On the Expressive Power of Variable-Confined Logics

Phokion G. Kolaitis*
University of California, Santa Cruz

Moshe Y. Vardi†
Rice University

Abstract

Finite-model theory is the study of first-order logic (FO) and its extensions on classes of finite structures. Least-fixpoint logic (LFP) and finite-variable infinitary logic ($\mathcal{L}^{\omega}_{\infty\omega}$) have turned out to be of particular importance in finite-model theory. In his 1986 Ph.D. dissertation, McColm formulated two conjectures concerning the interplay between FO, LFP, and $\mathcal{L}^{\omega}_{\infty\omega}$ on arbitrary classes of finite structures. In a nutshell, these conjectures assert that boundedness of LFP is the key to the collapse of $\mathcal{L}^{\omega}_{\infty\omega}$ and LFP to FO. More specifically, McColm's First Conjecture asserts that LFP has the same expressive power as FO on a class $\mathcal{C}$ of finite structures if and only if LFP is bounded on $\mathcal{C}$. McColm's Second Conjecture asserts that $\mathcal{L}^{\omega}_{\infty\omega}$ has the same expressive power as FO on $\mathcal{C}$ if and only if LFP is bounded on $\mathcal{C}$. The second conjecture was confirmed by Kolaitis and Vardi in 1992 and the first conjecture was refuted by Gurevich, Immerman, and Shelah in 1994. Both results use a finite but arbitrarily large number of variables.

In this paper we study the comparative expressive power of *variable-confined* logics, that is, logics with a fixed finite number of variables. This is motivated by the fact that the number of variables is considered a logical resource in descriptive complexity theory. We consider the expressive power of the logics FO^k (first-order logic with k variables), LFP^k (LFP with k variables, appropriately defined), and $\mathcal{L}^{k}_{\infty\omega}$ (infinitary logic with k variables) over classes of finite structures. While the definitions of FO^k and $\mathcal{L}^{k}_{\infty\omega}$ are quite clear, it turns out that ramifying LFP is a more delicate matter. We define LFP^k in terms of *systems of least fixpoints*, i.e., instead of taking the least fixpoint of a single positive first-order formula, we consider simultaneous least fixpoints of a vector of positive first-order formulas. As evidence that LFP^k, $k \geq 1$, is the right ramification of LFP we offer two main results. The first is a new proof of a theorem by Dawar, Lindell, and Weinstein to the effect that equivalence classes of finite structures with respect to the logic $\mathcal{L}^{k}_{\infty\omega}$ are expressible in FO^k. The second result, novel and technically difficult, is a characterization for each $k \geq 1$ of the collapse of $\mathcal{L}^{k}_{\infty\omega}$ to FO^k in terms of boundedness of LFP^k. More precisely, we establish the following stronger version of McColm's Second Conjecture: $\mathcal{L}^{k}_{\infty\omega} = \mathrm{FO}^k$ on a class $\mathcal{C}$ of finite structures if and only if LFP^k is *uniformly bounded* on $\mathcal{C}$.

1 Introduction

One of the distinguishing characteristics of finite-model theory is the fact that first-order logic has severely limited expressive power on finite structures (cf. [Fag75, AU79, Gai82]). This is the main reason why in finite-model theory the emphasis is on logics that go beyond first-order logic (FO). Adding fixpoint constructs is a natural way to increase the limited expressive power of FO. Such extensions have been the subject of extensive study – focusing on their expressive power, their relationship to complexity classes, in particular to PTIME and PSPACE, and their asymptotic probability properties (cf. [AV89, AV95, Cha88, CH82, Imm86, KV87, Var82]). Of particular importance among these extensions is *least-fixpoint logic* (LFP), which is obtained from FO by adding least fixpoints of positive first-order formulas.

It is by now widely recognized that it is extremely useful to view fixpoint logics as effective fragments of $\mathcal{L}^{\omega}_{\infty\omega}$, *finite-variable infinitary logic* [KV92c, Hod93].

*Department of Computer Science, UCSC, Santa Cruz, CA 95064; e-mail: `kolaitis@cse.ucsc.edu`. During the preparation of this paper this author was partially supported by NSF Grant CCR-9307758.

†Department of Computer Science, Rice University, Houston, TX 77005-1892; e-mail: `vardi@cs.rice.edu`; URL: `http://www.cs.rice.edu/~vardi`

More formally, $\mathcal{L}^{\omega}_{\infty\omega} = \bigcup_{k=1}^{\infty} \mathcal{L}^{k}_{\infty\omega}$, where $\mathcal{L}^{k}_{\infty\omega}$ is the collection of all formulas that may have infinite conjunctions and disjunctions, but at most k distinct variables. Although the infinitary logic $\mathcal{L}^{\omega}_{\infty\omega}$ was studied originally on infinite structures [Bar77], it turned out to have numerous uses in finite-model theory. Since $\mathcal{L}^{\omega}_{\infty\omega}$ constitutes a proper extension of fixpoint logics [KV92b], it can be used to derive both positive and negative results about fixpoint logics. On the positive side, structural results about $\mathcal{L}^{\omega}_{\infty\omega}$, such as the 0-1 law, transfer to similar results for fixpoint logics, while, on the negative side, lower-bound results for expressibility in $\mathcal{L}^{\omega}_{\infty\omega}$ yield, a fortiori, lower bounds for expressibility in fixpoint logics.

An issue of fundamental importance is the comparison between the expressive power of $\mathcal{L}^{\omega}_{\infty\omega}$, LFP, and FO over classes of finite structures. This issue is intimately related to open questions in complexity theory [KV92a, DH94]. It is known that on the class of all finite structures $\mathcal{L}^{\omega}_{\infty\omega}$ is strictly more expressive than LFP, which in turn is strictly more expressive than FO [CH82, KV92b]. In practice, however, one is often concerned with a restricted class of finite structures, such as the class of of planar graphs or the class of databases satisfying a constraint. Thus, there is a solid motivation to understand the relationship between $\mathcal{L}^{\omega}_{\infty\omega}$, LFP, and FO, on specific classes of finite structures. Note that these logics may have identical expressive power over some classes of finite structures. This holds, for example, on the class of finite equivalence relations, where both LFP and $\mathcal{L}^{\omega}_{\infty\omega}$ collapse to FO [Kol85].

Under what conditions does LFP or $\mathcal{L}^{\omega}_{\infty\omega}$ collapse to FO over a given class of finite structures? This problem was first raised and studied by McColm [McC86, McC90b] who formulated two conjectures asserting that the collapse of LFP and $\mathcal{L}^{\omega}_{\infty\omega}$ down to FO is related to a certain *boundedness* condition on LFP. Intuitively, LFP is *bounded* on a class $\mathcal{C}$ of finite structures if the "bottom-up" iteration of every positive first-order operator converges on $\mathcal{C}$ to its least fixpoint in a data independent fashion, i.e., the number of required iterations to compute the fixpoint is independent of the structures. McColm conjectured that boundedness of LFP is the key to both the collapse of LFP to FO and the collapse of $\mathcal{L}^{\omega}_{\infty\omega}$ to FO. More precisely, McColm's First Conjecture asserts that LFP has the same expressive power as FO on $\mathcal{C}$ if and only if LFP is bounded on $\mathcal{C}$. McColm's Second Conjecture asserts that $\mathcal{L}^{\omega}_{\infty\omega}$ has the same expressive power as FO on $\mathcal{C}$ if and only if LFP is bounded on $\mathcal{C}$. Note that the two conjectures combined together assert that LFP collapses to FO on a class $\mathcal{C}$ of finite structures if and only if $\mathcal{L}^{\omega}_{\infty\omega}$ collapses to FO on $\mathcal{C}$.

McColm's conjectures were subsequently studied in [KV92a, GIS94], where the second conjecture was confirmed and the first conjecture was refuted. First, in [KV92a] it was shown that if LFP is bounded on $\mathcal{C}$, then $\mathcal{L}^{\omega}_{\infty\omega}$ collapses to FO on $\mathcal{C}$. Then, in [GIS94] it was shown that there is a class $\mathcal{C}$ on which LFP does collapse to FO though LFP is not bounded on $\mathcal{C}$.

The results in both [KV92a] and [GIS94] take advantage of the fact that we are considering logics whose formulas have a *finite* but *arbitrarily large* number of variables. Indeed, boundedness of LFP with $2k$ variables is used in [KV92a] to show that $\mathcal{L}^{k}_{\infty\omega}$ collapses to FO. Furthermore, on the class $\mathcal{C}$ constructed in [GIS94] LFP with k variables collapses to FO with m variables, where m is a number that is exponentially bigger than k. It is an important realization in descriptive complexity theory that variables can be viewed as a logical "resource", which is closely related to the computational resource of space [Imm81, Imm91]. The importance of imposing bounds on the number of variables also emerged from the study of database query evaluation [Var95]. This motivates studying the comparative expressive power of *variable-confined* logics, that is to say, logics with a fixed finite number of variable. In this paper, we investigate the expressive power of the logics FO^k (first-order logic with k variables), LFP^k (LFP with k variables, appropriately defined), and $\mathcal{L}^{k}_{\infty\omega}$ (infinitary logic with k variables) over classes of finite structures. The study of variable-confined logics is also motivated by the fact that most results about $\mathcal{L}^{\omega}_{\infty\omega}$ are obtained from results about $\mathcal{L}^{k}_{\infty\omega}$ for $k \geq 1$, e.g., the 0-1 law for $\mathcal{L}^{\omega}_{\infty\omega}$ is obtained by proving 0-1 laws for $\mathcal{L}^{k}_{\infty\omega}$, $k \geq 1$. This happens because the expressive power of $\mathcal{L}^{\omega}_{\infty\omega}$ can be characterized in terms of k-pebble games for $\mathcal{L}^{k}_{\infty\omega}$, $k \geq 1$ [Bar77, Imm82, KV92b].

In order to study the expressive power of variable-confined LFP, a precise definition of LFP^k is needed. The definition should be such that on one hand we have that $\mathrm{LFP} = \bigcup_{k\geq 1} \mathrm{LFP}^k$ and on the other hand the collapse of $\mathcal{L}^k_{\infty\omega}$ to FO^k is related to a boundedness condition on LFP^k. The obvious definition for LFP^k is the closure of FO^k under positive fixpoints. As noted earlier, however, the collapse of $\mathcal{L}^k_{\infty\omega}$ to FO^k is then related to boundedness of LFP^{2k} rather than LFP^k.

In this paper, we argue that the correct way to define LFP^k is in terms of *systems* of fixpoints. Specifically, instead of taking the fixpoint of a single positive first-order formula, we consider *simultaneous* fixpoints of a vector of positive first-order formulas with k variables (we provide a precise definition later on). It is well known that this extension of FO coincides with LFP; moreover, systems of positive fixpoints provide a *normal form* for LFP [Mos74, Imm86]. We take LFP^k to be the fragment of LFP defined by systems of positive FO^k formulas (this is essentially the same as the logic IND^k defined in [McC90a]). It follows from the normal-form theorem that $\mathrm{LFP} = \bigcup_{k\geq 1} \mathrm{LFP}^k$. At first it may seem that LFP^k is not a variable-confined logic, since a system may consist of any finite number of formulas. Nevertheless, we show that this is not the case by proving that LFP^k can be viewed as a fragment of $\mathcal{L}^k_{\infty\omega}$.

We offer two main results as evidence that LFP^k is the right ramification of LFP. The first result is concerned with the definability of $\equiv^k_{\infty\omega}$-equivalence classes. Two structures $\mathbf{A}$ and $\mathbf{B}$ are $\mathcal{L}^k_{\infty\omega}$-equivalent, denoted $\mathbf{A} \equiv^k_{\infty\omega} \mathbf{B}$, if they agree on all sentences of $\mathcal{L}^k_{\infty\omega}$. The study of $\equiv^k_{\infty\omega}$-equivalence classes has significant applications to the study of the expressive power of finite-variables logics [KV92a, DLW95] with intriguing connections to complexity theory [AVV92, AV95, DH94]. In [DLW95] it was shown that every $\equiv^k_{\infty\omega}$-equivalence class is definable in FO^k; their proof is an adaptation of Scott's theorem to the effect that every countable structure can be described up to isomorphism by a formula of the infinitary logic $\mathcal{L}_{\omega_1\omega}$ [Sco65]. We give an alternative proof of this result that follows easily from a description of k-pebble games in LFP^k.

The second result, novel and technically difficult, is a characterization of the collapse of $\mathcal{L}^k_{\infty\omega}$ to FO^k in terms of boundedness of LFP^k. It should be pointed out that care has to be exercised in defining boundedness for systems of fixpoints. When a vector of first-order operations is iterated, not all the components need to change in each iteration. We define boundedness of LFP^k in terms of the maximum number of changes that each component of the system goes through before the least fixpoint of the system is reached. Using this definition of boundedness, we prove a ramified version of McColm's Second Conjecture: $\mathcal{L}^k_{\infty\omega} = \mathrm{FO}^k$ on a class $\mathcal{C}$ if and only if FP^k is *uniformly bounded* on $\mathcal{C}$, i.e., the bound is independent of the structures and the formulas. The proof involves a delicate combinatorial argument and is significantly more involved than that of McColm's Second Conjecture in [KV92a]. Note that our earlier confirmation of McColm's Second Conjecture can be derived from this new ramified version.

Since the refutation in [GIS94] of McColm's First Conjecture does not ramify, the above ramified result opens the possibility of revisiting McColm's First Conjecture and examining whether it holds in its ramified version, i.e., is it the case that LFP^k has the same expressive power as FO^k on a class $\mathcal{C}$ of finite structures if and only if LFP^k is uniformly bounded on $\mathcal{C}$? We conclude with some partial results on this problem.

2 Least-Fixpoint Logic and Infinitary Logic

In this paper we assume that all structures are finite structures over relational vocabularies.

Let k be a positive integer. A *k-ary query Q on* a vocabulary σ is a mapping on the class of all σ-structures such that if $\mathbf{A}$ is a σ-structure, then $Q(\mathbf{A})$ is a k-ary relation on A that is preserved under isomorphisms. A *Boolean query on σ* is a mapping Q from σ-structures to $\{0, 1\}$ such that if $\mathbf{A}$ and $\mathbf{B}$ are isomorphic, then $Q(\mathbf{A}) = Q(\mathbf{B})$. Equivalently, a Boolean query can be identified with a class of σ-structures that is closed under isomorphisms. Let $\mathcal{L}$ be a logic, k a positive integer, Q a k-ary query on σ, and $\varphi(x_1, \ldots, x_k)$ a formula of $\mathcal{L}$ with free variables among $x_1, \ldots, x_k$. We say that the formula $\varphi(x_1, \ldots, x_k)$ *defines the query Q* if for every σ-structure $\mathbf{A}$ and every sequence $(a_1, \ldots, a_k)$ from the universe A of $\mathbf{A}$ we have that $(a_1, \ldots a_k) \in Q(\mathbf{A})$ iff $\mathbf{A} \models \varphi(x_1/a_1, \ldots, x_k/a_k)$. The above definitions can be relativized to an arbitrary class $\mathcal{C}$ of finite structures that is closed under isomorphisms.

The infinitary logic $\mathcal{L}_{\infty\omega}$ is the extension of first-order logic that results by allowing infinite disjunctions and conjunctions in the syntax, while keeping the quantifier depth finite (cf. [BF85]). The *k-variable* *infinitary logic* $\mathcal{L}^k_{\infty\omega}$ consists of all formulas of $\mathcal{L}_{\infty\omega}$ with at most k distinct variables. The *finite-variable infinitary logic* $\mathcal{L}^\omega_{\infty\omega}$ is the union of all k-variable logics, i.e., $\mathcal{L}^\omega_{\infty\omega} = \bigcup_{k=1}^{\infty} \mathcal{L}^k_{\infty\omega}$. Note that each $\mathcal{L}^k_{\infty\omega}$, $k \geq 1$, is a variable-confined logic, while $\mathcal{L}^\omega_{\infty\omega}$ is not, as $\mathcal{L}^\omega_{\infty\omega}$-formulas may have an arbitrarily large finite number of variables. Finally, FO^k is the collection of all first-order formulas with at most k distinct variables.

The expressive power of the logics $\mathcal{L}^k_{\infty\omega}$, $k \geq 1$, is usually illustrated by the fact that for any $n \geq 1$ the property $p_n(x, y)$ asserting that "there is a path of length n from x to y" is expressible by a formula of $\mathcal{L}^3_{\omega\omega}$. Indeed, put $p_1(x, y) \equiv E(x, y)$ and assume, by induction on n, that $p_{n-1}(x, y)$ is equivalent to a formula of FO^3. Then the desired formula $p_n(x, y)$ is $(\exists z)[E(x, z) \wedge (\exists x)(x = z \wedge p_{n-1}(x, y))]$. As a result, the formula $(\forall x)(\forall y)(\bigvee_{n=1}^{\infty} p_n(x, y))$ of $\mathcal{L}^3_{\infty\omega}$ defines the *connectivity* query on graphs.

Let $\varphi(x_1, \ldots, x_n, T)$ be a first-order formula in which T is a n-ary relation symbol not included in σ. If $\mathbf{A}$ is a σ-structure, then φ gives rise to an operator Φ from n-ary relations R on the universe A of $\mathbf{A}$ to n-ary relations on A, where $\Phi(R) = \{(a_1, \ldots, a_n) : \mathbf{A} \models \varphi(x_1/a_1, \ldots, x_n/a_n, T/R)\}$. Such an operator Φ can be iterated and, thus, generates the sequence of *stages* Φ^m, $m \geq 0$, defined by the induction: $\Phi^0 = \emptyset$, $\Phi^{m+1} = \Phi(\Phi^m)$.

If $\varphi(x_1, \ldots, x_n, T)$ is *positive in* T (which means that every occurrence of T in φ is within an even number of negations), then Φ is *monotone*, i.e., $\Phi(R_1) \subseteq \Phi(R_2)$, whenever $R_1 \subseteq R_2$. In this case, the sequence of stages is increasing and, thus, for every σ-structure $\mathbf{A}$ there is an integer $m_0 \leq |A|^n$, such that $\Phi^{m_0} = \Phi^m$ for every $m \geq m_0$. Moreover, Φ^{m_0} is the *least fixpoint of* Φ on $\mathbf{A}$, i.e., it is the smallest n-ary relation T on A such that $\Phi(T) = T$. We write $\varphi^\infty(x_1, \ldots, x_n)$ to denote the least fixpoint of Φ and refer to it as the *least fixpoint* of the formula $\varphi(x_1, \ldots, x_n, T)$.

Least-fixpoint logic (LFP) is the extension of first-order logic that contains all first-order formulas and is closed under the operations of first-order logic and the following *least-fixpoint rule:* If $\varphi(x_1, \ldots, x_n, T)$ is a positive in T first-order formula, then $\varphi^\infty(x_1, \ldots, x_n)$ is also a formula of LFP. The least fixpoint $\varphi^\infty(x, y)$ of $E(x, y) \vee (\exists z)(T(x, z) \wedge T(z, y))$ provides the canonical example of a LFP-formula, since it defines the *transitive closure* of the *edge* relation E in a graph $G = (V, E)$. LFP enjoys a close relationship with polynomial-time computability. First, we have that LFP $\subseteq$ PTIME. On the class of all σ-structures this containment is proper, since LFP can not express *counting* queries, such as *even cardinality* [CH82]. In contrast, Immerman [Imm86] and Vardi [Var82] showed that if $\mathcal{C}$ is a class of *ordered* σ-structures, then LFP = PTIME on $\mathcal{C}$.

LFP embodies recursion through the iteration of single positive first-order formulas. One can also consider a richer recursion mechanism that embodies mutual recursion through the simultaneous iteration of a finite sequence of positive first-order formulas (cf. [Mos74]). A *system* of first-order formulas is a finite sequence

$$\begin{gathered}\varphi_1(x_1, \ldots, x_{n_1}, T_1, \ldots, T_l) \\ \vdots \\ \varphi_l(x_1, \ldots, x_{n_l}, T_1, \ldots, T_l)\end{gathered}$$

of first-order formulas such that each T_i is a relation symbol of arity n_i that is not in σ. If $\mathbf{A}$ is a σ-structure, then every system gives rise to an operator Φ from sequences $(R_1, \ldots, R_l)$ of relations R_i of arity n_i on the universe A of $\mathbf{A}$ to sequences of relations on A of the same arities. More precisely, $\Phi(R_1, \ldots, R_l) = (\Phi_1(R_1, \ldots, R_l), \ldots, \Phi_l(R_1, \ldots, R_l))$, where, for every $1 \leq i \leq l$, $\Phi_i(R_1, \ldots, R_l)$ is defined as

$$\begin{aligned}\{(a_1, \ldots, a_{n_i}) : \mathbf{A} \models\ & \varphi_i(x_1/a_1, \ldots, x_{n_i}/a_{n_i}, \\ & T_1/R_1, \ldots, T_l/R_l)\}.\end{aligned}$$

The *stages* $\Phi^m = (\Phi_1^m, \ldots, \Phi_l^m)$, $m \geq 0$, of Φ on a structure $\mathbf{A}$ are defined by the following induction on m simultaneously for all $i \leq l$:

$$\Phi_i^0 = (\emptyset, \ldots, \emptyset),\ \Phi_i^{m+1} = \Phi_i(\Phi_1^m, \ldots, \Phi_l^m).$$

If each formula $\varphi_i(x_1, \ldots, x_{n_i}, T_1, \ldots, T_l)$, $1 \leq i \leq l$, of a system is positive in every T_i, then Φ is monotone in each of its arguments and, as a result, the sequence of its stages is increasing in each component. Thus, for every σ-structure $\mathbf{A}$ there is a positive integer m_0 such that $\Phi^m = \Phi^{m_0}$ for every $m \geq m_0$. Moreover, the sequence $\Phi^{m_0} = (\Phi_1^{m_0}, \ldots, \Phi_l^{m_0})$ is the *least fixpoint of* Φ on $\mathbf{A}$, i.e., the smallest (with respect to componentwise containment) sequence $(R_1, \ldots, R_l)$ of relations on A such that $\Phi(R_1, \ldots, R_l) = (R_1, \ldots, R_l)$. We call this sequence the *least fixpoint of the system* $\varphi_1, \ldots, \varphi_l$ and denote it by $(\varphi_1^\infty, \ldots, \varphi_l^\infty)$. Among the components of the least fixpoint, there is a distinguished one in which we are interested. As an example, let

$$\begin{array}{rcl} \varphi_1(x, y, T_1, T_2) & \equiv & E(x, y) \vee \\ & & (\exists z)(T_1(x, z) \wedge T_1(z, y)), \\ \varphi_2(x, T_1, T_2) & \equiv & T_1(x, x). \end{array}$$

Then $\varphi_1^\infty(x, y)$ defines the *transitive closure* query, while $\varphi_2^\infty(x)$ defines the *cycle* query, which returns all nodes a such that there is a path from a to a.

The extension of first-order logic with least fixpoints of systems of positive first-order formulas has the same expressive power as LFP [Mos74, Imm86]. Indeed, every system of first-order formulas can be simulated by a single formula $\varphi(x_1, \ldots, x_q, T)$ in which T is a relation symbol of sufficiently large arity r, so that each component of the least fixpoint of the system can be defined from the least fixpoint $\varphi^\infty(x_1, \ldots, x_r)$ using first-order operations. Moreover, systems of positive first-order formulas provide a *normal form* for LFP, i.e., every LFP-definable query is also definable by the component φ_l^∞ of the least fixpoint of some system $(\varphi_1, \ldots, \varphi_l)$ of positive first-order formulas. In view of the above, from now on we allow least fixpoints of systems of positive first-order formulas in the syntax of LFP. Using this extended syntax, LFP can be ramified into a hierarchy of levels based on the number of variables occurring in systems (cf. [McC90a]).

Definition 2.1: Let σ be a vocabulary and k a positive integer. LFP^k is the fragment of fixpoint logic LFP having as formulas the components φ_i^∞, $1 \leq i \leq l$, of the least fixpoints of systems $\varphi_1, \ldots, \varphi_l$ of positive FO^k-formulas over σ, i.e., each φ_i is a first-order formula over σ whose total number of variables is at most k. ▮

As an example, the cycle query is in LFP^3. If $\varphi_1, \ldots, \varphi_l$ is a system of positive FO^k-formulas, then we can assume without loss of generality that the variables of each formula are among the variables $x_1, \ldots, x_k$; thus, the total number of variables occurring in all formulas of the system is at most k. Note also that LFP $= \bigcup_{k=1}^\infty$ LFP^k, since systems yield a normal form for LFP.

It is well known (cf. [KV92b]) that $\mathcal{L}^\omega_{\infty\omega}$ subsumes LFP on finite structures. Moreover, as shown in [KV92b], if $\varphi(x_1, \ldots, x_n, T)$ is a positive FO^k-formula, then the least fixpoint $\varphi^\infty(x_1, \ldots, x_n)$ is definable by a formula of $\mathcal{L}^k_{\infty\omega}$. Our first result in the present paper extends this theorem to systems of positive FO^k-formulas, which implies that $\mathcal{L}^k_{\infty\omega}$ subsumes LFP^k. It should be pointed out that the standard simulation of systems by single formulas in [Mos74, Imm86] increases the number of variables, so the subsumption of LFP^k by $\mathcal{L}^k_{\infty\omega}$ does not follow from the results in [Mos74, Imm86] and [KV92b]. Instead, we provide a more refined simulation based on the following result.

Lemma 2.2: *Let k be a positive integer, let $\pi : \{1, \ldots, k\} \mapsto \{1, \ldots, k\}$ be a function, and let Q be a k-ary query on σ. If Q is* FO^k*-definable, then the query Q_π is also* FO^k*-definable, where*

$$Q_\pi(\mathbf{A}) = \{(a_1, \ldots, a_k) : (a_{\pi(1)}, \ldots, a_{\pi(k)}) \in Q(\mathbf{A})\}.$$

The above Lemma 2.2 is proved by induction on the construction of FO^k-formulas simultaneously for all functions π. The proof of the next theorem involves repeated applications of Lemma 2.2 and a subtle induction on m simultaneously for all components of Φ. In what follows, by *quantifier rank of a system* $\varphi_1, \ldots, \varphi_l$ we mean the $\max\{\mathrm{qr}(\varphi_1), \ldots, \mathrm{qr}(\varphi_l)\}$, where $\mathrm{qr}(\varphi_i)$ is the quantifier rank of φ_i, that is, the depth of quantifier nesting in φ_i, $1 \leq i \leq l$.

Theorem 2.3: *Let k be a positive integer and let $\varphi_1, \ldots, \varphi_l$, $1 \leq i \leq l$, be a system of positive* FO^k*-formulas of quantifier rank d. Then the following are true for the above system.*

- *Each component Φ_i^m of every stage $\Phi^m = (\Phi_1^m, \ldots, \Phi_l^m)$, $m \geq 0$, of the operator Φ is definable by an* FO^k*-formula of quantifier rank at most dm.*
- *Each component φ_i^∞ of the least fixpoint $(\varphi_1^\infty, \ldots, \varphi_l^\infty)$ of the system is $\mathcal{L}^k_{\infty\omega}$-definable. As a result, we have that* LFP$^k \subseteq \mathcal{L}^k_{\infty\omega}$.

3 Definability of $\mathcal{L}^k_{\infty\omega}$-Equivalence

The model theory of a logic $\mathcal{L}$ can not be developed without analyzing the concept of *$\mathcal{L}$-equivalence*. In recent years, the study of $\mathcal{L}^k_{\infty\omega}$-equivalence has been the focus of intensive investigation [KV92b, Daw93, DLW95].

Let k be a positive integer and let **A**, **B** be two σ-structures.

• **A** is $\mathcal{L}^k_{\infty\omega}$*-equivalent* to **B**, denoted by $\mathbf{A} \equiv^k_{\infty\omega} \mathbf{B}$, if **A** and **B** satisfy the same $\mathcal{L}^k_{\infty\omega}$-sentences.

• Let **a** and **b** be two k-tuples from **A** and **B**. We say that $(\mathbf{A}, \mathbf{a})$ is $\mathcal{L}^k_{\infty\omega}$*-equivalent* to $(\mathbf{B}, \mathbf{b})$, and write $(\mathbf{A}, \mathbf{a}) \equiv^k_{\infty\omega} (\mathbf{B}, \mathbf{b})$, if $(\mathbf{A}, \mathbf{a})$ and $(\mathbf{B}, \mathbf{b})$ satisfy the same $\mathcal{L}^k_{\infty\omega}$-formulas.

The analysis of $\mathcal{L}^k_{\infty\omega}$-equivalence is derived from the fact that the relation $\mathbf{A} \equiv^k_{\infty\omega} \mathbf{B}$ can be characterized in terms of a k*-pebble game* ([Bar77, Imm82]), a variant of which we describe next. Let $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$ be two k-tuples from **A** and **B**. The k*-pebble game* between the *Spoiler* and the *Duplicator* on $(\mathbf{A}, a_1, \ldots, a_k)$ and $(\mathbf{B}, b_1, \ldots, b_k)$ has the following rules: There are k white pebbles $W_1, \ldots, W_k$ and k black pebbles $B_1, \ldots, B_k$. Throughout the game, the white pebbles are placed on elements of **A** and the black pebbles on elements of **B**. Initially, pebble W_j is placed on a_j and pebble B_j is placed on b_j, $1 \leq j \leq k$. In each round of the game, the Spoiler chooses one of the two colors and moves a pebble of that color. The Duplicator responds by moving the pebble of the other color that has the same index as the pebble moved by the Spoiler.

Let c_j and d_j be the elements of **A** and **B** pebbled by W_j and B_j, $1 \leq j \leq k$, at the end of some round of the game. The Spoiler wins the game at this point if the correspondence $c_j \mapsto d_j$, $1 \leq j \leq k$, fails to be an isomorphism between the substructures of **A** and **B** with universes $\{c_1, \ldots, c_k\}$ and $\{d_1, \ldots, d_k\}$. Otherwise, the game goes on. The Duplicator *wins the* k*-pebble game* if he can continue playing "forever", i.e., the Spoiler can never win a round of the game.

The crucial connection between k-pebble games and $\mathcal{L}^k_{\infty\omega}$-equivalence is provided by the following result due to Barwise [Bar77] (cf. [Imm82]). A detailed proof can be found in [KV92b].

Theorem 3.1: *Let* **A** *and* **B** *be two* σ*-structures, and let* $(a_1, \ldots, a_k)$ *and* $(b_1, \ldots, b_k)$ *be two* k*-tuples from the universes of* **A** *and* **B**, *respectively. Then the following are equivalent:*

• $(\mathbf{A}, a_1, \ldots, a_k) \equiv^k_{\infty\omega} (\mathbf{B}, b_1, \ldots, b_k)$.

• *The Duplicator wins the* k*-pebble game on* $(\mathbf{A}, a_1, \ldots, a_k)$ *and* $(\mathbf{B}, b_1, \ldots, b_k)$.

If $\mathcal{L}$ is an arbitrary logic, then there is no a priori reason for $\mathcal{L}$-equivalence to be definable in $\mathcal{L}$ itself. In particular, it it not hard to show that neither FO nor $\mathcal{L}^\omega_{\infty\omega}$ can express their own equivalence. On the other hand, as observed in [KV92a], Theorem 3.1 makes it possible to show that $\mathcal{L}^k_{\infty\omega}$-equivalence is expressible in LFP. In fact, for every $k \geq 1$ there is a positive LFP^{2k}-formula such that its least fixpoint defines query: "Is $(\mathbf{A}, a_1, \ldots, a_k) \not\equiv^k_{\infty\omega} (\mathbf{B}, b_1, \ldots, b_k)$?" Note that the increase in the number of variables is inevitable, since this is a $2k$-ary query. Dawar et al. [DLW95] pursued further the analysis of $\mathcal{L}^k_{\infty\omega}$-equivalence by focusing on the definability of the equivalence classes of $\equiv^k_{\infty\omega}$. Before describing their results in more detail, we need to make the concepts and the notation precise.

Definition 3.2: Let k be a positive integer, **B** a σ-structure, and **b** a k-tuple from **B**.

• $Q^k_{\mathbf{B},\mathbf{b}}$ denotes the k-ary query on σ such that, given a σ-structure **A**, we have that

$$Q^k_{\mathbf{B},\mathbf{b}}(\mathbf{A}) = \{\mathbf{a} : (\mathbf{A}, \mathbf{a}) \equiv^k_{\infty\omega} (\mathbf{B}, \mathbf{b})\}.$$

• $Q^k_{\mathbf{B}}$ denotes the $\equiv^k_{\infty\omega}$-equivalence class of **B**, which means that $Q^k_{\mathbf{B}} = \{\mathbf{A} : \mathbf{A} \equiv^k_{\infty\omega} \mathbf{B}\}$. Note that $Q^k_{\mathbf{B}}$ is a Boolean query on σ. ∎

Dawar et al. [DLW95] established that for every σ-structure **B** and every k-tuple **b** from **B** the queries $Q^k_{\mathbf{B},\mathbf{b}}$ and $Q^k_{\mathbf{B}}$ are FO^k-definable. This was achieved by a painstaking adaptation to $\mathcal{L}^k_{\infty\omega}$ of Scott's Theorem [Sco65] to the effect that the isomorphism type of every countable structure is definable in the infinitary logic $\mathcal{L}_{\omega_1\omega}$. In what follows in this section, we give a different proof of the results of Dawar et al., a proof that demonstrates the expressive power of systems of positive FO^k-formulas and makes a case that LFP^k, $k \geq 1$, is indeed the right ramification of LFP.

Theorem 3.3 : *Let* k *be a positive integer,* **B** *a* σ*-structure, and* $\mathbf{b}_1, \ldots, \mathbf{b}_l$ *an enumeration of all* k*-tuples from* **B**. *Then there is a system* $\mathsf{S}_{\mathbf{B}} = (\varphi_{\mathbf{B},\mathbf{b}_i}(x_1, \ldots, x_k, T_{\mathbf{b}_1}, \ldots, T_{\mathbf{b}_l}), 1 \leq i \leq l)$ *of* FO^k*-formulas that are positive in* $T_{\mathbf{b}_1}, \ldots, T_{\mathbf{b}_l}$ *and such that* $\varphi^\infty_{\mathbf{B},\mathbf{b}_i}(x_1, \ldots, x_k)$ *defines the complement of the query* $Q^k_{\mathbf{B},\mathbf{b}_i}$, $1 \leq i \leq l$. *In other words, for every* σ*-structure* **A** *and every* k*-tuple* **a** *from* **A**

$$(\mathbf{A}, \mathbf{a}) \equiv^k_{\infty\omega} (\mathbf{B}, \mathbf{b}) \Longleftrightarrow \mathbf{A}, \mathbf{a} \models \neg\varphi^\infty_{\mathbf{B},\mathbf{b}_i}(x_1, \ldots, x_k).$$

Sketch of Proof: For every $i \leq l$, let $\chi_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k)$ be the conjunction of all atomic or negated atomic formulas $\eta(x_1,\ldots,x_k)$ such that $\mathbf{B},\mathbf{b}_i \models \eta(x_1,\ldots,x_k)$. Moreover, for every $1 \leq j \leq k$ and every element b from the universe of **B**, we let $\mathbf{b}_i[j/b]$ be the k-tuple obtained from the k-tuple $\mathbf{b}_i = (b^i_1,\ldots,b^i_k)$ by replacing b^i_j by b. We then consider the system $\mathsf{S}_\mathbf{B} = (\varphi_{\mathbf{B},\mathbf{b}_i}, 1 \leq i \leq l)$, where $\varphi_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k,T_{\mathbf{b}_1}\ldots,T_{\mathbf{b}_l})$ is the formula

$$\begin{aligned}&\neg\chi_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k)\\ &\vee[\textstyle\bigvee_{j=1}^k(\exists x_j)\bigwedge_{b\in B}T_{\mathbf{b}_i[j/b]}(x_1,\ldots,x_k)]\\ &\vee[\textstyle\bigvee_{j=1}^k\bigvee_{b\in B}(\forall x_j)T_{\mathbf{b}_i[j/b]}(x_1,\ldots,x_k)].\end{aligned}$$

By induction on m simultaneously for all $i \leq l$, it can be shown that on every σ-structure **A** the component $\Phi^m_{\mathbf{B},\mathbf{b}_i}$, $1 \leq i \leq l$, of the m-th stage of the system $\mathsf{S}_\mathbf{B}$ consists of all k-tuples **a** from **A** such that the Spoiler can win within m rounds the k-pebble game on $(\mathbf{A},\mathbf{a})$ and $(\mathbf{B},\mathbf{b})$. ∎

By combining Theorem 2.3 with Theorem 3.3, we can now give a different proof of the Dawar et al. [DLW95] result to the effect that every $\equiv^k_{\infty\omega}$-equivalence class is FO^k-definable.

Theorem 3.4: *Let k be a positive integer, **B** a σ-structure, and **b** a k-tuple from **B**.*

- *There is a formula $\theta_{\mathbf{B},\mathbf{b}}(x_1,\ldots,x_k)$ of* FO^k *that defines the query $Q^k_{\mathbf{B},\mathbf{b}}$.*
- *There is a sentence $\theta_\mathbf{B}$ of* FO^k *that defines the query $Q^k_\mathbf{B}$.*

Sketch of Proof: Let $\mathbf{b}_1,\ldots,\mathbf{b}_l$ be an enumeration of all k-tuples from the universe of **B** and let $\mathsf{S}_\mathbf{B} = (\varphi_{\mathbf{B},\mathbf{b}_i}, 1 \leq i \leq l)$, be the system of positive FO^k-formulas used in the proof of Theorem 3.3. By Theorem 2.3, for every $i \leq l$ and every $m \geq 1$, there is a FO^k-formula $\varphi^m_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k)$ that defines the component $\Phi^m_{\mathbf{B},\mathbf{b}_i}$ of the m-th stage of this system.

Apply now the system $\mathsf{S}_\mathbf{B}$ to the structure **B** itself. Then there is a positive integer m_0 such that on **B** the least fixpoint of this system is equal to its m_0-th stage, that is, for every $i \leq l$ and every k-tuple **b** from **B**

$$\mathbf{B},\mathbf{b} \models \varphi^{m_0}_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k) \leftrightarrow \varphi^{m_0+1}_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k).$$

It can be shown that $Q^k_{\mathbf{B},\mathbf{b}}$ is definable by the following FO^k-formula $\theta_{\mathbf{B},\mathbf{b}}(x_1,\ldots,x_k)$:

$$\begin{aligned}&\neg\varphi^{m_0}_{\mathbf{B},\mathbf{b}}(x_1,\ldots,x_k)\wedge\\ &\textstyle\bigwedge_{i=1}^l(\forall x_1\ldots\forall x_k)(\varphi^{m_0}_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k)\leftrightarrow\\ &\qquad\varphi^{m_0+1}_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k)).\end{aligned}$$

Finally, the query $Q^k_\mathbf{B}$ is definable by the FO^k-sentence $(\exists x_1\ldots\exists x_k)(\bigvee_{i=1}^l\theta_{\mathbf{B},\mathbf{b}_i}(x_1,\ldots,x_k))$. ∎

4 FO vs. LFP and $\mathcal{L}^\omega_{\infty\omega}$

If $\mathcal{C}$ is an arbitrary class of σ-structures, then FO $\subseteq$ LFP $\subseteq \mathcal{L}^\omega_{\infty\omega}$ on $\mathcal{C}$. It is known that these containments are proper on the class $\mathcal{F}$ of all σ-structures and the class $\mathcal{O}$ of all ordered σ-structures (cf. [KV92b]). In contrast, FO $=$ LFP $= \mathcal{L}^\omega_{\infty\omega}$ on the class $\mathcal{E}$ of all finite equivalence relations [Kol85]. Based on the empirical evidence that was available at the time, McColm [McC86, McC90b] conjectured that both the collapse of LFP to first-order logic and the collapse of $\mathcal{L}^\omega_{\infty\omega}$ to first-order logic on an arbitrary class $\mathcal{C}$ of σ-structures are tantamount to a *boundedness* condition for LFP on $\mathcal{C}$. A positive first-order formula $\varphi(\mathbf{x},T)$ is said to be *bounded* on $\mathcal{C}$ if there is some $m \geq 0$ such that $\Phi^m = \Phi^\infty$ on every σ-structure in $\mathcal{C}$. For example, the formula $E(x,y) \vee (\exists z)(T(x,z) \wedge T(z,y))$ is bounded on every class of finite graphs having fixed diameter, but it is unbounded on the class of all finite graphs. LFP is *bounded* on $\mathcal{C}$ if every positive first-order formula $\varphi(\mathbf{x},T)$ is bounded on $\mathcal{C}$. With this terminology at hand, we can state McColm's conjectures precisely.
McColm's First Conjecture: LFP *collapses to* FO *on a class $\mathcal{C}$ of σ-structures if and only if* LFP *is bounded on $\mathcal{C}$.*
McColm's Second Conjecture: *$\mathcal{L}^\omega_{\infty\omega}$ collapses to* FO *on a class $\mathcal{C}$ of σ-structures if and only if* LFP *is bounded on $\mathcal{C}$.*
As mentioned earlier, the first conjecture was refuted by Gurevich, Immerman, Shelah [GIS94], while the second was affirmed by us in [KV92a]. Both papers make crucial use of the fact that McColm's conjectures are about logics whose formulas have a finite but arbitrarily large number of variables. In particular, McColm's Second Conjecture is proved in [KV92a] by establishing that the boundedness on $\mathcal{C}$ of positive $\mathcal{L}^{2k}_{\omega\omega}$-formulas implies that $\mathcal{L}^k_{\infty\omega} =$ FO^k on $\mathcal{C}$, which, in turn, implies the boundedness on $\mathcal{C}$ of positive FO^k-formulas. This leaves a gap between $2k$ and k, which

can not be bridged by modifying the proof in any straightforward way.

How do the variable-confined fragments of first-order logic, LFP, and $\mathcal{L}^{\omega}_{\infty\omega}$ compare on arbitrary classes of finite structures? In particular, is the collapse of $\mathcal{L}^{k}_{\infty\omega}$ to FO^k on a class $\mathcal{C}$ of σ-structures equivalent to a boundedness condition for LFP^k on $\mathcal{C}$? Note that in McColm's conjectures the concept of boundedness of LFP was defined in terms of least fixpoints of single positive first-order formulas. When it comes to systems of positive first-order formulas, however, more than one concepts of boundedness arise, because some components of a system may not change when passing from one stage of the system to the next. To illustrate this point, consider again the system

$$\begin{aligned}\varphi_1(x,y,T_1,T_2) &\equiv E(x,y)\vee\\ &\quad (\exists z)(T_1(x,z)\wedge T_1(z,y)),\\ \varphi_2(x,T_1,T_2) &\equiv T_1(x,x).\end{aligned}$$

in which $\varphi_1^{\infty}(x,y)$ defines the *transitive closure* query and $\varphi_2^{\infty}(x)$ defines the *cycle* query. It is possible that in a graph G there are two nodes of distance n, but there is no cycle of length n, which means that $G \models (\varphi_1^n \neq \varphi_1^{n-1}) \wedge (\varphi_2^n = \varphi_2^{n-1})$. In particular, if G is an acyclic graph, then $\varphi_2^m = \emptyset$ for all $m \geq 0$. This phenomenon suggests that we consider the following two functions.

Definition 4.1: Let $\mathcal{F}$ be the class of all σ-structures and let $\mathsf{S} = (\varphi_1, \ldots, \varphi_l)$ be a system of positive first-order formulas over the vocabulary σ.

• The *closure function* $\mathrm{cl}(\mathsf{S}) : \mathcal{F} \mapsto \{1, 2, \ldots\}$, given a σ-structure $\mathbf{A}$, takes value

$$\mathrm{cl}(\mathsf{S})(\mathbf{A}) = \min\{m : \mathbf{A} \models \bigwedge_{i=1}^{l} (\varphi_i^m = \varphi_i^{m-1})\}.$$

• The *change function* $\mathrm{ch}(\mathsf{S}) : \mathcal{F} \mapsto \{1, 2, \ldots\}$, given a σ-structure $\mathbf{A}$, takes value

$\mathrm{ch}(\mathsf{S})(\mathbf{A}) = \max_{1\leq i\leq l}\{$the number of distinct stages φ_i^m, $m \geq 0$, of component φ_i on $\mathbf{A}\}$.

In words, the closure function gives the number of distinct stages of the system, while the change function gives the maximum number of changes of a component of the system. ∎

If φ is a single formula, then $\mathrm{cl}(\varphi)(\mathbf{A}) = \mathrm{ch}(\varphi)(\mathbf{A})$ for every σ-structure $\mathbf{A}$. It is also easy to see that if $\mathsf{S} = (\varphi_1, \ldots, \varphi_l)$ is a system and $\mathbf{A}$ is a σ-structure, then $\mathrm{ch}(\mathsf{S})(\mathbf{A}) \leq \mathrm{cl}(\mathsf{S})(\mathbf{A}) \leq l \cdot \mathrm{ch}(\mathsf{S})(\mathbf{A})$. Note that the change function of a system may take values that are always less than the values of the closure function. As an example, for every $l \geq 1$, let $Y_l = (\psi_1, \ldots, \psi_l)$ be the system (for $2 \leq i \leq l$):

$$\begin{aligned}\psi_1(x,T_1,\ldots,T_l) &\equiv (x = x),\\ \psi_i(x,T_1,\ldots,T_l) &\equiv (x = x)\wedge(\exists x)T_{i-1}(x)\end{aligned}$$

If $\mathbf{A}$ is a σ-structure, then $\mathrm{ch}(Y_l)(\mathbf{A}) = 2$, while $\mathrm{cl}(Y_l)(\mathbf{A}) = l$. Next, we introduce the key concept that will enable us to show that the collapse of $\mathcal{L}^{k}_{\infty\omega}$ to FO^k on a class $\mathcal{C}$ is equivalent to a *uniform boundedness* condition for the stage functions of systems of arbitrary FO^k-formulas on $\mathcal{C}$.

Definition 4.2: Let k be a positive integer and $\mathcal{C}$ a class of σ-structures. We say that LFP^k is *uniformly bounded* on $\mathcal{C}$ if there is a positive integer m_0 such that for every system $\mathsf{S} = (\varphi_1, \ldots, \varphi_l)$ of positive FO^k-formulas and for every σ-structure $\mathbf{A}$ in $\mathcal{C}$ we have that $\mathrm{ch}(\mathsf{S})(\mathbf{A}) \leq m_0$. ∎

Note that it is not meaningful to formulate a concept of uniform boundedness in terms of the closure function, since there is no uniform bound on the closure functions of the aforementioned systems Y_l, $l \geq 1$. We are now ready to establish the main result of this section, which reveals that a ramified version of McColm's Second Conjecture governs the relationship between FO^k, LFP^k, and $\mathcal{L}^{k}_{\infty\omega}$. Below, if $\mathbf{B}$ is a σ-structure and $\mathbf{b}_1, \ldots, \mathbf{b}_l$ is an enumeration of all k-tuples from $\mathbf{B}$, then $\mathsf{S}_{\mathbf{B}} = (\varphi_{\mathbf{B},\mathbf{b}_i}, 1 \leq i \leq l)$ is the system in Theorem 3.3 that computes the complements of the queries $Q^k_{\mathbf{B},\mathbf{b}_i}$, $1 \leq i \leq l$.

Theorem 4.3: *Let k be a positive integer and let $\mathcal{C}$ be a class of σ-structures. Then the following are equivalent:*

(1) $\mathcal{L}^{k}_{\infty\omega} =$ FO^k *on* $\mathcal{C}$.

(2) LFP^k *is uniformly bounded on* $\mathcal{C}$.

(3) *The change functions of the systems* $\mathsf{S}_{\mathbf{B}}$ *are uniformly bounded on* $\mathcal{C}$*, that is, there is a positive integer* m_0 *such that if* $\mathbf{A}$ *and* $\mathbf{B}$ *are members of* $\mathcal{C}$*, then* $\mathrm{ch}(\mathsf{S}_{\mathbf{B}})(\mathbf{A}) \leq m_0$.

(4) *The closure functions of the systems* $\mathsf{S}_{\mathbf{B}}$ *are uniformly bounded on* $\mathcal{C}$*, that is, there is a positive integer* m_1 *such that if* $\mathbf{A}$ *and* $\mathbf{B}$ *are members of* $\mathcal{C}$*, then* $\mathrm{cl}(\mathsf{S}_{\mathbf{B}})(\mathbf{A}) \leq m_1$.

Sketch of Proof: We establish a cycle of implications; the most difficult one is $(3) \Rightarrow (4)$.

$(3) \Rightarrow (4)$: Assume that the change functions of the systems $\mathsf{S}_{\mathbf{B}}$ are uniformly bounded on $\mathcal{C}$. We have to show that the closure functions of these systems are uniformly bounded on $\mathcal{C}$. By examining the proofs of Theorems 3.3 and 3.4 we can show that if $\mathbf{A}$ is an arbitrary σ-structure, then the value $\mathrm{cl}(\mathsf{S}_{\mathbf{B}})(\mathbf{A})$ is bounded by a quantity that depends only on $\mathrm{cl}(\mathsf{S}_{\mathbf{B}})(\mathbf{B})$. Thus, to establish the uniform boundedness of the closure functions of the systems $\mathsf{S}_{\mathbf{B}}$, it suffices to show that there is a positive integer m_1 such that for every structure $\mathbf{B}$ in $\mathcal{C}$ we have $\mathrm{cl}(\mathsf{S}_{\mathbf{B}})(\mathbf{B}) \leq m_1$.

It is well known that the complement of the least fixpoint of any positive formula $\varphi(x_1, \ldots, x_n, T)$ is the greatest fixpoint of the *dual* formula $\tilde{\varphi}(x_1, \ldots, x_n, T) \equiv \neg\varphi(x_1, \ldots, x_n, \neg T)$, which is a positive formula as well. Moreover, the greatest fixpoint of $\tilde{\varphi}$ can be obtained via a top-down (co-inductive) evaluation. More precisely, if we put $\tilde{\Phi}^0 = A^n$ and $\tilde{\Phi}^{m+1} = \neg\Phi(\neg\tilde{\Phi}^m)$, $m \geq 0$, then the greatest fixpoint of $\tilde{\varphi}$ is the first stage $\tilde{\Phi}^{m_0}$ such that $\tilde{\Phi}^{m_0} = \tilde{\Phi}^{m_0+1}$. This extends naturally to systems of positive formulas with notational modifications only. For every structure $\mathbf{B}$ in $\mathcal{C}$, we now consider the dual system $\tilde{\mathsf{S}}_{\mathbf{B}} = (\tilde{\varphi}_{\mathbf{B},\mathbf{b}_i}, 1 \leq i \leq l)$, of the system $\mathsf{S}_{\mathbf{B}}$. Thus, the components of the greatest fixpoint of the dual system $\tilde{\mathsf{S}}_{\mathbf{B}}$ define the queries $Q^k_{\mathbf{B},\mathbf{b}_i}$, $1 \leq i \leq l$. In particular, if $\tilde{\mathsf{S}}_{\mathbf{B}}$ is evaluated on $\mathbf{B}$ itself, then the components of its greatest fixpoint give the $\equiv^k_{\infty\omega}$-equivalence classes of k-tuples from $\mathbf{B}$.

We will analyze the top-down evaluation of the greatest fixpoint of $\tilde{\mathsf{S}}_{\mathbf{B}}$ on $\mathbf{B}$. This evaluation can be viewed as an iterative process that computes equivalence classes of successively finer equivalence relations on the set B^k of all k-tuples from $\mathbf{B}$. Indeed, from the proof of Theorem 3.3 it follows that for every $m \geq 0$ the components $\tilde{\Phi}^m_{\mathbf{B},\mathbf{b}_i}$, $1 \leq i \leq l$, of the m-stage of the dual system define the equivalence classes of the equivalence relation $\equiv^{k,m}$ on B^k, where $\mathbf{b} \equiv^{k,m} \mathbf{b}'$ if and only if the Duplicator can win m rounds of the k-pebble game on $(\mathbf{B}, \mathbf{b})$ and $(\mathbf{B}, \mathbf{b}')$. Thus, in the beginning (at stage 0) the equivalence relation $\equiv^{k,0}$ relation has B^k as its single equivalence class; moreover, in going from stage m to stage $m+1$, the equivalence relation $\equiv^{k,m+1}$ is obtained from the equivalence relation $\equiv^{k,m}$ by splitting one or more $\equiv^{k,m}$-equivalence classes. We now claim that the *index* (i.e., the number of equivalence classes) of each equivalence relation $\equiv^{k,m}$ is bounded by a quantity $g(m)$ that depends only on m and not on the structure $\mathbf{B}$ itself ($g(m)$ also depends on k and the arities of the relation symbols in the vocabulary σ, but these quantities have been fixed throughout this discussion). We proceed by induction on m. For $m = 0$ there is only one equivalence class, while the index of $\equiv^{k,1}$ is bounded by the number of quantifier-free formulas with k-variables over the vocabulary σ. Assume that for some $m \geq 1$ the index of $\equiv^{k,m}$ is bounded by $g(m)$. Note that a $\equiv^{k,m}$-equivalence class can be split in the next stage only if it contains two k-tuples that were not distinguishable within m rounds of the k-pebble game, but are distinguishable within $m+1$ rounds. It follows that an $\equiv^{k,m}$-equivalence class can be split in at most $2^{kg(m)}$ different $\equiv^{k,m+1}$-equivalence classes. Thus, the index of $\equiv^{k,m+1}$ is bounded by $g(m)2^{kg(m)}$.

It is now time to use the hypothesis that the change functions of the systems $\mathsf{S}_{\mathbf{B}}$ (and, hence, also of the dual systems $\tilde{\mathsf{S}}_{\mathbf{B}}$) are uniformly bounded on $\mathcal{C}$ by a positive integer m_0. Let $\mathbf{B}$ be a structure in $\mathcal{C}$ and let m be a non-negative integer. For each such pair $(\mathbf{B}, m)$, we introduce a m_0+1-tuple $(d_0, \ldots, d_{m_0})$ of non-negative integers, which we call the *trace* of $(\mathbf{B}, m)$ and denote by $tr(\mathbf{B}, m)$. This tuple contains information about the number of posssible future splittings of each $\equiv^{k,m}$-equivalence class on $\mathbf{B}$. More precisely, a non-zero entry d_i in $tr(\mathbf{B}, m)$ means that there are d_i distinct $\equiv^{k,m}$-equivalence classes on $\mathbf{B}$ each of which can split at most $m_0 + 1 - i$ times until the evaluation of the greatest fixpoint of the system $\tilde{\mathsf{S}}_{\mathbf{B}}$ terminates on $\mathbf{B}$. Thus, $tr(\mathbf{B}, 0) = (1, 0, \ldots, 0)$ and $tr(\mathbf{B}, 1) = (0, d_1, 0, \ldots, 0)$, where d_1 is the number of quantifier-free formulas with k variables that are inequivalent on $\mathbf{B}$. Let $\prec$ be the *strict lexicographic order* on m_0+1-tuples of non-negative integers. Then for every structure $\mathbf{B}$ in $\mathcal{C}$ there is some $m_0' < m_0$ such that for $0 \leq m \leq m_0'$ that $tr(\mathbf{B}, m+1) \prec tr(\mathbf{B}, m)$. Note also if $tr(\mathbf{B}, m) = (d_0, \ldots, d_{m_0})$ is a trace, then $d_0 + \ldots + d_{m_0}$ is the index of $\equiv^{k,m}$ on $\mathbf{B}$ and, thus,

$d_0 + \ldots + d_{m_0} \leq g(m)$.

Next, we define a directed acyclic graph with the traces as its nodes. For every $m \geq 0$, let $V_m = \{tr(\mathbf{B}, m) : \mathbf{B} \in \mathcal{C}\}$ and let $V = \bigcup_{m \geq 0} V_m$. We equip V with an edge relation E defined as follows. If $(d_0, \ldots, d_{m_0})$ is a member of V_m and $(e_0, \ldots, e_{m_0})$ is a member of V_{m+1}, then we put an edge between these two traces provided there is a $\mathbf{B}$ in $\mathcal{C}$ and an $m \geq 0$ such that $tr(\mathbf{B}, m) = (d_0, \ldots, d_{m_0})$ and $tr(\mathbf{B}, m+1) = (e_0, \ldots, e_{m_0})$; no other edges are allowed. Note that each set V_m is finite, since, as seen above, if $(d_0, \ldots, d_{m_0}) \in V_m$, then $d_0 + \ldots + d_{m_0} \leq g(m)$. As a result of this, the graph (V, E) is *locally finite*, i.e., each node has a finite number of edges incident to it. Moreover, the graph (V, E) contains no infinite one-way path, because the edge relation E agrees with the strict lexicographic order $\prec$, which is a well-founded partial order. We can now apply König's lemma and conclude that the graph V is finite, which means that there is a positive integer m_1 such that $V_m = V_{m_1+1}$ for every $m \geq m_1$. Consequently, for every structure $\mathbf{B}$ in $\mathcal{C}$ the evaluation of the greatest fixpoint of $\tilde{\mathsf{S}}_{\mathbf{B}}$ terminates within m_1 steps. Thus, the closure functions of the systems $\mathsf{S}_{\mathbf{B}}$ are bounded by m_1 uniformly on $\mathcal{C}$.

$(4) \Rightarrow (1)$ Let m_1 be a positive integer such that for every σ-structure $\mathbf{B}$ in $\mathcal{C}$ the closure function $\mathrm{cl}(\mathsf{S}_{\mathbf{B}})$ of the associated system $\mathsf{S}_{\mathbf{B}} = (\varphi_{\mathbf{B},\mathbf{b}_i}, 1 \leq i \leq l)$ is bounded by m_1. Note that the quantifier rank of every system $\mathsf{S}_{\mathbf{B}}$ is 1. From Theorem 2.3 and from the proofs of Theorems 3.3 and 3.4, it follows that, for every σ-structure $\mathbf{B}$ in $\mathcal{C}$ and every k-tuple $\mathbf{b}$ from $\mathbf{B}$, the query $Q^k_{\mathbf{B},\mathbf{b}}$ is definable by a FO^k-formula of quantifier rank at most $m_1 + k + 1$. This means that every $\equiv^k_{\infty\omega}$-equivalence class is definable by a FO^k-formula of fixed quantifier rank. There are, however, only finitely many FO^k-formulas of fixed quantifier rank. Thus, there are finitely many distinct k-ary queries that are FO^k-definable on $\mathcal{C}$ and, consequently, $\mathcal{L}^k_{\infty\omega} = \mathrm{FO}^k$ on $\mathcal{C}$. ∎

The preceding Theorem 4.3 makes a compelling case as to why systems of positive FO^k-formulas constitute the right way to ramify LFP. Indeed, it is the use of systems that made it possible to distinguish the change function from the closure function and to discover the uniform boundedness condition on the change functions. As mentioned earlier, the closure functions are never uniformly bounded and, thus, the collapse of $\mathcal{L}^k_{\infty\omega}$ to FO^k can not be characterized in terms of uniform boundedness of the closure function. Theorem 4.3 asserts that a ramified version of McColm's Second Conjecture holds. As a matter of fact, this ramified version is finer than the original conjecture, since in the full paper we show that the confirmation of McColm's Second Conjecture can be derived easily from Theorem 4.3.

We now revisit McColm's First Conjecture and ask whether it holds in its ramified version.

Open Problem: Let k be a positive integer and let $\mathcal{C}$ be a class of σ-structures. Are the following two statements equivalent?

(1) $\mathcal{L}^k_{\infty\omega} = \mathrm{FO}^k$ on $\mathcal{C}$.

(2) $\mathrm{LFP}^k = \mathrm{FO}^k$ on $\mathcal{C}$.

It should be pointed out that Gurevich, Immerman, and Shelah [GIS94] present two counterexamples to McColm's First Conjecture, but neither one yields a class $\mathcal{C}$ that violates the equivalence between the two statements above. We conclude by stating a partial result that sheds some light on this open problem. In the result below, the proof of which will be given in the full paper, the notation FO^{k,d} stands for the collection of all FO^k-formulas of quantifier rank at most d.

Theorem 4.4: *Let k be a positive integer and let $\mathcal{C}$ be a class of σ-structures. Then the following statements are equivalent:*

(1) $\mathcal{L}^k_{\infty\omega} = \mathrm{FO}^k$ *on* $\mathcal{C}$.

(2) *There is a positive integer d such that* $\mathcal{L}^k_{\infty\omega} = \mathrm{FO}^{k,d}$ *on* $\mathcal{C}$.

(3) *There is a positive integer d such that* $\mathrm{LFP}^k = \mathrm{FO}^{k,d}$ *on* $\mathcal{C}$.

Thus, the ramified version of McColm's First Conjecture reduces to the following equivalence: LFP^k collapses to FO^k on $\mathcal{C}$ if and only if there is a d such that LFP^k collapses to FO^{k,d} on $\mathcal{C}$.

5 Concluding Remarks

Motivated by considerations from descriptive complexity theory, we studied here the expressive power of variable-confined logics. Our main assertion is that to study variable-confined least-fixpoint logic we need to consider variable-confined *systems* of positive first-order formulas. As we saw, "measuring" the "convergence" of simultanous fixpoints is quite subtle, as one has to distinguish between the closure function and the change function. We characterized the collapse of $\mathcal{L}^k_{\infty\omega}$ to FO^k in terms of the uniform boundedness of the change function, though the closure function is never uniformly bounded.

While our results constitute an important step in understanding the behavior of variable-confined logics, several important questions remain open. We already mentioned the ramified version of McColm's First Conjecture. The precise relationship between boundedness and uniform boundedness is also open: can LFP^k be bounded on a class $\mathcal{C}$ without being uniformly bounded on $\mathcal{C}$?

References

[AU79] A. V. Aho and J. D. Ullman. Universality of data retrieval languages. In *Proc. 6th ACM Symp. on Principles of Programming Languages*, pages 110–117, 1979.

[AV89] S. Abiteboul and V. Vianu. Fixpoint extensions of first-order logic and Datalog-like languages. In *Proc. 4th IEEE Symp. on Logic in Computer Science*, pages 71–79, 1989.

[AV95] S. Abiteboul and V. Vianu. Computing with first-order logic. *Journal of Computer and System Sciences*, 50(2):309–335, April 1995.

[AVV92] S. Abiteboul, Moshe Y. Vardi, and V. Vianu. Fixpoint logics, relational machines, and computational complexity. In *Proc. 7th IEEE Symp. on Structure in Complexity Theory*, 1992.

[Bar77] J. Barwise. On Moschovakis closure ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.

[BF85] J. Barwise and S. Feferman, editors. *Model-Theoretic Logics*. Springer-Verlag, 1985.

[CH82] A. Chandra and D. Harel. Structure and complexity of relational queries. *Journal of Computer and System Sciences*, 25:99–128, 1982.

[Cha88] A. Chandra. Theory of database queries. In *Proc. 7th ACM Symp. on Principles of Database Systems*, pages 1–9, 1988.

[Daw93] A. Dawar. *Feasible Computation through Model Theory*. PhD thesis, University of Pennsylvania, 1993.

[DH94] A. Dawar and L. Hella. The expressive power of finitely many generalized quantifiers. In *Proceedings of the Ninth IEEE Symposium on Logic and Computer Science*, pages 20–29, 1994.

[DLW95] A. Dawar, S. Lindell, and S. Weinstein. Infinitary logic and inductive definability over finite structures. *Information and Computation*, 119:160–175, 1995.

[Fag75] R. Fagin. Monadic generalized spectra. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 21:89–96, 1975.

[Gai82] H. Gaifman. On local and nonlocal properties. In J. Stern, editor, *Logic Colloquium '81*, pages 105–135. North Holland, 1982.

[GIS94] Y. Gurevich, N. Immerman, and S. Shelah. Mccolm's conjecture. In *Proc. 7th IEEE Symp. on Logic in Comp. Sci.*, pages 10–19, 1994.

[Hod93] I. Hodkinson. Finite variable logics. *Bulletin of the European Association for Theoretical Computer Science*, 51:111–140, October 1993.

[Imm81] N. Immerman. Number of quantifiers is better than number of tape cells. *Journal of Computer and System Sciences*, 22:384–406, 1981.

[Imm82] N. Immerman. Upper and lower bounds for first-order expressibility. *Journal of Computer and System Sciences*, 25:76–98, 1982.

[Imm86] N. Immerman. Relational queries computable in polynomial time. *Information and Control*, 68:86–104, 1986.

[Imm91] N. Immerman. Dspace[n^k] = var[$k+1$]. In *Proc. 6th IEEE Symp. on Structure in Complexity Theory*, pages 334–340, 1991.

[Kol85] Ph. G. Kolaitis. On asymptotic probabilities of inductive queries and their decision problem. In R. Parikh, editor, *Logics of Programs '85, Lecture Notes in Computer Science 193*, pages 153–166. Springer-Verlag, 1985.

[KV87] Ph. G. Kolaitis and M. Y. Vardi. The decision problem for the probabilities of higher-order properties. In *Proc. 19th ACM Symp. on Theory of Computing*, pages 425–435, 1987.

[KV92a] Ph. G. Kolaitis and M. Y. Vardi. Fixpoint logic vs. infinitary logic in finite-model theory. In *Proc. 6th IEEE Symp. on Logic in Comp. Sci.*, pages 46–57, 1992.

[KV92b] Ph. G. Kolaitis and M. Y. Vardi. Infinitary logic and 0-1 laws. *Information and Computation*, 98:258–294, 1992. Special issue: Selections from the Fifth Annual IEEE Symposium on Logic in Computer Science.

[KV92c] Ph. G. Kolaitis and M. Y. Vardi. Infinitary logic for computer science. In *Proc. 19th International Colloq. on Automata, Languages, and Programming*, pages 450–473. Springer-Verlag, Lecture Notes in Computer Science 623, 1992.

[McC86] G.L. McColm. *Simple and simultaneous recursive fixed points*. PhD thesis, University of California, Los Angeles, 1986.

[McC90a] G. L. McColm. Parametrization over inductive relations of a bounded number of variables. *Annals of Pure and Applied Logic*, 48:103–134, 1990.

[McC90b] G. L. McColm. When is arithmetics possibe? *Annals of Pure and Applied Logic*, 50:29–51, 1990.

[Mos74] Y. N. Moschovakis. *Elementary Induction on Abstract Structures*. North Holland, 1974.

[Sco65] D. Scott. Logic with denumerably long formulas and finite strings of quantifiers. In J.W. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 320–341. North Holland, 1965.

[Var82] M. Y. Vardi. The complexity of relational query languages. In *Proc. 14th ACM Symp. on Theory of Computing*, pages 137–146, 1982.

[Var95] M.Y. Vardi. On the complexity of bounded-variable queries. In *Proc. 14th ACM Symp. on Principles of Database Systems*, pages 266–276, 1995.