Beth Definability for the Guarded Fragment

Eva Hoogland[1], Maarten Marx[2] and Martin Otto[3]

[1] Dept. of WINS, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands, ehooglan@wins.uva.nl
[2] Dept. of Artificial Intelligence, Faculty of Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands, marx@wins.uva.nl
[3] Mathematical Foundations of Computer Science, RWTH Aachen, Ahornstr.55, 52074 Aachen, Germany, otto@informatik.rwth-aachen.de

Abstract. The guarded fragment (GF) was introduced in [ABN98] as a fragment of first order logic which combines a great expressive power with nice modal behavior. It consists of relational first order formulas whose quantifiers are relativized by atoms in a certain way. While GF has been established as a particularly well-behaved fragment of first order logic in many respects, interpolation fails in restriction to GF, [HM99]. In this paper we consider the Beth property of first order logic and show that, despite the failure of interpolation, it is retained in restriction to GF. Being a closure property w.r.t. definability, the Beth property is of independent interest, both theoretically and for typical potential applications of GF, e.g., in the context of description logics. The Beth property for GF is here established on the basis of a limited form of interpolation, which more closely resembles the interpolation property that is usually studied in modal logics. From this we obtain that, more specifically, even every n-variable guarded fragment with up to n-ary relations has the Beth property.

1 Introduction

The Guarded Fragment It has proven useful to view modal logics not only as systems in themselves but also as fragments of first order logic. As is well-known, the basic modal logic K can be seen as a fragment of first order logic via the translation t which maps a proposition letter p to the atom Px, which commutes with the Boolean connectives, and which maps formulas of the form $\Diamond\varphi$ to $\exists y(Rxy \wedge \varphi^t(y))$ and $\Box\varphi$ to $\forall y(Rxy \rightarrow \varphi^t(y))$. The image of K under this translation is referred to as the *modal fragment*. This fragment turns out to behave excellently. It shares several nice model-theoretic properties with full first order logic (e.g., interpolation, Beth definability, or the Łos-Tarski property), and has in addition good algorithmic qualities: it is decidable and every satisfiable modal formula has a finite model and a tree model (in other words, the modal fragment has the finite model property and the tree model property). Moreover, the decidability of this fragment is robust in the sense that various extensions remain decidable. For example, adding features like counting quantifiers or fixed points to the modal fragment does not affect decidability.

The usefulness of the modal fragment brought logicians to search for generalizations of this fragment which retain the afore-mentioned nice properties. An obvious candidate for such a generalization is the two variable fragment of first order logic, denoted by L_2. Although this logic is decidable and has the finite model property, it does not have interpolation nor the Beth property. Neither does it have the tree model property, and also its decidability is not as robust as that of the modal fragment [Var98][GO99].

In [ABN98] it is argued that the distinguishing characteristic of the modal fragment is not its restriction to two variables but its restriction on quantifiers, namely to quantifier patterns $\exists y(Rxy \wedge \varphi(y))$ or $\forall y(Rxy \rightarrow \varphi(y)))$. This brings Andréka, van Benthem and Németi to investigate to what extent these quantifier restrictions can be loosened while retaining the attractive modal behavior. The outcome is the *guarded fragment* (GF) which allows for quantifications of the form $\exists \bar{y}(R\bar{x}\bar{y} \wedge \varphi(\bar{x}, \bar{y}))$ and $\forall \bar{y}(R\bar{x}\bar{y} \rightarrow \varphi(\bar{x}, \bar{y}))$, where $\bar{x}, \bar{y}$ are finite sequences of variables and φ is a guarded formula with free variables among $\bar{x}, \bar{y}$ which *all* must appear in the atomic formula $R\bar{x}\bar{y}$.

In [ABN98] this fragment is shown to have the finite model property, the Los-Tarski property and, most importantly, to be decidable. Grädel [Grä97] improves on this result by classifying the satisfiability problem for GF to be complete for deterministic double exponential time; satisfiability for the finite variable guarded fragments is even in $\mathsf{Exptime}$, in fact $\mathsf{Exptime}$-complete. This is worth comparing with the satisfiability problem for L_2 which is known to be $\mathsf{Nexptime}$-complete [GKV97]. What is more, GF has a certain tree model property. Since the tree model property of the modal fragment can be seen as the main reason behind the robustness of the decidability of that fragment (cf. e.g., [Var98]), this gives hope as to the robustness of GF. And indeed, adding least and greatest fixed points to GF yields a decidable extension [GW99].

However, as shown in [HM99], the interpolation theorem of first order logic fails for GF. In the present paper it will be shown that GF does have an alternative interpolation property, which closely resembles the interpolation property usually studied in modal logics. This result turns out to be strong enough to entail the Beth definability theorem for GF.

The Beth (Definability) Property In a slogan, the Beth definability property states that implicit definability equals explicit definability. Generally, this property may be regarded as an indication that there is a good balance between syntax and semantics of a logic: the semantic phenomenon that the meaning of a basic relation is implicitly determined, guarantees that there is an explicit syntactic expression for that relation. Intuitively, an *implicit definition* of a relation R is a definition of R, in the sense that it fixes the interpretation of R, in which the relation symbol R may occur. For example, consider the conjunction Σ of formulas saying that $<$ is an irreflexive linear order, there exists a first element and this element has property R, and an element has property R iff its successor does not have property R. Note that these statements can be formulated in first order logic (with equality) using the predicates $<$ and R. It is obvious that on every finite irreflexive linear order the interpretation of the relation R is fixed. In other words, on finite models, Σ implicitly defines R. On the other hand, as first observed by [Háj77], there is no first order formula $\theta(x)$ which does not mention R and which would *explicitly define* R over the finite models of Σ. I.e., there is *no* formula $\theta(x)$ using just $<$ such that $\Sigma \models Rx \leftrightarrow \theta(x)$ would be true over all finite models. Obviously, every relation that is explicitly definable is also implicitly definable. As the above example showed, the converse is in general not true. However, in the classical context of not necessarily finite models, implicit definability and explicit definability in first order logic coincide. This property of first order logic has first been observed by E.W. Beth (see [Bet53]). Nowadays, logics for which an analogous statement holds are said to have the *Beth (definability) property*. So the above-mentioned example shows that first order logic restricted to finite models does not have the Beth property. Another logic which fails to have this property is L_2 (cf. [Sai90], see also Remark 2). Besides first order logic, logics with the Beth property include classical (and intuitionistic) propositional calculus, or the modal logics K, $\mathsf{K4}$ and $\mathsf{S5}$.

Note that for GF, and the modal logics, as long as we consider *finite* sets of sentences Σ it does not make a difference for the Beth property whether we are in the classical context of not necessarily finite models, or regard finite models only. For, as these logics have the finite model property, a finite set of sentences Σ implicitly defines a relation over finite models if and only if it does so over all models. The same for explicit definitions.

Description Logics Description logics were designed for the purpose of knowledge representation. Roughly speaking, a description logic starts from some set of primitive concepts (which are unary predicates) and roles (binary predicates). The logic then specifies (or defines) complex concepts out of these primitives and makes assertions about these specifications, mostly in terms of modally expressible dependencies between concepts via roles. E.g., the logic can assert that a certain object, or all objects related to it via a designated role, belongs to a certain concept. Although they originated from entirely different backgrounds, there is therefore a close correspondence between description logics and modal logics. For example the description logic $\mathcal{ALC}$ is nothing but a syntactic variant of the basic multi-modal logic $\mathsf{K_n}$ [Sch91]. Hence the guarded fragment can also be seen as a general framework for description logics, which may express more than the ordinary modal dependencies. In particular, it may go beyond the built-in arity restriction of modal logics, so that one can speak of higher-arity concepts and roles. The interested reader is referred to [Grä98] for a proposal of GF as a framework for description logic and for further references. In the description logic context, the Beth property seems particularly desirable as it guarantees explicit definability of concepts (and roles): e.g., concept specification in the framework of GF is closed in the sense that any concept that can implicitly be characterized can actually be defined explicitly within the logic.

Outline of the Paper Ever since 1956 when W. Craig gave an alternative proof of the Beth theorem for first order logic via his interpolation theorem, these two properties are almost always studied simultaneously. This paper forms no exception. In Section 3 we will prove a certain interpolation property for GF from which the Beth property for GF will be derived in Section 4. Even better, both these properties will be shown to hold for each of the n-variable fragments of GF individually, in the presence of at most n-ary relations.

2 Preliminaries

In this section we will collect all the necessary preliminaries. It also serves to fix notation and terminology.

Convention 1 By a *language* $\mathcal{L}$ we will henceforth understand a relational first order language without function- or constant symbols. Besides variables, and the parentheses $),($, we consider as *logical symbols* the connectives $\wedge$, $\neg$, the existential quantifier $\exists$ and the identity symbol $=$. $\dashv$

Notation 1 Models are denoted by calligraphic letters like $\mathcal{M}$, $\mathcal{N}$, and their respective universes by M, N, etc. The interpretation of an n-ary predicate R in the model $\mathcal{M}$ (notation: $I^{\mathcal{M}}(R) \subseteq M^n$) is defined as usual. Moreover, we extend this terminology to sets. That is, for $X \subseteq M$ we write $X \in I^{\mathcal{M}}(R)$ if the elements of X are R-related, in any order or multiplicity. E.g., if R is ternary and $\langle n, m, n \rangle \in I^{\mathcal{M}}(R)$, then $\{m, n\} \in I^{\mathcal{M}}(R)$. For a model $\mathcal{M}$, $m_1, \ldots, m_n \in M$ and a formula φ with free variables among $\{v_1, \ldots, v_n\}$, we write $\mathcal{M} \models \varphi[m_1, \ldots, m_n]$ iff each assignment which maps v_i to m_i satisfies φ in $\mathcal{M}$. If Σ is a formula (or a set of formulas) and ψ a formula, then $\Sigma \models \psi$ denotes the *consequence* relation. That is, $\Sigma \models \psi$ iff any assignment into a model $\mathcal{M}$ which satisfies (all formulas in) Σ also satisfies ψ. In particular, $\varphi \models \psi$ is the same as to say that $\varphi \to \psi$ is valid, i.e., $\models \varphi \to \psi$. For any formula φ, by $free(\varphi)$ we denote the set of free variables occurring in φ. By $\mathcal{L}_\varphi$ (read: *the language of* φ) we denote the set of relation symbols occurring in φ. $\dashv$

2.1 Defining the Guarded Fragment: Syntax

Definition 1 (Guarded formula). Let $\mathcal{L}$ be a language. The *atomic $\mathcal{L}$-formulas* (or, $\mathcal{L}$-atoms) are of the usual form:

1. $v_1 = v_2$, for variables v_1, v_2.
2. $Pv_1 \cdots v_n$, for n-ary $P \in \mathcal{L}$ and variables $v_1, \ldots, v_n$, not necessarily distinct.

The *guarded $\mathcal{L}$-formulas* are defined by induction as follows.

1. Any atomic $\mathcal{L}$-formula is a guarded $\mathcal{L}$-formula.
2. If φ, ψ are guarded $\mathcal{L}$-formulas, then $\varphi \wedge \psi$ and $\neg\varphi$ are guarded $\mathcal{L}$-formulas.

3. Let $\bar{v}$ be a finite, non-empty sequence of variables, ψ a guarded $\mathcal{L}$-formula, and G an $\mathcal{L}$-atom such that $free(\psi) \subseteq free(G)$. Then $\exists\bar{v}(G \wedge \psi)$ is a guarded $\mathcal{L}$-formula. In this case, the atom G is called the *guard* of the quantifier. $\dashv$

Note that as a dual of guarded existential quantification we also get guarded universal quantification, of the form $\forall\bar{v}(G \rightarrow \psi)$.

A typical example of a guarded formula is the one expressing symmetry of a relation: $\forall v_1v_2(Rv_1v_2 \rightarrow Rv_2v_1)$. On the other hand, the formula $\forall v_1v_2v_3((Rv_1v_2 \wedge Rv_2v_3) \rightarrow Rv_1v_3)$, which expresses the transitivity of the relation R, is not guarded, as $Rv_1v_2 \wedge Rv_2v_3$ is not a guard.

Remark 1. For readers familiar with [ABN98] we note that contrary to that paper, Definition 1 allows for identity atoms as guards. Since this issue does not affect decidability nor interpolation, we decided to concentrate on this slightly more general fragment. This also places us in line with [Grä97]. $\dashv$

Guarded formulas are obviously first order formulas. The fragment of first order logic consisting of guarded formulas is called the *guarded fragment* (GF). We understand by $\mathrm{GF}_n, n \in \omega$, the fragment of GF that consists of formulas whose variables (free or bound) are among $v_1, \ldots, v_n$. The collection of formulas in GF_n which are built up from at most k-ary relation symbols is denoted by GF_n^k.

2.2 Semantic Characterization of the Guarded Fragment

Similar to modal logics, the guarded fragment can be semantically analyzed via a suitable notion of bisimulation. This has been done in [ABN98]. Here we will recapitulate as much of these results as needed for the purposes of the present paper.

Definition 2 (live set). Let Z be a finite subset of a model $\mathcal{M}$. The set Z is called *live* in $\mathcal{M}$ if Z is either a singleton, or there exists a relation R and a set X such that $Z \subseteq X \in I^{\mathcal{M}}(R)$. In this case we will say that Z is *R-live* (in $\mathcal{M}$). For any language $\mathcal{L}$ we use the notation $Z \subseteq^l_{\mathcal{L}} \mathcal{M}$ to denote that Z is $\mathcal{L}$-live in $\mathcal{M}$. That is, Z is R-live in $\mathcal{M}$ for some $R \in \mathcal{L}$. We will omit the subscript $\mathcal{L}$ if it does not cause confusion. $\dashv$

Note that by definition subsets of live sets are again live.

Below, by a finite *partial $\mathcal{L}$-isomorphism* we mean a finite one-to-one partial map between two models which preserves the relations in $\mathcal{L}$ both ways. By the *image* of a map $f : X \longrightarrow Y$ we understand the set $\{f(x) : x \in X\}$, and we refer to X as its *domain*.

Definition 3 (Guarded bisimulation). A *guarded $\mathcal{L}$-bisimulation* between two models $\mathcal{M}$ and $\mathcal{N}$ is a non-empty set F of finite partial $\mathcal{L}$-isomorphisms between $\mathcal{M}$ and $\mathcal{N}$ such that for any $f : X \longrightarrow Y \in F$ the following hold:

1. For any $Z \subseteq^l_{\mathcal{L}} \mathcal{M}$ there is a $g \in F$ with domain Z such that g and f agree on the intersection of their domains. (The *zig*-condition)
2. For any $W \subseteq^l_{\mathcal{L}} \mathcal{N}$ there is a $g \in F$ with image W such that g^{-1} and f^{-1} agree on the intersection of their domains. (The *zag*-condition) $\dashv$

Guarded bisimulations are defined in such a way as to preserve guarded formulas. That is, for a guarded $\mathcal{L}$-formula φ with free variables among $\{v_1, \ldots, v_k\}$, a guarded $\mathcal{L}$-bisimulation F between models $\mathcal{M}$, $\mathcal{N}$, an $f \in F$, and $m_1, \ldots, m_k \in dom(f)$ it is the case that $\mathcal{M} \models \varphi[m_1, \ldots, m_k] \;\Leftrightarrow\; \mathcal{N} \models \varphi[f(m_1), \ldots, f(m_k)]$. This can be shown by a straightforward induction on the complexity of φ. The zig- and zag-conditions precisely take care of the induction step for existential quantification. Indeed, preservation under guarded bisimulations is the characteristic feature of GF, in the sense of the following Characterization Theorem from [ABN98]: up to logical equivalence, GF precisely consists of those first order formulas that are preserved under guarded bisimulations.

Note that in the definition of a guarded bisimulation that can be found in [ABN98], the above role of live sets is taken over by what ABN call *guarded* sets. These are subsets Z of a model $\mathcal{M}$ such that $Z \in I^{\mathcal{M}}(R)$, for some relation R. Mutatis mutandis, all arguments in [ABN98] and in particular the characterization theorem also apply to guarded formulas and guarded bisimulations as defined in this paper. Note e.g., that a guarded formula of the form $\exists v(v = v \wedge \neg Pv)$ which is not guarded in the ABN-sense, is preserved under guarded bisimulations in our sense by virtue of the fact that singletons are live.

For further use we exhibit, for any relation R, a formula $\lambda_R(v_1, \ldots, v_l)$ which defines the set of R-live l-tuples. More precisely, for models $\mathcal{M}$ and $m_1, \ldots, m_l \in M$: $\mathcal{M} \models \lambda_R(v_1, \ldots, v_l)[m_1, \ldots, m_l]$ iff the set $\{m_1, \ldots, m_l\}$ is R-live in $\mathcal{M}$.

Let s be the arity of R. Let e range over all complete equality types in variables $v_1, \ldots, v_l$. We regard e both as a quantifier-free formula $e(v_1, \ldots, v_l)$ in the empty vocabulary and as an equivalence relation on the set $\{1, \ldots, l\}$ according to $(j, i) \in e$ iff $e \models v_j = v_i$. Let $\rho\colon \{1, \ldots, s\} \to \{1, \ldots, l+s\}$ be a mapping that is onto $\{1, \ldots, l\}/e$, i.e., for every $j \in \{1, \ldots, l\}$ there is some $i \in \{1, \ldots, s\}$ such that $\rho(i)$ is in the same e equivalence class with j. Put, for any such pair of e and ρ,

$$\gamma_{e,\rho} = e(v_1, \ldots, v_l) \wedge \exists \bar{v}\big(Rv_{\rho(1)} \ldots v_{\rho(s)} \wedge \mathit{true}\big), \tag{1}$$

where $\bar{v}$ consist of those $v_{\rho(i)}$ for which $\rho(i) > l$ (if there are such; else no quantification is necessary and $\gamma_{e,\rho}$ is actually atomic). The desired formula $\lambda_R(v_1, \ldots, v_l)$ is obtained as the disjunction over all $\gamma_{e,\rho}$ for matching pairs (e, ρ).

For any *finite* language $\mathcal{L}$ we further obtain a formula $\lambda_{\mathcal{L}}(v_1, \ldots, v_l)$ defining the set of $\mathcal{L}$-live l-tuples by putting

$$\lambda_{\mathcal{L}}(v_1, \ldots, v_l) = \Big(\bigwedge_{1 \le i,j \le l} v_i = v_j\Big) \vee \bigvee_{R \in \mathcal{L}} \lambda_R(v_1, \ldots, v_l), \tag{2}$$

where the first disjunct reflects the fact that all singleton sets are regarded as live (namely, as guarded by equality).

We finally note that $\gamma_{e,\rho}$ is equivalent to a formula in $\mathrm{GF}_{\max(l,s)}$. We conclude that for every finite language $\mathcal{L}$ which contains at most k-ary relations and for any $l \leq k$ we may assume $\lambda_{\mathcal{L}}(v_1, \ldots, v_l) \in \mathrm{GF}_k^k$.

3 Interpolation for the Guarded Fragment

As shown in [HM99], GF does not have the interpolation property.

Theorem 1 (Failure of interpolation in GF, [HM99]). *There exist sentences φ, $\psi \in \mathrm{GF}_3^3$ such that $\models \varphi \to \psi$, without a guarded interpolant (in any number of variables). That is, there does* not *exist a guarded formula ϑ built up from relation symbols which occur both in φ and ψ such that $\models \varphi \to \vartheta$ and $\models \vartheta \to \psi$.*

To see why this property fails for GF, it is useful to compare it to the interpolation property studied in modal logic. In modal logic, the interpolant is usually confined to proposition letters in the common language but may contain non-shared modalities. Strengthening the requirement on the common language to also include common modalities results in a much stronger interpolation property. [Ben99] shows this property for the basic multi-modal K. [Mar99] generalizes this result to Sahlqvist axiomatizable multi-modal logics whose axioms correspond to universal Horn frame conditions which do not specify any interaction between the different accessibility relations (e.g., bi-modal S5). When we have interaction, the stronger interpolation property is easily lost as the following example from [Ben99] shows. Consider the multi-modal logic defined by the axiom $\Diamond_1 p \to \Diamond_2 p$. This logic does not have the stronger interpolation property. For, in this logic $\Diamond_1 True \to \Diamond_2 True$ is a theorem whereas the only formulas in the common language (in the strong sense) are $True$ and $False$, which are obviously not interpolants. However, this logic does have the usual interpolation property (cf. [MV97, Corollary B.4.1]).

Thinking of guarded formulas as translations of modal formulas, we see that Theorem 1 formulates exactly this stronger version of interpolation, where 'common language' means the set of common relation symbols, which includes both the relations which are translated proposition letters and the relations that are obtained in translating the modalities. This suggests to consider an alternative interpolation property for GF that more closely resembles the one that is usually studied in modal logic. For this we will distinguish occurrences of relation symbols as guards from other occurrences.

Notation 2 For any guarded formula φ we understand by $\mathcal{L}_{G(\varphi)}$ the set of relations that occur in φ as the guard of some quantifier. $\dashv$

Note that the relations in $\mathcal{L}_{G(\varphi)}$ may at the same time occur in φ at non-guard positions. For example, in $\varphi = \exists x(Px \wedge \forall y(Sxy \to Py))$, the relation P occurs both as a guard and as a non-guard.

Definition 4 (Guarded $\mathcal{L}_1/\mathcal{L}_2$-bisimulation). Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$ be languages. A *guarded $\mathcal{L}_1/\mathcal{L}_2$-bisimulation* between models $\mathcal{M}$ and $\mathcal{N}$ is defined as a non-empty set of finite partial $\mathcal{L}_2$-isomorphisms between $\mathcal{M}, \mathcal{N}$ with zig- and zag- condition stipulated for $\mathcal{L}_1$-live sets only. $\dashv$

This type of bisimulation supports a characterization theorem for that fragment of GF in which only $\mathcal{L}_1$-predicates may be used as guards, but all predicates in $\mathcal{L}_1$ and $\mathcal{L}_2$ may occur at non-guard positions. Analogously to the characterization theorem for GF, the following characterization can be shown, using [HM99, Proposition 3.11].

Proposition 1. *Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$ be languages which contain at most k-ary relation symbols. A first order sentence φ is preserved under guarded $\mathcal{L}_1/\mathcal{L}_2$-bisimulations iff φ is logically equivalent to an $\mathcal{L}_2$-sentence $\psi \in \mathrm{GF}_k^k$ with $\mathcal{L}_{G(\psi)} \subseteq \mathcal{L}_1$.*

Notation 3 For models $\mathcal{M}, \mathcal{N}$, and $m_1, \ldots, m_k \in M$, $n_1, \ldots, n_k \in N$, we write $\mathcal{M}, m_1 \cdots m_k \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, n_1 \cdots n_k$, if $\mathcal{M} \models \theta[m_1, \ldots, m_k]$ iff $\mathcal{N} \models \theta[n_1, \ldots, n_k]$, for all $\mathcal{L}$-formulas $\theta \in \mathrm{GF}_k$. $\dashv$

Recall that for any formula ϑ, by $\mathcal{L}_\vartheta$ we denote the language consisting of all the relation symbols occurring in ϑ. The theorem below states that GF_k^k (and hence GF) has interpolation provided an interpolant for $\varphi \to \psi$ is allowed to contain relations in $\mathcal{L}_{G(\varphi)}$ and $\mathcal{L}_{G(\psi)}$ which are not necessarily in the common language. Modally speaking, an interpolant may use non-shared modalities.

Theorem 2 (GF_k^k has interpolation w.r.t. non-guard occurrences). *Let $k \in \omega$. For any $\varphi, \psi \in \mathrm{GF}_k^k$ such that $\models \varphi \to \psi$, there exists a $\vartheta \in \mathrm{GF}_k^k$ such that*

1. *$\mathcal{L}_\vartheta \subseteq (\mathcal{L}_\varphi \cap \mathcal{L}_\psi) \cup \mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}$, and*
2. *$\models \varphi \to \vartheta$ and $\models \vartheta \to \psi$.*

Proof of Theorem 2: We will show 'amalgamation via bisimulation' in the same spirit as e.g., the proof of interpolation for the basic modal logic K in [ABN98, Theorem 2.5]. Its main construction is a deviation of a fairly standard amalgamation method as can be found e.g., in [Mar95] and [Ném85].

For the course of this proof, let $k \in \omega$ be fixed but arbitrary. Consider $\varphi, \psi \in \mathrm{GF}_k^k$ such that $\models \varphi \to \psi$. For brevity, write $\mathcal{L}$ for $(\mathcal{L}_\varphi \cap \mathcal{L}_\psi) \cup \mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}$. Let $\Theta \stackrel{\text{def}}{=} \{\vartheta \in \mathrm{GF}_k \ : \ \mathcal{L}_\vartheta \subseteq \mathcal{L} \ \& \ \models \varphi \to \vartheta\}$. Our aim is to show that

Claim 1 *$\Theta \models \psi$.*

Note that in the formulation of this claim free variables play the role of individual constants. By compactness, it then follows that ψ is implied by some finite conjunction ϑ of formulas in Θ. Note that ϑ is again an $\mathcal{L}$-formula in GF_k^k. Hence ϑ is an interpolant for φ, ψ, and we are done.

To prove Claim 1, consider an arbitrary $\mathcal{L}_\psi \cup \mathcal{L}$-model $\mathcal{N}$, and $b_1, \ldots, b_k \in N$ such that $\mathcal{N} \models \vartheta[b_1, \ldots, b_k]$, for every $\vartheta \in \Theta$. Our task is to show that $\mathcal{N} \models \psi[b_1, \ldots, b_k]$.

We first note that there exists some $\mathcal{L}_\varphi \cup \mathcal{L}$-model $\mathcal{M}$ and $a_1, \ldots, a_k \in M$ such that $\mathcal{M} \models \varphi[a_1, \ldots, a_k]$ and $\mathcal{M}, a_1 \cdots a_k \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, b_1 \cdots b_k$. For, reasoning to contraposition, assume such $\mathcal{M}$, $a_1, \ldots, a_k$ do not exist. In that case, $\Phi \stackrel{\text{def}}{=} \{\vartheta \in \mathrm{GF}_k : \mathcal{L}_\vartheta \subseteq \mathcal{L} \;\&\; \mathcal{N} \models \vartheta[b_1, \ldots, b_k]\} \models \neg\varphi$. By compactness it follows that $\models \varphi \to \neg \bigwedge \Phi_0$, for some finite conjunction of formulas in Φ. Therefore $\neg \bigwedge \Phi_0 \in \Theta$, whence $\mathcal{N} \models \neg \bigwedge \Phi_0[b_1, \ldots, b_k]$. Quod non.

By passing to ω-saturated elementary extensions of $\mathcal{M}$ and $\mathcal{N}$, we may w.l.o.g. assume that $\mathcal{M}$, $\mathcal{N}$ are ω-saturated. As shown in the proof of the Characterization Theorem for GF in [ABN98, Theorem 4.2.2], the relation of guarded $\mathcal{L}$-equivalence between ω-saturated structures induces a guarded $\mathcal{L}$-bisimulation. The same is true for the relation of $\equiv_{\mathcal{L}}^{\mathrm{GF}_k}$ if $\mathcal{L}$ contains at most k-ary relations. Hence our assumption in particular implies the existence of a guarded $\mathcal{L}$-bisimulation between $\mathcal{M}, \mathcal{N}$ which links $\langle a_1, \ldots, a_k\rangle$ and $\langle b_1, \ldots, b_k\rangle$.

The aim of the rest of this proof is to amalgamate the models $\mathcal{M}$ and $\mathcal{N}$ in such a way that we can define guarded $\mathcal{L}_{G(\varphi)}/\mathcal{L}_\varphi$- (resp. $\mathcal{L}_{G(\psi)}/\mathcal{L}_\psi$-) bisimulations between the amalgamated model and $\mathcal{M}$ (resp. $\mathcal{N}$) which, when composed, will map $\langle a_1, \ldots, a_k\rangle$ to $\langle b_1, \ldots, b_k\rangle$. Chasing the resulting diagram and using the fact that $\varphi \models \psi$ will yield the desired conclusion that $\mathcal{N} \models \psi[b_1, \ldots, b_k]$. This will be made precise in the sequel.

We define a model over the set MN consisting of pairs $\langle m, n\rangle \in M \times N$ whose components cannot be distinguished by $\mathcal{L}$-formulas in GF_k. The interpretation of the predicates is read off coordinatewise. More precisely,

- $MN \stackrel{\text{def}}{=} \{\langle m, n\rangle \in M \times N \;:\; \mathcal{M}, m \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, n\}$.
- For l-ary $R \in \mathcal{L}_\varphi$, set $\langle\langle m_1, n_1\rangle, \ldots, \langle m_l, n_l\rangle\rangle \in I^{\mathcal{MN}}(R)$ iff
  - $\mathcal{M}, m_1 \cdots m_l \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, n_1 \cdots n_l$,
    (i.e., the m_i and n_i are not only pairwise equivalent but jointly so), and
  - $\langle m_1, \ldots, m_l\rangle \in I^{\mathcal{M}}(R)$.
- The interpretation of relations in $\mathcal{L}_\psi$ is defined similarly.

Note that the interpretation of relations in the common language is well-defined thanks to the requirement on live subsets of $\mathcal{MN}$ to be jointly $\mathcal{L}$-equivalent. The upshot of amalgamating our models into a product is that we can take projection functions as building blocks for the desired bisimulations. This is the purport of the following lemma, where π_i, $i = 1, 2$, denotes the projection function to the i-th coordinate. Define $F_{\pi_1} \stackrel{\text{def}}{=} \{\pi_1 : X \longrightarrow Y \;:\; X \subseteq^l_{\mathcal{L}_{G(\varphi)}} \mathcal{MN}$ or $X = \{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\}\}$, where the elements $a_1, \ldots, a_k \in M, b_1, \ldots, b_k \in N$ are the ones picked at the very beginning of this proof.

Lemma 1 (Amalgamation lemma). *The set* F_{π_1} *is a guarded* $\mathcal{L}_{G(\varphi)}/\mathcal{L}_{\varphi}$*-bisimulation between* $\mathcal{MN}$ *and* $\mathcal{M}$*. The analogously defined set* F_{π_2} *is a guarded* $\mathcal{L}_{G(\psi)}/\mathcal{L}_{\psi}$*-bisimulation between* $\mathcal{MN}$ *and* $\mathcal{N}$*.*

Before proving the lemma, let us first demonstrate its use and finish the proof of Claim 1. Recall that the model $\mathcal{M}$ and the sequence $a_1, \ldots, a_k \in M$ were chosen in such a way that $\mathcal{M} \models \varphi[a_1, \ldots, a_k]$. We took care to include $\pi_1 : \{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\} \longrightarrow \{a_1, \ldots, a_k\}$ in F_{π_1}. Since φ is invariant under guarded $\mathcal{L}_{G(\varphi)}/\mathcal{L}_{\varphi}$-bisimulations, it follows from the amalgamation lemma that $\mathcal{MN} \models \varphi[\langle a_1, b_k\rangle, \ldots, \langle a_k, b_k\rangle]$. By assumption then $\mathcal{MN} \models \psi[\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle]$. Since we included $\pi_2 : \{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\} \longrightarrow \{b_1, \ldots, b_k\}$ in F_{π_2}, the second part of the amalgamation lemma allows us to conclude that $\mathcal{N} \models \psi[b_1, \ldots, b_k]$.
Q.E.D. Claim 1.

Now we turn to the proof of the amalgamation lemma.

Proof of Lemma 1: We will prove the first part of the lemma concerning F_{π_1}. The second statement about F_{π_2} can be shown similarly.

F_{π_1} is obviously non-empty. Let $\pi_1 : X \longrightarrow Y \in F_{\pi_1}$. Then $X = \{x_1, \ldots, x_l\}$, for some $l \leq k$, and $\mathcal{M}, \pi_1(x_1) \cdots \pi_1(x_l) \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, \pi_2(x_1) \cdots \pi_2(x_l)$. By construction this implies that for any n-ary $R \in \mathcal{L}_{\varphi}$, and any $\langle x_{i_1}, \ldots, x_{i_n}\rangle \in X^n$ it is the case that $\langle x_{i_1}, \ldots, x_{i_n}\rangle \in I^{\mathcal{MN}}(R)$ iff $\langle \pi_1(x_{i_1}), \ldots, \pi_1(x_{i_n})\rangle \in I^{\mathcal{M}}(R)$. In other words, π_1 is a partial $\mathcal{L}_{\varphi}$-isomorphism.

For the zag-condition, consider $\pi_1 : X \longrightarrow Y \in F_{\pi_1}$, and $W \subseteq_R^l \mathcal{M}$, for some $R \in \mathcal{L}_{G(\varphi)}$. As above, $X = \{x_1, \ldots, x_l\}$, for some $l \leq k$, and $\mathcal{M}, \pi_1(x_1) \cdots \pi_1(x_l) \equiv_{\mathcal{L}}^{\mathrm{GF}_k} \mathcal{N}, \pi_2(x_1) \cdots \pi_2(x_l)$. Recall that the relation $\equiv_{\mathcal{L}}^{\mathrm{GF}_k}$ forms a guarded $\mathcal{L}$-bisimulation between $\mathcal{M}$ and $\mathcal{N}$. We saw that the partial map f from $\mathcal{M}$ to $\mathcal{N}$ which maps $\pi_1(x)$ to $\pi_2(x)$, for $x \in X$, is an element of this bisimulation. By the zig-condition, there exists a partial $\mathcal{L}$-isomorphism g in this bisimulation with domain W which agrees with f on the intersection of their domains. Let $W^* = \{\langle w, g(w)\rangle : w \in W\} \subseteq_R^l \mathcal{MN}$. Then W^* is the desired pre-image for W. As the zig-condition is trivially fulfilled, this completes the proof.
Q.E.D. Lemma 1.
Q.E.D. Theorem 2.

Corollary 1. GF *has interpolation w.r.t. non-guard occurrences.*

4 The Beth Theorem for the Guarded Fragment

In general, an important reason to investigate the interpolation property is that it can be seen as an intermediate stage in proving the Beth definability theorem. It will be shown that the limited form of interpolation expressed in Theorem 2 still serves this purpose for GF.

Let $\mathcal{L}_0$ be a language and R and R' distinct relation symbols of the same arity that are not in $\mathcal{L}_0$. Let $\mathcal{L} = \mathcal{L}_0 \cup \{R\}$. Let Σ be a set of guarded sentences in the language $\mathcal{L}$, and let Σ' denote the result of renaming R to R' in Σ.

Theorem 3 (Beth Theorem for GF_k^k). *Let $\mathcal{L}_0$, $\mathcal{L}$, R, R', Σ and Σ' be as above. Let $k \in \omega$ be such that $\Sigma \cup \{R\bar{v}\} \subseteq \mathrm{GF}_k^k$. If Σ* implicitly defines *R, i.e., if $\Sigma, \Sigma' \models \forall\bar{v}(R\bar{v} \leftrightarrow R'\bar{v})$, then there exists some $\varphi(\bar{v}) \in \mathrm{GF}_k^k$ in the language $\mathcal{L}_0$ such that $\Sigma \models \forall\bar{v}(R\bar{v} \leftrightarrow \varphi(\bar{v}))$. This formula φ is called an* explicit definition *for R relative to Σ.*

Proof of Theorem 3: Let all data be as in the theorem, and assume that

$$\Sigma, \Sigma' \models \forall\bar{v}(R\bar{v} \leftrightarrow R'\bar{v}). \tag{3}$$

We first show that (3) implies that any R-live set in a model for Σ is $\mathcal{L}_0$-live.

Claim 2 *Let $\mathcal{M}$ be a model of Σ, and let $Y \subseteq_R^l \mathcal{M}$. Then $Y \subseteq_{\mathcal{L}_0}^l \mathcal{M}$.*

Proof of Claim 2: Let $\mathcal{M}$ be a model of Σ, and let $Y_0 \subseteq_R^l \mathcal{M}$. Reasoning to contraposition, suppose $Y_0 \not\subseteq_{\mathcal{L}_0}^l \mathcal{M}$. We will derive a contradiction from this.

Let $\underline{2}$ denote the two-element universal $\mathcal{L}$-model with domain $\{0,1\}$. That is, $\bar{s} \in I^{\underline{2}}(P)$, for every l-ary $P \in \mathcal{L}$ and every $\bar{s} \in \{0,1\}^l$. Let $\mathcal{M}\times\underline{2}$ denote the usual product model. Writing π_1 for the projection on the first coordinate, this definition entails that $\bar{s} \in I^{\mathcal{M}\times\underline{2}}(P)$ iff $\langle \pi_1(s_1), \ldots, \pi_1(s_l)\rangle \in I^{\mathcal{M}}(P)$, for all $\bar{s} \in (M \times \{0,1\})^l$ and l-ary $P \in \mathcal{L}$. As the reader can easily verify, this in its turn implies that $F_1 \stackrel{\mathrm{def}}{=} \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq_{\mathcal{L}}^l \mathcal{M} \times \underline{2}\}$ is a guarded $\mathcal{L}$-bisimulation between $\mathcal{M}\times\underline{2}$ and $\mathcal{M}$. Since $\mathcal{M} \models \Sigma$, we conclude that $\mathcal{M}\times\underline{2} \models \Sigma$.

Our aim is to modify the interpretation of R on $\mathcal{M}\times\underline{2}$ in such a way that the resulting structure is again a model for Σ, contradicting the fact that Σ implicitly defines R. For this, we consider $X_0 \stackrel{\mathrm{def}}{=} Y_0 \times \{0\}$. Let $(\mathcal{M}\times\underline{2})'$ be the model which differs from $\mathcal{M}\times\underline{2}$ only in that $X_0 \notin I^{(\mathcal{M}\times\underline{2})'}(R)$. We claim that $F_1' \stackrel{\mathrm{def}}{=} \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq_{\mathcal{L}}^l (\mathcal{M}\times\underline{2})'\}$ is a guarded $\mathcal{L}$-bisimulation between $(\mathcal{M}\times\underline{2})'$ and $\mathcal{M}$.

F_1' is certainly not empty. Consider some $\pi_1 : X \longrightarrow Y$ in F_1'. If $X_0 \not\subseteq X$, then $\mathcal{L}$-relations are obviously preserved by π_1 in both ways. But we changed the interpretation of R such that X_0 is not R-live in $(\mathcal{M}\times\underline{2})'$. As Y_0 is not $\mathcal{L}_0$-live, it follows that X_0 is not $\mathcal{L}_0$-live in $(\mathcal{M}\times\underline{2})'$ either, and hence no superset of X_0 is the domain of some $\pi_1 \in F_1'$.

The zig-condition needs no comment. For the zag-condition, consider some $\pi_1 : X \longrightarrow Y$ in F_1', and $W \subseteq_{\mathcal{L}}^l \mathcal{M}$. If $W \subseteq Y$, the condition is trivially fulfilled. If not, then $\pi_1^{-1}[Y \cap W]$ can be extended in at least two ways to a set Z for which $\pi_1[Z] = W$. For $W \neq Y_0$, either one of these two extensions constitutes the domain of a projection in F_1' fulfilling the zag condition for π_1, W. For $W = Y_0$ any extension other than X_0 can be taken as such.

This shows that $(\mathcal{M}\times\underline{2})' \models \Sigma$. Summarizing, we see that $\mathcal{M}\times\underline{2} \models \Sigma$, $(\mathcal{M}\times\underline{2})' \models \Sigma$ but $I^{\mathcal{M}\times\underline{2}}(R) \neq I^{(\mathcal{M}\times\underline{2})'}(R)$. This contradicts the fact that Σ implicitly defines

R. We conclude that Y_0 is indeed $\mathcal{L}_0$-live, as was to be shown.
Q.E.D. Claim 2.

By compactness we may assume Σ to be a single sentence, and $\mathcal{L}_0$ finite. Assume R is l-ary, and let $\lambda_0(v_1, \ldots, v_l)$ be the canonical $\mathcal{L}_0$-formula in GF_k^k saying that the set $\{v_1, \ldots, v_l\}$ is $\mathcal{L}_0$-live (cf. page 6). For brevity, we write λ_0. For all $\mathcal{L}$-models $\mathcal{M}$ we define $R_0 \stackrel{\mathrm{def}}{=} \{\bar{m} \in M^l \ : \ \mathcal{M} \models R \wedge \lambda_0[\bar{m}]\}$. Hence for every $\mathcal{L}_0$-model $\mathcal{M}$, $\bar{m} \in M^l$, and every interpretation of R in $\mathcal{M}$: $(\mathcal{M}, R) \models R \wedge \lambda_0[\bar{m}]$ iff $(\mathcal{M}, R_0) \models R[\bar{m}]$. Note that by Claim 2, $I^{\mathcal{M}}(R) = R_0$, for models $\mathcal{M}$ of Σ. Let Σ_0 be the result of replacing in Σ any occurrence of R by the conjunction $R \wedge \lambda_0$. It is now straightforward to check the following:

(i) Σ_0 is preserved under guarded $\mathcal{L}_0/\mathcal{L}$-bisimulations. Hence, by Proposition 1, Σ_0 is equivalent to an $\mathcal{L}$-formula Γ in GF_k^k with $\mathcal{L}_{G(\Gamma)} \subseteq \mathcal{L}_0$.
(ii) $\Sigma \models \Sigma_0$, by Claim 2.
(iii) For every $\mathcal{L}_0$-model $\mathcal{M}$, and every interpretation of R in $\mathcal{M}$: if $(\mathcal{M}, R) \models \Sigma_0$, then $(\mathcal{M}, R_0) \models \Sigma$.

Let Σ_0' be the result of replacing R by R' in Σ_0. It follows from (3) and (iii) that

$$\Sigma_0 \wedge R \wedge \lambda_0 \models \Sigma_0' \rightarrow R'. \tag{4}$$

For, consider an $\mathcal{L}_0 \cup \{R, R'\}$-model $(\mathcal{M}, R, R')$ and some $\bar{m} \in M^l$ such that $(\mathcal{M}, R, R') \models (\Sigma_0 \wedge R \wedge \lambda_0 \wedge \Sigma_0')[\bar{m}]$. We have to show that $(\mathcal{M}, R, R') \models R'[\bar{m}]$. It follows from (iii) that $(\mathcal{M}, R_0, R_0') \models (\Sigma \wedge R \wedge \Sigma')[\bar{m}]$. By (3), then $(\mathcal{M}, R_0, R_0') \models R'[\bar{m}]$. Hence, certainly, $(\mathcal{M}, R, R') \models R'[\bar{m}]$.

Replace Σ_0 and Σ_0' in (4) by $\mathcal{L}_0$-guarded formulae Γ and Γ' according to (i). We then apply Theorem 2 to obtain, as an interpolant for (4), a formula ϑ in GF_k^k such that $\mathcal{L}_\vartheta \subseteq \mathcal{L}_0$ and $(\Sigma_0 \wedge R \wedge \lambda_0) \models \vartheta$ and $\vartheta \models \Sigma_0' \rightarrow R'$. Applying (ii) and Claim 2, we find that $\Sigma \wedge R \models \vartheta$. Renaming R' back into R in the second implication and one more appeal to (ii) gives us that $\vartheta \models \Sigma \rightarrow R$. Hence $\Sigma \models R \leftrightarrow \vartheta$ and ϑ provides the desired explicit definition of R relative to Σ.
Q.E.D. Theorem 3.

Corollary 2. GF *has the Beth definability theorem.*

Remark 2. Theorem 3 shows that the guarded finite variable fragments behave much nicer w.r.t. definability than the full finite variable fragments of first order logic (FO). For not only does the Beth theorem fail for any n-variable fragment of FO, $n \geq 2$, it fails drastically. In fact, even FO_2^2 (using the terminology FO_n^k for fragments of FO similar to our use of GF_n^k for guarded fragments) does not have the Beth property. For more information, the reader is referred to [Sai90] or [Hod93]. ⊣

References

[ABN98] H. Andréka, J. van Benthem, and I. Németi. Modal logics and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27(3):217–274, 1998.

[Ben99] J. van Benthem. Modal foundations for predicate logic. In E. Orlowska, editor, *Memorial Volume for Elena Rasiowa*, Studies in fuziness and soft computing, pages 39–55. Physica-Verlag, Heidelberg, New York, 1999.

[Bet53] E. W. Beth. On Padoa's method in the theory of definition. *Nederl. Akad. Wetensch. Proc. Ser. A.* **56** = *Indagationes Math.*, 15:330–339, 1953.

[GKV97] E. Grädel, P. Kolaitis, and M. Vardi. On the decision problem for two–variable first order logics. *Bulletin of Symbolic Logic*, 3:53–69, 1997.

[GO99] E. Grädel and M. Otto. On logics with two variables. *Theoretical Computer Science*, 1999. to appear.

[Grä97] E. Grädel. On the restraining power of guards. Technical report, RWTH Aachen, Lehrgebiet Mathematische Grundlagen der Informatik, 1997. To appear in the Journal of Symbolic Logic.

[Grä98] E. Grädel. Guarded fragments of first-order logic: a perspective for new description logics? In *Proc. of 1998 Int. Workshop on Description Logics DL '98, Trento*, CEUR Electronic Workshop Proceedings, 1998. Extended abstract, available at `http://sunsite.informatik.rwth-aachen.de/Publications/ CEUR-WS/Vol-11`.

[GW99] E. Grädel and I Walukiewicz. Guarded fixed point logic. In *Proc. 14th Symp. on Logic in Computer Science, LICS'99*, pages 45–54, 1999.

[Háj77] P. Hájek. Generalized quantifiers and finite sets. Ser. Konfer. No 1 14, Prace Nauk. Inst. Mat. Politech. Wroclaw, 1977.

[HM99] E. Hoogland and M. Marx. Interpolation in guarded fragments. Technical report, Institute for Logic, Language and Computation, University of Amsterdam, 1999.

[Hod93] I. Hodkinson. Finite variable logics. *Bull. Europ. Assoc. Theor. Comp. Sci.*, 51:111–140, 1993. Addendum in vol. 52. Also available at `http://www.doc.ic.ac.uk/~imh/papers/yuri.html`.

[Mar95] Maarten Marx. *Algebraic Relativization and Arrow Logic*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1995.

[Mar99] M. Marx. Interpolation in (fibered) modal logic. In A. Haeberer, editor, *Proc. of AMAST 1998, Amazonia–Manaus, Brazil, 4–8 January 1999*, 1999.

[MV97] M. Marx and Y. Venema. *Multi-dimensional Modal Logic*. Applied Logic Series. Kluwer Academic Publishers, 1997.

[Ném85] I. Németi. Cylindric relativised set algebras have strong amalgamation. *Journal of Symbolic Logic*, 50(3):689–700, 1985.

[Sai90] I. Sain. Beth's and Craig's properties via epimorphisms and amalgamation in algebraic logic. In C. Bergman, R. Maddux, and D. Pigozzi, editors, *Algebraic logic and universal algebra in computer science*, volume 24 of *Lecture Notes in Computer Science*, pages 209–226. Springer Verlag, Berlin, Heidelberg, New York, 1990.

[Sch91] K. Schild. A correspondence theory for terminological logics. In *Proceedings of the 12th IJCAI*, pages 466–471, 1991.

[Var98] M. Vardi. Why is modal logic so robustly decidable? In *Descriptive Complexity and Finite Models: Proceedings of a DIMACS Workshop*, volume 31 of *Series in Discrete Mathematics and Theoretical Computer Science*, pages 149–184. American Mathematical Society, 1998.