start(\sigma)$ and $end(\sigma)$ to denote the first and the last letter in σ. The length of σ is denoted as $|\sigma|$. For any σ, e.g. $\sigma = 1\cdot 2$, the mapping $\phi_k^{(\sigma)}(\omega)$ returns a structure, which agent 1 believes that agent 2 considers as possible. Clearly, the mappings are defined recursively from longer words to their prefix.

Lemma 2. *For any* $\omega \in \Omega$*,* $k \in \mathbb{N}$ *and* $\sigma \in \mathrm{Ag}^+$*,*

- *For any* $j \neq start(\sigma)$*,* $\phi_k^{(\sigma)}(\omega) = \phi_{k+1}^{(j\cdot\sigma)}(\omega)$
- $\phi_{k+1}^{(\sigma)}(\omega)\downarrow = \phi_k^{(\sigma)}(\omega)$

[3] Note the difference between Kripke's worlds and Levesque's. To avoid confusion, we use the Latin letter w for Levesque's world and the Greek letter ω for Kripke's.
[4] We only consider σ, where the adjacent letters are distinct i.e. $\sigma \notin (\mathrm{Ag}^*\cdot i\cdot i\cdot \mathrm{Ag}^*)$ for any $i \in \mathrm{Ag}$.

Proof. When $k - |\sigma| = 0$, the proof is trivial. Suppose that the lemma holds for any k' and $|\sigma'|$ s.t. $k' - |\sigma'| < m$. For k, σ s.t. $k - |\sigma| = m$, let $\phi_k^{(\sigma)}(\omega) = (w, e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)$, then for all $l \neq end(\sigma)$, by the induction hypothesis we have

$$e_l = \{\phi_k^{(\sigma \cdot l)}(\omega') \mid \omega R_l \omega'\} = \{\phi_{k+1}^{(j \cdot \sigma \cdot l)}(\omega') \mid \omega R_l \omega'\}$$
$$e_l = \{\phi_k^{(\sigma \cdot l)}(\omega') \mid \omega R_l \omega'\} = \{\phi_{k+1}^{(\sigma \cdot l)}(\omega') \downarrow \mid \omega R_l \omega'\}$$

Thus $\phi_k^{(\sigma)}(\omega) = \phi_{k+1}^{(j \cdot \sigma)}(\omega)$ and $\phi_{k+1}^{(\sigma)}(\omega) \downarrow = \phi_k^{(\sigma)}(\omega)$. □

With the above mappings, we show the bi-simulation between Kripke models and $\mathcal{ML}_n$ models up to any finite depth.

Lemma 3. *For any $K45_n$ Kripke model M, $\omega \in \Omega$ and $k \in \mathbb{N}^+$, let $\Phi_k(\omega) = (w, \vec{e})$. For any formula α s.t. $dep[\alpha] \leq k$,*

$$(M, \omega) \models \alpha \text{ iff } (w, \vec{e}) \models \alpha \tag{1}$$

Proof. We prove it via induction on the structure of α.

- Base case: Given $\alpha := p$ for $p \in AP$, then for any k

$$\begin{aligned} (M, \omega) \models p &\Longleftrightarrow p \in \pi(\omega) && \text{(Kripke semantics)} \\ &\Longleftrightarrow p \in w \text{ for } (w, \vec{e}) = \Phi_k(\omega) && \text{(Def. } \Phi_k) \\ &\Longleftrightarrow w, \vec{e} \models p && (\mathcal{ML}_n \text{ semantics}) \end{aligned}$$

- Induction on $\wedge$ and $\neg$ is straightforward.
- Induction on $\boldsymbol{K}_i$: suppose that Eq. 1 holds for all ω and k. Let $(w, \vec{e}) = \Phi_{k+1}(\omega)$,

$$\begin{aligned} & (M, \omega) \models \boldsymbol{K}_i \alpha \\ \Longleftrightarrow & (M, \omega') \models \alpha \text{ for all } \omega' \text{ s.t. } \omega R_i \omega' && \text{(Kripke semantics)} \\ \Longleftrightarrow & w', \vec{e'} \models \alpha \text{ for all } \omega' \text{ s.t. } \omega R_i \omega', \Phi_k(\omega') = (w', \vec{e'}) && \text{(I.H.)} \\ \Longleftrightarrow & \text{ for all } s \in e_i, s \cup_i e_i' \models \alpha && (\#) \\ \Longleftrightarrow & \text{ for all } s \in e_i, s \cup_i e_i \models \alpha && (e_i' = e_i \downarrow, \text{Lem.1}) \\ \Longleftrightarrow & (w, \vec{e}) \models \boldsymbol{K}_i \alpha && (\mathcal{ML}_n \text{ semantics}) \end{aligned}$$

To explain step (#): consider $(w', \vec{e'}) = \Phi_k(\omega')$ of a fixed ω' s.t. $\omega R_i \omega'$. By definition, $e_i' = \{\phi_k^{(i)}(\omega'') \mid \omega' R_i \omega''\}$. Since R_i is transitive and Euclidean, for any $\tilde{\omega}$, $\phi_{k+1}^{(i)}(\tilde{\omega}) \in e_i$ iff $\phi_k^{(i)}(\tilde{\omega}) \in e_i'$. Thus e_i' is the regression of e_i, i.e. $e_i \downarrow = e_i'$.
For $j \neq i$, by Lem. 2 we have $e_j' = \{\phi_{k+1}^{(i \cdot j)}(\omega'') \mid \omega' R_j \omega''\}$. Thus $(w', e_1', \dots, e_{i-1}', e_{i+1}', \dots e_n') = \phi_{k+1}^{(i)}(\omega')$ and by definition of e_i, we have $\phi_{k+1}^{(i)}(\omega') \in e_i$. Conversely, for each $s \in e_i$, $s = \phi_{k+1}^{(i))}(\omega')$ with some ω' s.t. $\omega R_i \omega'$. By Lem.2, $s \cup_i e_i' = \Phi_k(\omega')$. □

2.4 Properties of the Logic

$\mathcal{ML}_n$ precisely captures the $K45_n$ properties in the sense that $K45_n$ modal system forms a sound and complete axiomatization for $\mathcal{ML}_n$. The soundness is implied by the following theorem:

Theorem 2. *For any $\alpha, \beta \in \mathcal{ML}_n$ and $i \in$ Ag,*

- *(Prop) If α is a propositional tautology, then $\models \alpha$*
- *(MP) If $\models \alpha$ and $\models \alpha \to \beta$ then $\models \beta$*
- *(Dist) $\models (\boldsymbol{K}_i \alpha \wedge \boldsymbol{K}_i(\alpha \to \beta)) \to \boldsymbol{K}_i \beta$*
- *(Nec) If $\models \alpha$ then $\models \boldsymbol{K}_i \alpha$*
- *(4) $\models \boldsymbol{K}_i \alpha \to \boldsymbol{K}_i \boldsymbol{K}_i \alpha$*
- *(5) $\models \neg \boldsymbol{K}_i \alpha \to \boldsymbol{K}_i \neg \boldsymbol{K}_i \alpha$*

Proving the completeness is equivalent to proving that every $K45_n$-consistent formula is $\mathcal{ML}_n$-satisfiable. By [17] every $K45_n$-consistent formula α is satisfiable in some Kripke structure with transitive, Euclidean relations. Then by Lem. 3 we prove that α is $\mathcal{ML}_n$-satisfiable. With the soundness and completeness, we have the following result:

Theorem 3. *α is $K45_n$-consistent iff α is $\mathcal{ML}_n$-satisfiable .*

2.5 Adaptation for Other Modal Systems

In Kripke's semantics, the consistency of belief and the truthfulness of knowledge can be interpreted as the seriality and the reflexivity of the accessible relations. We show that in our set-based semantics, they amount to considering a subset of $\mathcal{ML}_n$ models.
The consistency of beliefs, also known as the D-property, requires that for any $i \in$ Ag and formula α in the language,

$$\models \boldsymbol{K}_i \alpha \to \neg \boldsymbol{K}_i \neg \alpha$$

It can be achieved by ruling out models with empty (sub-)structures.

Definition 5 (D-state). *The set of D-states is defined as follows:*

- *If $e \in \mathcal{E}^1$, then e is a D-state iff $e \neq \{\}$.*
- *If $e \in \mathcal{E}^k$, where $k > 1$, then e is a D-state iff $e \neq \{\}$ and for all $(w', e_1', \dots, e_{i-1}', e_{i+1}', \dots, e_{n-1}') \in e$ and $j \neq i$, e_j' is a D-state.*

We call $(w, \vec{e})$ a D-model if for all $i \in$ Ag, e_i is a D-state. We say a formula α is $\mathcal{ML}_n^D$-satisfiable iff α is satisfied in some D-models.

The truthfulness of knowledge requires that for $i \in$ Ag and α,

$$\models \boldsymbol{K}_i \alpha \to \alpha$$

The truthfulness of knowledge requires that the actual world is always considered possible and the actual epistemic state of an agent is always considered possible by any other agents. We say a model $(w, \vec{e})$ is *homogeneous* if all e_i are of the same depth, i.e. there is a k s.t. $e_i \in \mathcal{E}^k$ for all $i \in$ Ag.

Definition 6. *A T-model is a homogeneous model $(w, \vec{e})$ s.t.*

- *$e_i = \{\varnothing\}$ for all $i \in$ Ag, or*
- *F.a. $i \in$ Ag, $s \in e_i$, $s \cup_i (e_i \downarrow)$ is a T-model and $(w, e_1 \downarrow, \dots, e_{i-1} \downarrow, e_{i+1} \downarrow, \dots, e_n \downarrow) \in e_i$*

We say α is $\mathcal{ML}_n^T$-satisfiable iff α is satisfied in some T-models.

Theorem 4. *For any $\alpha \in \mathcal{ML}_n$,*

- *α is $KD45_n$-consistent iff α is $\mathcal{ML}_n^D$-satisfiable;*
- *α is $S5_n$-consistent iff α is $\mathcal{ML}_n^T$-satisfiable.*

The proof is analogous to Thm. 3. Soundness is proved by the validity of axioms (similar to Thm. 2) and completeness is proved by constructing a D-model (T-model) given a Kripke frame and the formula to be satisfied (similar to Lem. 3).

3 TRANSLATION

In this section, we propose a new approach to translate modal formulae into first-order ones. The translation is mainly based on explicitly specifying the inclusion relation of $\mathcal{ML}_n$ structures. The form is similar to the standard translation, yet some essential differences exist: Binary predicates of the form E_σ are used to represent the inclusion relation instead of the accessibility of Kripke states.

For $\sigma \in \mathrm{Ag}^*$ s.t. $end(\sigma) = i$, $E_\sigma(u, v)$ means that the epistemic state at the i-th position of structure v contains structure u. We prove that such a translation preserves satisfiability, without explicitly expressing the frame conditions as the standard translation does. For any $\mathcal{ML}_n$ formula, the translation results in a first-order formula in the GF^2 fragment.

3.1 Translation Function

We define a recursive translation function $\mathcal{R}[\alpha, \sigma, u, v]$. Here α is the $\mathcal{ML}_n$ formula to be translated, $\sigma \in \mathrm{Ag}^*$ is a finite word, u, v are variables used for the translation.

Definition 7 (translation function). *Given $\sigma \in \mathrm{Ag}^*$,*

(1) *For $p \in AP$, $\mathcal{R}[p, \sigma, u, v] := P(u)$*
(2) $\mathcal{R}[\neg\alpha, \sigma, u, v] := \neg\mathcal{R}[\alpha, \sigma, u, v]$
(3) *For $\odot \in \{\wedge, \vee, \to, \leftrightarrow\}$,*
$\mathcal{R}[\alpha_1 \odot \alpha_2, \sigma, u, v] := \mathcal{R}[\alpha_1, \sigma, u, v] \odot \mathcal{R}[\alpha_2, \sigma, u, v]$
(4) *If $end(\sigma) = i$,*
$\mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] := \forall u\ E_\sigma(u, v) \to (\mathcal{R}[\alpha, \sigma, u, v])$
(5) *If $end(\sigma) \neq i$,*
$\mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] := \forall v\ E_{\sigma\cdot i}(v, u) \to (\mathcal{R}[\alpha, \sigma\cdot i, v, u])$

Where P is a unary predicate associated with proposition p.

Example 3. *Let $\alpha := \boldsymbol{K}_1(\neg p \wedge \neg\boldsymbol{K}_2(\neg p))$ The translation of α is*

$$\begin{aligned}&\mathcal{R}[\alpha, \epsilon, u, v]\\ =&\mathcal{R}[\boldsymbol{K}_1(\neg p \wedge \neg\boldsymbol{K}_2(\neg p)), \epsilon, u, v]\\ =&\forall v\ E_1(v, u) \to \mathcal{R}[\neg p \wedge \neg\boldsymbol{K}_2(\neg p), 1, v, u]\\ =&\forall v\ E_1(v, u) \to (\neg P(v) \wedge \neg\forall u(E_{1\cdot 2}(u, v) \to \mathcal{R}[\neg p, 1\cdot 2, u, v]))\\ =&\forall v\ E_1(v, u) \to (\neg P(v) \wedge \neg\forall u(E_{1\cdot 2}(u, v) \to \neg P(u))\end{aligned}$$

For $KD45_n$ system, we define function $\mathcal{R}_D$ which is identical to $\mathcal{R}$ except for items (4) and (5):

(4') If $end(\sigma) = i$, $\mathcal{R}_D[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$(\forall u\ E_\sigma(u, v) \to (\mathcal{R}_D[\alpha, \sigma, u, v])) \wedge \exists u\ E_\sigma(u, v)$
(5') If $end(\sigma) \neq i$, $\mathcal{R}_D[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$(\forall v\ E_{\sigma\cdot i}(v, u) \to (\mathcal{R}_D[\alpha, \sigma\cdot i, v, u])) \wedge \exists v\ E_{\sigma\cdot i}(v, u)$

For $S5_n$, we define $\mathcal{R}_T$ which also differs only on (4) and (5):

(4") If $end(\sigma) = i$, $\mathcal{R}_T[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$(\forall u\ E_\sigma(u, v) \to (\mathcal{R}_D[\alpha, \sigma, u, v])) \wedge E_\sigma(v, v)$
(5") If $end(\sigma) \neq i$, $\mathcal{R}_D[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$(\forall v\ E_{\sigma\cdot i}(v, u) \to (\mathcal{R}_D[\alpha, \sigma\cdot i, v, u])) \wedge E_{\sigma\cdot i}(u, u)$

3.2 Construction of Canonical FOL Model

To prove that $\mathcal{ML}_n$ can be embedded into FOL, we construct a canonical FOL model as follows:

Definition 8. *We define $M^c = \langle \mathcal{D}^c, I\rangle$ as follows:*

C1 *Domain of discourse $\mathcal{D}^c$ is the set of $\mathcal{ML}_n$ models $\mathbb{M}$.*
C2 *For $P \in AP$ and $(w, \vec{e}) \in \mathcal{D}^c$, $(w, \vec{e}) \in P^{M^c}$ iff $P \in w$*
C3 *For $\omega, \omega' \in \mathcal{D}^c$, $\sigma \in \mathrm{Ag}^+$ s.t. $end(\sigma) = i$. $(\omega', \omega) \in E_\sigma^{M^c}$ iff $\omega = (w, \vec{e})$ and $\omega' = s \cup_i e_i$ for some $s \in e_i$*

Since C1-C3 specifies different aspects of the model, the existence of the canonical model is obvious. For each satisfiable formula in $\mathcal{ML}_n$, we show the satisfiability of the resulting formula.

Theorem 5. *For any compatible model $(w, \vec{e})$ and sentence α,*

(1) *$w, \vec{e} \models \alpha$ iff $M^c, \mu \models_{FO} \mathcal{R}[\alpha, \epsilon, u, v]$ for $\mu(u) = (w, \vec{e})$.*
(2) *If $(w', \vec{e'})$ exists s.t. $(w, \vec{e}) = s \cup_i e_i'$ for some $s \in e_i$, then*

$$w, \vec{e} \models \alpha \text{ iff } M^c, \mu \models_{FO} \mathcal{R}[\alpha, \sigma\cdot i, u, v]$$

Where $\mu(u) = (w, \vec{e})$ and $\mu(v) = (w', \vec{e'})$.

Proof. We prove this via induction on the formula structure. As for the base case, let $p \in AP$, $\sigma \in \mathrm{Ag}^*$

$$\begin{aligned}w, \vec{e} \models p \Longleftrightarrow & p \in w && (\mathcal{ML}_n \text{ semantics})\\ \Longleftrightarrow & \mu(u) \in P^{M^c} && (\text{C2})\\ \Longleftrightarrow & M^c, \mu \models P(u) && (\text{FOL semantics})\\ \Longleftrightarrow & M^c, \mu \models \mathcal{R}[p, \sigma, u, v] && (\text{Def. } \mathcal{R}[\cdot])\end{aligned}$$

The induction on $\neg, \wedge$ is straightforward. For induction on the $\boldsymbol{K}_i$-operator where $end(\sigma) \neq i$, as hypothesis we assume that for any $(w, \vec{e})$, $s \in e_i$,

$$s \cup_i e_i \models \alpha \text{ iff } M^c, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, u', u] \tag{2}$$

Where $\mu'(u') = (s \cup_i e_i)$, $\mu'(u) = (w, \vec{e})$, then

$$\begin{aligned}&w, \vec{e} \models \boldsymbol{K}_i\alpha\\ \Longleftrightarrow&\text{for all } s' \in e_i, s' \cup_i e_i \models \alpha && (\mathcal{ML}_n \text{ semantics})\\ \Longleftrightarrow&\text{for all } s' \in e_i \text{ and } \mu' \text{ s.t. } \mu'(u') = (s' \cup_i e_i),\\ &\mu'(u) = (w, \vec{e}), M^c, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, u', u] && (\text{I.H.})\\ \Longleftrightarrow&\text{for all } \mu' \text{ s.t. } \mu'(u) = (w, \vec{e}) \text{ and } M^c, \mu' \models E_{\sigma\cdot i}(u', u),\\ &M^c, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, u', u] && (\text{cond. C3})\\ \Longleftrightarrow&M^c, \mu \models \forall u' E_{\sigma\cdot i}(u', u) \to \mathcal{R}[\alpha, \sigma\cdot i, u', u] && (\text{FOL semantics})\\ \Longleftrightarrow&M^c, \mu \models \mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] && (\text{Def. } \mathcal{R}[\cdot])\end{aligned}$$

Considering the case $\sigma = \epsilon$, then it is proved that for $\mu(u) = (w, \vec{e})$,

$$(w, \vec{e}) \models \boldsymbol{K}_i\alpha \text{ iff } M^c, \mu \models \mathcal{R}[\boldsymbol{K}_i\alpha, \epsilon, u, v], \tag{3}$$

which is part (1) of the theorem. To complete the induction on $\boldsymbol{K}_i$ for part (2), it remains to prove that for any $\sigma \in \mathrm{Ag}^*$,

$$w, \vec{e} \models \boldsymbol{K}_i\alpha \text{ iff } M^c, \mu \models_{FO} \mathcal{R}[\boldsymbol{K}_i\alpha, \sigma\cdot i, u, v] \tag{4}$$

$$\begin{aligned}&M^c, \mu \models \mathcal{R}[\boldsymbol{K}_i\alpha, \sigma\cdot i, u, v]\\ \Longleftrightarrow&M^c, \mu \models \forall u'.E_{\sigma\cdot i}(u', v) \to \mathcal{R}[\alpha, \sigma\cdot i, u', v] && (\text{Def. } \mathcal{R}[\cdot])\\ \Longleftrightarrow&M^c, \mu' \models E_{\sigma\cdot i}(u', v) \to \mathcal{R}[\alpha, \sigma\cdot i, u', v] \text{ for all } \mu' \sim_{u'} \mu\\ &&& (\text{FOL semantics})\\ \Longleftrightarrow&\text{ for all } \mu' \sim_{u'} \mu, M^c, \mu' \models E_{\sigma\cdot i}(u', v) \Rightarrow M^c, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, u', v]\\ \Longleftrightarrow&\text{ for all } \mu' \sim_{u'} \mu, \text{ if } \mu'(u') = (s \cup_i e_i) \text{ for some } s \in e_i\\ &\text{then } M^c, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, u', v] && (\text{cond. C3}, e_i' = e_i)\\ \Longleftrightarrow&\text{ for all } \mu' \sim_{u'} \mu, \text{ if } \mu'(u') = (s \cup_i e_i) \text{ for some } s \in e_i\\ &\text{then } s \cup_i e_i \models \alpha && (\text{I.H case 2})\\ &\text{Since for all } s \in e_i, \mu' \text{ exists s.t. } \mu' \sim_{u'} \mu \text{ and } \mu'(u') = s \cup_i e_i\\ \Longleftrightarrow&(s \cup_i e_i) \models \alpha \text{ for all } s \in e_i\\ \Longleftrightarrow&w, \vec{e} \models \boldsymbol{K}_i\alpha && (\mathcal{ML}_n \text{ semantics})\end{aligned}$$

□

3.3 Preserving Satisfiability and Decidability

Thm. 5 implies that if the original formula is $\mathcal{ML}_n$-satisfiable, then the translation is FO-satisfiable. In particular, it is satisfied by the canonical model M^c. We are also interested in the other direction, namely, when the translation is satisfied in some model M (not necessarily the canonical one), whether a model for the modal formula exists. We construct a set of mappings from the domain of M to structures and models of $\mathcal{ML}_n$. It is similar to the proof of Lem. 3: Given a FO model M, element $\omega, \omega' \in \mathcal{D}^M$ and $k > 0$, we define mappings to structures $\phi^{(\sigma)}_{M,k}$ and mapping to models $\Phi^{(\sigma)}_{M,k}$.

- When $|\sigma| = k$, $\phi^{(\sigma\cdot i)}_{M,k}(\omega) = \varnothing$;
- When $|\sigma| < k$, $\phi^{(\sigma\cdot i)}_{M,k} = (w, e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n)$, where $w = \{p \mid \omega \in P^M\}$, $e_j = \{\phi^{(\sigma\cdot i\cdot j)}_{M,k}(\omega') \mid (\omega', \omega) \in E^M_{\sigma\cdot i\cdot j}\}$.

Given $\omega, \omega' \in \mathcal{D}^M$, we define $\Phi^{(\sigma)}_{M,k}(\omega, \omega') = (w, \vec{e})$ as follows

- $w = \{p \in AP \mid \omega \in P^M\}$
- For any $j \neq end(\sigma)$, $e_j = \{\phi^{(\sigma\cdot j)}_{M,k}(\omega'') \mid (\omega'', \omega) \in E^M_{\sigma\cdot j}\}$
- For $i = end(\sigma)$, $e_i = \{\phi^{(\sigma)}_{M,k}(\omega'') \mid (\omega'', \omega') \in E^M_{\sigma}\}$

For simplicity, we omit the subscripts M, k when the context is clear. When $\sigma = \epsilon$, the model is completely determined by ω. By definition of $\phi^{(\sigma)}$ and $\Phi^{(\sigma)}$, we have the following property:

Lemma 4. *$(\omega, \omega') \in E^M_{\sigma\cdot i}$ if and only if $\Phi^{(\sigma\cdot i)}(\omega, \omega') = s \cup_i e_i$ for some $s \in e_i$ and $e_i = \{\phi^{(\sigma\cdot i)}(\omega'') \mid (\omega'', \omega') \in E^M_{\sigma\cdot i}\}$*

As we will show, given that (M, μ) satisfies the translating formula $\mathcal{R}[\alpha, \epsilon, u, v]$, the original formula will be satisfied by the $\mathcal{ML}_n$ model $(w, \vec{e}) = \Phi^{(\epsilon)}(\mu(u), \mu(v))$.

Lemma 5. *Given $\alpha \in \mathcal{ML}_n$ and $\sigma \in \mathrm{Ag}^*$, Then for μ such that $\mu(u) = \omega$, $\mu(v) = \omega'$,*

(1) *If $(w, \vec{e}) = \Phi^{(\epsilon)}(\omega, \omega')$, then*
$$M, \mu \models \mathcal{R}[\alpha, \epsilon, u, v] \text{ iff } (w, \vec{e}) \models \alpha$$
(2) *If $(w, \vec{e}) = \Phi^{(\sigma\cdot i)}(\omega, \omega')$ and $(\omega, \omega') \in E^M_{\sigma\cdot i}$*
$$M, \mu \models \mathcal{R}[\alpha, \sigma\cdot i, u, v] \text{ iff } (w, \vec{e}) \models \alpha$$

Proof. Since induction proofs with similar spirit are shown repeatedly, we only give the induction step for $\boldsymbol{K}$-operators. For each $i \in \mathrm{Ag}$, when $i \neq end(\sigma)$,

$$\begin{aligned}
& M, \mu \models \mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] \\
\Longleftrightarrow\ & M, \mu \models \forall v\, E_{\sigma\cdot i}(v, u) \rightarrow \mathcal{R}[\alpha, \sigma\cdot i, v, u] && \text{(Def. } \mathcal{R}[\cdot]) \\
\Longleftrightarrow\ & \text{for all } \mu' \sim_v \mu, \text{ if } (\mu'(v), \mu'(u)) \in E^M_{\sigma\cdot i} \text{ then} \\
& M, \mu' \models \mathcal{R}[\alpha, \sigma\cdot i, v, u] && \text{(FOL semantics)} \\
\Longleftrightarrow\ & \text{for all } \mu' \sim_v \mu, \text{ if } (\mu'(v), \mu'(u)) \in E^M_{\sigma\cdot i} \text{ then} \\
& w', \vec{e'} \models \alpha \text{ for } (w', \vec{e'}) = \Phi^{(\sigma\cdot i)}(\mu'(v), \mu'(u)) && \text{(I.H. case 2)} \\
\Longleftrightarrow\ & \text{for all } s \in e_i, s \cup_i e_i \models \alpha && \text{(Lem. 4)} \\
\Longleftrightarrow\ & e_i \models \boldsymbol{K}_i\alpha && (\mathcal{ML}_n \text{ semantics}) \\
\Longleftrightarrow\ & w, \vec{e} \models \boldsymbol{K}_i\alpha
\end{aligned}$$

Part (1) is covered when $\sigma = \epsilon$. Analogously we proof the induction for the case when $i = end(\sigma)$. □

We introduce a constant c and define $\mathcal{T}[\alpha]$ as a sentence replacing every free appearance of u in $\mathcal{R}[\alpha, \epsilon, u, v]$ by c.

Theorem 6. *α is $\mathcal{ML}_n$-satisfiable iff $\mathcal{T}[\alpha]$ is FO-satisfiable.*

Proof. Suppose that α is $\mathcal{ML}_n$-satisfiable, by Thm. 5 M^c satisfies $\mathcal{R}[\alpha, \epsilon, u, v]$ with some μ. By Skolemization, we replace the free variable u with constant c_w and it results in $\mathcal{T}[\alpha]$ (v never occurs in $\mathcal{R}[\alpha, \epsilon, u, v]$ as a free variable so it can be ignored). For the other direction, suppose that $\mathcal{T}[\alpha]$ is FO-satisfiable, then $\mathcal{R}[\alpha, \epsilon, u, v]$ is satisfied in some (M, μ). By Lem. 5, α is $\mathcal{ML}_n$-satisfiable. □

With satisfiability-preservation, the properties of validity checking and theorem proving can also be derived:

Corollary 1. *Given finite Σ and $\alpha \in \mathcal{ML}_n$,*

- *$\Sigma \models \alpha$ iff $\mathcal{T}[\Sigma] \models \mathcal{T}[\alpha]$;*
- *α is valid iff $\mathcal{T}[\alpha]$ is FO-valid.*

We also show that the translation will convert any $\mathcal{ML}_n$ formula to a decidability-preserving formula in linear time and space.

Theorem 7. *For any $\alpha \in \mathcal{ML}_n$, $\mathcal{T}[\alpha]$ can be computed in $O(|\alpha|)$, $\mathcal{T}[\alpha] \in GF^2$ and the size of $\mathcal{T}[\alpha]$ is linear in $|\alpha|$.*

Proof. Let $\varphi = \mathcal{T}[\alpha]$, it is obvious that $\varphi \in FO^2$ since the translation only uses variables u, v. By induction on the structure of φ we prove $\varphi \in GF$. As for the base case, $\mathcal{R}[P, \sigma, u, v] = P(u)$, which is guarded. Negation or conjunction does not affect the guarded condition. Quantifiers only occur in translating a (sub-)formula of form $\boldsymbol{K}_i\alpha$. Apparently, the resulting formula has a guard E_σ containing all free variables in the conclusion. Linear complexity in terms of time and size is straightforward. By Def. 7, the number of translation steps is linear in $|\alpha|$, and for each step, only a constant number of symbols will be added to the resulting formulae. □

Since the translation results in GF^2, as the straightforward result, the satisfiability of the resulting formulae is decidable.

Corollary 2. *Given $\alpha \in \mathcal{ML}_n$, the satisfiability of $\mathcal{T}[\alpha]$ can be decided in EXPTIME.*

Note that the size of standard translation ST_u also grows linearly, our approach $\mathcal{T}$ is as succinct as ST_u. We use the following example to compare two translations.

Example 4. *Let $\alpha := \boldsymbol{K}_1 p \wedge \neg \boldsymbol{K}_1 \boldsymbol{K}_1 p$. Apparently, α is not $K45_n$-satisfiable. The standard translation $ST_u[\alpha]$ is*
$$\forall v(R_1(u, v) \rightarrow P(v)) \wedge \neg\forall v(R_1(u, v) \rightarrow \forall u(R_1(v, u) \rightarrow P(v)))$$
The resulting formula is FO-satisfiable. To reason about the $K45_n$-properties via the standard translation, for each $i \in \mathrm{Ag}$ the following axioms need to be added:
$$\forall u \forall v \forall v'(R_i(u, v) \wedge R_i(v, v')) \rightarrow R_i(u, v') \qquad \text{(Transitivity)}$$
$$\forall u \forall v \forall v'(R_i(u, v) \wedge R_i(u, v')) \rightarrow R_i(v, v') \qquad \text{(Euclideanity)}$$
Apparently, these axioms are not in FO^2 since three variables are needed, and they are not in GF since a guard containing all free variables does not exist.[5] Considering the same formula α, our translation $\mathcal{R}[\alpha, \epsilon, u, v]$ is as follows:
$$\forall v(E_1(v, u) \rightarrow P(v)) \wedge \neg\forall v(E_1(v, u) \rightarrow \forall v(E_1(v, u) \rightarrow P(v)))$$

[5] It can be shown that these axioms are not even in the *loosely guarded fragment* [34]

Which preserves the $K45_n$-(un)satisfiability and no extra axioms are needed.

Let $\mathcal{T}_D[\alpha]$ be the sentence which replaces every free appearance of u in $\mathcal{R}_D[\alpha, \epsilon, u, v]$ by c (and $\mathcal{T}_T[\alpha]$ is defined similarly). The translation for $KD45_n/S5_n$ systems is obtained.

THEOREM 8. *For any $\alpha \in \mathcal{ML}_n$,*

- *α is $KD45_n$-consistent iff $\mathcal{T}_D[\alpha]$ is FO-satisfiable.*
- *α is $S5_n$-consistent iff $\mathcal{T}_T[\alpha]$ is FO-satisfiable.*

THEOREM 9. *For $* \in \{T, D\}$, $\alpha \in \mathcal{ML}_n$, $\mathcal{T}_*[\alpha]$ can be computed in $O(|\alpha|)$, $|\mathcal{T}_*[\alpha]| \in O(|\alpha|)$ and $\mathcal{T}_*[\alpha] \in GF^2$.*

4 FIRST-ORDER MODAL LOGIC

Our translation method can be extended to first-order modal logic (FOML). We lift $\mathcal{ML}_n$ to $\mathcal{KL}_n$, which can be considered as the multi-agent extension of the logic $\mathcal{KL}$ [24] and the fragment of $\mathcal{OL}_n$ [3] without mentioning only-knowing or the "knowing at most" modalities.

4.1 Syntax and Semantics of logic $\mathcal{KL}_n$

Compared with ordinary first-order (modal) logic, $\mathcal{KL}_n$ contains a fixed, countably infinite domain of discourse $\mathcal{N}$, where each element refers to a unique *standard name*. There is a countably infinite supply of predicate and functional symbols of every arity. Formally, we have the following definition:

- By *terms* of the language we mean the minimal set which contains all variables, standard names, and expressions of form $f(t_1, \dots, t_m)$ where f is a functional symbol and each t_j itself is a term.
- The *well-formed formulae* are inductively defined as follows:
  - For terms t and t', $t = t'$ is a formula,
  - For predicate symbol P and terms $t_1, \dots, t_m$, $P(t_1, \dots, t_m)$ is a formula,
  - If α, β are formulae, then $\neg\alpha, \alpha \wedge \beta, \forall x\, \alpha, \boldsymbol{K}_i\alpha$ are formulae.

We say a term is *ground* if it contains no variables. A ground term containing only a single function symbol is called *primitive*. A formula of form $P(t_1, \dots, t_l)$ is called an *atomic formula* or simply an atom. Furthermore, an atom with no variables is called *ground* and a ground atom with no function symbols is called *primitive*.

Similar to $\mathcal{ML}_n$, the semantics of $\mathcal{KL}_n$ is also based on k-structures, but the worlds in $\mathcal{KL}_n$ are defined as assignments of primitive terms to the co-referring standard names and primitive atoms to $\{0, 1\}$. For any ground term t, the value of t at world w, written as $|t|_w$, is defined as follows:

- If t is a standard name, then $|t|_w = t$;
- For function f, $|f(t_1, \dots, t_m)|_w = w[f(|t_1|_w, \dots, |t_m|_w)]$.

The satisfaction relation is defined as follows:

- $(w, \vec{e}) \models P(t_1, \dots, t_m)$ iff $w[P(|t_1|_w, \dots, |t_m|_w)] = 1$;
- $(w, \vec{e}) \models (t_1 = t_2)$ iff $|t_1|_w$ is the same name as $|t_2|_w$;
- $(w, \vec{e}) \models \neg\alpha$ iff it is not the case that $(w, \vec{e}) \models \alpha$
- $(w, \vec{e}) \models \alpha \wedge \beta$ iff $(w, \vec{e}) \models \alpha$ and $(w, \vec{e}) \models \beta$
- $(w, \vec{e}) \models \forall x\, \alpha$ iff $(w, \vec{e}) \models \alpha^x_n$ for all $n \in \mathcal{N}$
- $(w, \vec{e}) \models \boldsymbol{K}_i\varphi$ iff for all $s \in e_i$, $s \cup_i e_i \models \varphi$

Here α^x_n means the formula obtained by substituting each occurrence of free x in α by n. Notions such as satisfiability, validity, or entailment are defined similar to $\mathcal{ML}_n$. The properties of $\mathcal{KL}_n$ and the comparison with the ordinary first-order logic are discussed in [3, 24]. It is shown that $\mathcal{KL}_n$ satisfies the $K45_n$ properties and first-order properties such as the *Barcan formula*:

- $\models \forall x\, \boldsymbol{K}_i\alpha \to \boldsymbol{K}_i\forall x\, \alpha$
- $\models \exists x\, \boldsymbol{K}_i\alpha \to \boldsymbol{K}_i\exists x\, \alpha$

4.2 Translation for $\mathcal{KL}_n$

We extend the translation $\mathcal{R}$ to deal with first-order modal formulae. Note that the translation implicitly results in a two-sorted first-order logic[5]: Quantifiers in the original formulae are of sort *object* and quantifiers introduced during translation are of sort *state*. It is commonplace that many-sorted logic is easily convertible to ordinary first-order logic [10]. Here we present the version where one-sorted conversion has been embedded:

(1) For atomic formula α, $\mathcal{R}[\alpha, \sigma, u, v] := \alpha \uparrow^u$, where:
  - For $n \in \mathcal{N}$, $n \uparrow^u := n$
  - For variable x, $x \uparrow^u := x$
  - $f(t_1, \dots t_m) \uparrow^u := f(t_1 \uparrow^u, \dots, t_m \uparrow^u, u)$
  - $P(t_1, \dots t_m) \uparrow^u := P(t_1 \uparrow^u, \dots, t_m \uparrow^u, u)$
  - $(t_1 = t_2) \uparrow^u := (t_1 \uparrow^u = t_2 \uparrow^u)$

(2) $\mathcal{R}[\neg\alpha, \sigma, u, v] := \neg\mathcal{R}[\alpha, \sigma, u, v]$
(3) For $\odot \in \{\wedge, \vee, \to, \leftrightarrow\}$,
$\mathcal{R}[\alpha_1 \odot \alpha_2, \sigma, u, v] := \mathcal{R}[\alpha_1, \sigma, u, v] \odot \mathcal{R}[\alpha_2, \sigma, u, v]$
(4) If $end(\sigma) = i$, $\mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$\forall u\, (E_\sigma(u, v) \wedge S(u)) \to (\mathcal{R}[\alpha, \sigma, u, v])$
(5) If $end(\sigma) \neq i$, $\mathcal{R}[\boldsymbol{K}_i\alpha, \sigma, u, v] :=$
$\forall v\, (E_{\sigma \cdot i}(v, u) \wedge S(v)) \to (\mathcal{R}[\alpha, \sigma \cdot i, v, u])$
(6) $\mathcal{R}[\forall x\, \alpha, \sigma, u, v] := \forall x\, O(x) \to \mathcal{R}[\alpha, \sigma, u, v]$
(7) $\mathcal{R}[\exists x\, \alpha, \sigma, u, v] := \exists x\, O(x) \wedge \mathcal{R}[\alpha, \sigma, u, v]$

Predicates O and S denote the membership of sort object and state. Let $\mathcal{T}[\alpha]$ be to replace every free appearance of u in $\mathcal{R}[\alpha, \epsilon, c, v]$.

THEOREM 10. *If α is $\mathcal{KL}_n$-satisfiable then $\mathcal{T}[\alpha]$ is FO-satisfiable.*

Similar to Thm. 5, we prove it by constructing a canonical model. Different from the propositional case, the other direction does not hold naively. A counter-example can be easily constructed:

EXAMPLE 5. *Let $\alpha := \forall x\, (x = n_1 \vee x = n_2)$, then $\mathcal{T}[\alpha]$ is satisfiable but α is not $\mathcal{KL}_n$-satisfiable.*

Note that the difference has already been discussed in [24]. It is mainly because $\mathcal{KL}_n$ interprets equality in the usual sense, and requires a countably infinite domain of discourse where each element has a unique name. To achieve satisfiability-preservation, it requires additional assumptions Σ_{eq} about equality:[6]

- $(x = x)$
- $(x = y) \to (y = x)$
- $((x = y) \wedge (y = z)) \to (x = z)$
- for any function symbol f in $\mathcal{T}[\alpha]$:
$((x_1 = y_1) \wedge \dots \wedge (x_m = y_m)) \to f(x_1, \dots, x_m) = f(y_1, \dots, y_m)$
- for any predicate symbol P in $\mathcal{T}[\alpha]$:
$((x_1 = y_1) \wedge \dots \wedge (x_m = y_m)) \to P(x_1, \dots, x_m) \equiv P(y_1, \dots, y_m)$

$\Sigma_{una} = \{(n \neq n') \mid n, n'$ are distinct standard names in $\mathcal{N}\}$.
Σ_{sort} includes the following formulae:

[6] Here all free variables are implicitly universally quantified.

- $\exists x\, O(x)$
- $S(c)$
- For each function symbol f in $\mathcal{T}[\alpha]$:
  $(O(x_1) \wedge \ldots \wedge O(x_m) \wedge S(u)) \rightarrow O(f(x_1, \ldots x_m, u))$
- $O(n)$ for all $n \in \mathcal{N}$

Intuitively, Σ_{eq} means that the symbol = is interpreted in its usual sense. Σ_{una} requires the domain of discourse to be infinite. Σ_{sort} preserves the satisfiability in the conversion from many-sorted to one-sorted (See [10, 24] for details). Let $\Sigma = \Sigma_{eq} \cup \Sigma_{una} \cup \Sigma_{sort}$, similar with Lem. 5 we have

Lemma 6. *Given $\alpha \in \mathcal{KL}_n$, if $\mathcal{T}[\alpha] \cup \Sigma$ is FO-satisfiable, then α is $\mathcal{KL}_n$-satisfiable.*

By Thm. 10 and Lem. 6, we derive the following theorem.

Theorem 11. *α is $\mathcal{KL}_n$-satisfiable iff $\mathcal{T}[\alpha] \cup \Sigma$ is FO-satisfiable.*

As shown in the following corollary, the infinite set of axioms Σ is needed only when equality or standard names occur in the formula. Let Σ'_{sort} be the subset of Σ_{sort} without $O(n)$ of any n:

Corollary 3. *For $\alpha \in \mathcal{KL}_n$ without standard names or equality, α is $\mathcal{KL}_n$-satisfiable iff $\mathcal{T}[\alpha] \cup \Sigma'_{sort}$ is FO-satisfiable.*

Since FOL is a fragment of $\mathcal{KL}_n$, it is clear that the decision problems of $\mathcal{KL}_n$ are undecidable. Thus it is valuable to investigate decidable first-order modal fragments. However, as shown in previous work including [5], the full modal extensions of many decidable first-order fragments, including the monadic fragment (with only unary predicates), *GF* and FO^2, cannot preserve decidability. With an analysis of the translation, we study the decidability of a fragment, which we call the *modal bounded guarded fragment* (*MBGF*).

Definition 9 (modal bounded guarded fragment). *A formula $\alpha \in \mathcal{KL}_n$ is in MBGF, if it contains no standard names or functional symbols,[7] and one of the following holds:*

- *α is an atom $P(\vec{t})$;*
- *α is a Boolean combination of formulae in MBGF;*
- *α is of form $\boldsymbol{K}_i\varphi$, where φ is a bounded MBGF;*
- *α is of form $\forall\vec{x}.G(\vec{x},\vec{y}) \rightarrow \varphi$, where φ is objective and is in MBGF, $G(\vec{x},\vec{y})$ is an atom and $Free(\varphi) \subseteq \{\vec{x},\vec{y}\}$, i.e. all free variables in φ are in $\vec{x}$ or $\vec{y}$.*

Intuitively, a formula is in *MBGF* if all the objective sub-formulae are in *GF*, and all subjective sub-formulae are bounded. For example, a formula of form $\boldsymbol{K}_1(\alpha \wedge (\boldsymbol{K}_2\beta \rightarrow \boldsymbol{K}_1\gamma))$ is in *MBGF*, provided that α, β, γ are all bounded formulae in *GF*. However, $\boldsymbol{K}_1\forall x(G(x) \rightarrow \boldsymbol{K}_1 F(x))$ is not in *MBGF*.

We consider the loosely guarded fragment (*LGF*), where the notion of guard is relaxed. In *LGF* quantifiers can be guarded by a conjunction of atomic formulae, provided that each quantified variable co-exists with each free or bounded variable in some atom. *LGF* has very similar properties with *GF*. It has been proved that the satisfiability problem for both *GF* and *LGF* is 2EXPTIME-complete [13].

Theorem 12. *For any $\alpha \in MBGF$, $\mathcal{R}[\alpha, \sigma, u, v] \in LGF$.*

[7] Constants, which can be understood as nullary functions, can be accepted though.

Proof. When $\alpha := P(\vec{t})$, $\mathcal{R}[\alpha, \sigma, u, v] = P(\vec{t} \uparrow^u, u)$. Thus the theorem holds for all atomic formulae. The induction on $\vee, \wedge, \neg$ is trivial. Suppose that α is of form $\boldsymbol{K}_i\varphi$ where φ is a bounded *MBGF*, i.e. φ has no free variables. When $end(\sigma) = i$,

$$\mathcal{R}[\alpha, \sigma, u, v] = \forall u\ (E_\sigma(u, v) \wedge S(u)) \rightarrow \mathcal{R}[\varphi, \sigma, u, v]$$

By induction hypothesis $\mathcal{R}[\varphi, \sigma, u, v] \in LGF$ and has no free variables other than u, v. Thus $\mathcal{R}[\alpha, \sigma, u, v] \in LGF$. The proof is similar when $end(\sigma) \neq i$. Suppose that α is of form $\forall\vec{x}.G(\vec{x},\vec{y}) \rightarrow \varphi$, then

$$\begin{aligned} &\mathcal{R}[\forall\vec{x}.G(\vec{x},\vec{y}) \rightarrow \varphi, \sigma, u, v] \\ =&\forall\vec{x}(O(x_1) \wedge \ldots \wedge O(x_m) \wedge G(\vec{x},\vec{y},u)) \rightarrow \mathcal{R}[\varphi, \sigma, u, v] \end{aligned}$$

By induction hypothesis $\mathcal{R}[\varphi, \sigma, u, v] \in LGF$ and $\mathcal{R}[\varphi, \sigma, u, v]$ has free variables among $\{\vec{x},\vec{y},u\}$. Since each pair in $\{\vec{x},\vec{y}\}$ co-occurs in $G(\vec{x},\vec{y},u)$, we have $\mathcal{R}[\alpha, \sigma, u, v] \in LGF$. □

Corollary 4. *Given $\alpha \in MBGF$, the satisfiability of α can be decided in 2EXPTIME.*

Proof. By Cor. 3, α is satisfiable iff $\mathcal{T}[\alpha] \cup \Sigma'_{sort}$ is FO-satisfiable. Thm. 12 shows that $\mathcal{T}[\alpha] \in LGF$. Since $\alpha \in MBGF$, α contains at most nullary functions and hence $\mathcal{T}[\alpha]$ contains functional terms with at most one variable. Thus $\Sigma'_{sort} \in LGF$, and $\mathcal{T}[\alpha] \cup \Sigma'_{sort}$ can be decided in 2EXPTIME. □

5 CONCLUSION

With the representation of a set-based semantics in first-order logic, we propose a decidability-preserving translation for multi-agent logics of knowledge and belief, and extend the approach to first-order modal logics. To the best of our knowledge, this approach is the first that preserves the concision of the standard translation and maintains the translation in a decidable fragment. Our approach enables the translation-based reasoning for multi-agent modal systems, including $K45_n$, $KD45_n$ and $S5_n$, which has been open for decades. In addition, the methodology we follow shows the potential of non-Kripke semantics and it appeals to more attention to them.

As for future work, we plan to extend our approach to consider temporal modalities. Translation for common knowledge or probabilistic belief will also be possible directions. For the study on complexity and decidability, we believe there could be a first-order fragment which is more essential for representing modal logics. On the one hand, the translation does not utilize the full language of GF^2 and it can be further restricted to the so-called *monadic* GF^2_- with binary guards and no constants. On the other hand, the decision problem of propositional modal logic is known to be PSPACE-complete [17]. With the conjecture PSPACE⊊EXPTIME, translation into a more essential fragment would be possible.

ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/2 'UnRAVeL' and the EU ICT-48 2020 project TAILOR (No. 952215). It is also supported by Mitacs Canada through the Mitacs Globalink Research Internship program.

REFERENCES

[1] Yves Auffray and Patrice Enjalbert. 1989. Modal theorem proving: an equational viewpoint. In *Proceedings of the 11th international joint conference on Artificial intelligence-Volume 1*. 441–445.

[2] Vaishak Belle and Gerhard Lakemeyer. 2010. Multi-agent only-knowing revisited. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning*. 49–59.

[3] Vaishak Belle and Gerhard Lakemeyer. 2015. Semantical considerations on multiagent only knowing. *Artificial Intelligence* 223 (2015), 1–26.

[4] Thomas Bolander. 2017. A gentle introduction to epistemic planning: The DEL approach. *arXiv preprint arXiv:1703.02192* (2017).

[5] Torben Brauner and Silvio Ghilardi. 2007. First Order Modal Logic. *STUDIES IN LOGIC AND PRACTICAL REASONING* 3 (2007), 549–620.

[6] Laurent Catach. 1991. TABLEAUX: A general theorem prover for modal logics. *Journal of automated reasoning* 7 (1991), 489–510.

[7] Brian F Chellas. 1980. *Modal logic: an introduction*. Cambridge university press.

[8] Hans De Nivelle, Renate A Schmidt, and Ulrich Hustadt. 2000. Resolution-based methods for modal logics. *Logic Journal of the IGPL* 8, 3 (2000), 265–292.

[9] Stéphane Demri and Hans De Nivelle. 2005. Deciding regular grammar logics with converse through first-order logic. *Journal of Logic, Language and Information* 14 (2005), 289–329.

[10] Herbert B Enderton. 2001. *A mathematical introduction to logic*. Elsevier.

[11] Kit Fine. 1975. Some connections between elementary and modal logic. In *Studies in Logic and the Foundations of Mathematics*. Vol. 82. Elsevier, 15–31.

[12] H Ganzinger, C Meyer, and M Veanes. 1999. The Two-Variable Guarded Fragment with Transitive Relations. In *Proceedings of the 14th Annual IEEE Symposium on Logic in Computer Science*. 24–34.

[13] Erich Grädel. 1999. On the restraining power of guards. *The Journal of Symbolic Logic* 64, 4 (1999), 1719–1742.

[14] Erich Grädel and Martin Otto. 1999. On logics with two variables. *Theoretical computer science* 224, 1-2 (1999), 73–113.

[15] Joseph Y Halpern. 1995. The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence* 75, 2 (1995), 361–372.

[16] Joseph Y Halpern and Yoram Moses. 1990. Knowledge and common knowledge in a distributed environment. *Journal of the ACM (JACM)* 37, 3 (1990), 549–587.

[17] Joseph Y Halpern and Yoram Moses. 1992. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial intelligence* 54, 3 (1992), 319–379.

[18] Jaakko Hintikka. 1962. Knowledge and Belief: An Introduction to the Logic of the Two Notions.

[19] Laura Kovács and Andrei Voronkov. 2013. First-order theorem proving and Vampire. In *International Conference on Computer Aided Verification*. Springer, 1–35.

[20] Saul A Kripke. 1963. Semantical analysis of modal logic i normal modal propositional calculi. *Mathematical Logic Quarterly* 9, 5-6 (1963), 67–96.

[21] Richard E Ladner. 1977. The computational complexity of provability in systems of modal propositional logic. *SIAM journal on computing* 6, 3 (1977), 467–480.

[22] Hector Joseph Levesque. 1982. A formal treatment of incomplete knowledge bases.

[23] Hector Joseph Levesque. 1990. All I know: a study in autoepistemic logic. *Artificial intelligence* 42, 2-3 (1990), 263–309.

[24] Hector J Levesque and Gerhard Lakemeyer. 2001. *The logic of knowledge bases*. Mit Press.

[25] William McCune. 2005–2010. Prover9 and Mace4. (2005–2010). http://www.cs.unm.edu/ mccune/prover9/.

[26] Morgan. 1976. Methods for automated theorem proving in nonclassical logics. *IEEE Trans. Comput.* 100, 8 (1976), 852–862.

[27] Andreas Nonnengart. 1993. First-order modal logic theorem proving and functional simulation. In *Proceedings of the 13th international joint conference on Artifical intelligence-Volume 1*. 80–85.

[28] Renate A Schmidt. 1999. Decidability by resolution for propositional modal logics. *Journal of Automated Reasoning* 22 (1999), 379–396.

[29] Renate A Schmidt and Ullrich Hustadt. 2003. A principle for incorporating axioms into the first-order translation of modal formulae. In *International Conference on Automated Deduction*. Springer, 412–426.

[30] Renate A Schmidt and Ullrich Hustadt. 2007. The axiomatic translation principle for modal logic. *ACM Transactions on Computational Logic (TOCL)* 8, 4 (2007), 19–es.

[31] Renate A Schmidt and Ullrich Hustadt. 2013. First-order resolution methods for modal logics. In *Programming Logics: Essays in Memory of Harald Ganzinger*. Springer, 345–391.

[32] Luc Segoufin and Balder ten Cate. 2013. Unary negation. *Logical Methods in Computer Science* 9 (2013).

[33] Balder Ten Cate and Luc Segoufin. 2011. Unary negation. In *28th International Symposium on Theoretical Aspects of Computer Science (STACS 2011)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik.

[34] Johan Van Benthem. 1997. Dynamic bits and pieces. (1997).

[35] JFAK van Benthem et al. 1977. Modal Correspondence Theory. (1977).

[36] Hans Van Ditmarsch, Wiebe van der Hoek, Joseph Y Halpern, and Barteld Kooi. 2015. *Handbook of epistemic logic*. College Publications.

[37] Moshe Y Vardi. 1996. Why is modal logic so robustly decidable? *Descriptive complexity and finite models* 31 (1996), 149–184.