INSTITUT FÜR INFORMATIK

UND PRAKTISCHE MATHEMATIK

First-Order Closure and the Monadic Second-Order Alternation Hierarchy

Oliver Matz

Bericht Nr. 9807
Mai 1998

CHRISTIAN-ALBRECHTS-UNIVERSITÄT

KIEL

First-Order Closure and the Monadic Second-Order Alternation Hierarchy

Oliver Matz

e-mail: oma@informatik.uni-kiel.de

Technical Report No. 9807
Institut für Informatik und Praktische Mathematik
Christian-Albrechts-Universität Kiel
24098 Kiel, Germany

Abstract. The first-order closure of a class $\mathcal{F}$ of formulas is the smallest superclass of $\mathcal{F}$ that is closed under first-order quantification. We consider (properties definable in) the first-order closure of different levels of the monadic second-order quantifier alternation hierarchy over grids or graphs.

We show that in this hierarchy, the first-order closure of the boolean closure of Σ_k is less expressive than the first-order closure of the boolean closure of Σ_{k+1}, more expressive than the boolean closure of Σ_k, but not more expressive than Δ_{k+2}.

To show this, we consider sets of rectangular grids where the width is a function of the height. If such a set is definable in the first-order closure of the boolean closure of Σ_k, then the corresponding function is at most $(k+1)$-fold exponential (this generalizes a result of [AFS97] from $k = 1$ to arbitrary k). Moreover, we prove that this bound is asymptotically optimal (this solves, for the case $k = 1$, an open question of [AFS97]).

1 Introduction

The subject of this paper is monadic second-order logic over finite graphs. The monadic second-order quantifier alternation hierarchy over graphs classifies formulas (and hence graph properties) wrt. the alternation depth of set quantifiers.

Σ_k (respectively Π_k) contains all formulas in prenex normal form that have a prefix of k alternating blocks of set quantifiers starting with an existential (respectively universal) block, followed by a first-order kernel. For every class $\mathcal{F}$ of formulas, $\underline{\mathcal{F}}$ denotes the set of all properties definable by an $\mathcal{F}$-sentence. Let $\underline{\Delta_k} = \underline{\Sigma_k} \cap \underline{\Pi_k}$.

With a certain amount of arbitrariness we choose $\underline{\Sigma_k}$ to be referred to as "level k of the monadic hierarchy". The choice of $\underline{\Pi_k}$, $\underline{\Delta_k}$, or the boolean closure of $\underline{\Sigma_k}$ would have been as canonical.

The first-order closure of a class $\mathcal{F}$ of formulas contains all formulas that result from a $\mathcal{F}$-formula by prefixing it with first-order quantifications. In other words, the first-order (respectively first-order/boolean) closure of $\mathcal{F}$ is the smallest superclass of $\mathcal{F}$ that is closed under first-order quantifications. The first-order closure of the boolean closure of $\mathcal{F}$ is called "first-order/boolean" closure of $\mathcal{F}$.

In [MST98] it is shown that the monadic second-order quantifier alternation hierarchy over graphs is strict in the sense that every level is more expressive than the boolean closure of the previous level, more precisely: the boolean closure of $\underline{\Sigma_k}$ is strictly included in $\underline{\Delta_{k+1}}$.

In this paper we show firstly that, for every k, there is a $\underline{\Delta_{k+2}}$-property not expressible by a sentence in the first-order/boolean closure of Σ_k, secondly that for each k there is a property which is expressible both in the first-order closure of Σ_k and in the first-order closure of Π_k, but not in the boolean closure of Σ_k, and thirdly that the first-order closures both of Σ_{k+1} and Π_{k+1} are more expressive than the first-order/boolean closure of Σ_k. The latter result may informally be rephrased as: The monadic hierarchy remains strict when each level is closed under first-order quantifications. The proofs refine automata theoretic techniques of [MT97] and techniques from [Sch97].

In contrast to these results, which show that in some sense the first-order closure is weak compared to monadic quantifier alternation, it is shown in [MST98] that there are properties definable in the first-order closures both of Σ_1 and of Π_1 and yet arbitrarily

high in the monadic hierarchy. This indicates that the alternation of first-order quantifications does mean an essential (wrt. the monadic quantifier alternation hierarchy) increase of expressive power in case the first-order quantifier block preceeds some second-order quantifications.

As in [MT97,Sch97,MST98] our proofs use a class of structures different from graphs, namely the class of (rectangular) grids, i.e. finite structures with two distinct binary relations that can be considered the vertical and horizontal successor relations when the universe of the structure is displayed as a matrix. As in the above papers, we associate to every function f from natural numbers to natural numbers the set L_f of all grids that are of size $m \times f(m)$ for some m. In [MT97] it has been shown that if L_f is definable in the boolean closure of Σ_k, then f is at most k-fold exponential. In this paper we show that if L_f is definable in the first-order/boolean closure of Σ_k, then f is at most $(k+1)$-fold exponential.

On the other hand, it has been shown in [Sch97,MST98] that for every k, there is a k-fold exponential function f such that the grid property L_f is in Δ_k. In this paper we will use Schweikardt's techniques to show that there is a $(k+1)$-fold exponential function f such that the grid property L_f is expressible in the first-order closure both of Σ_k and of Π_k. This solves an open question of [AFS97].

The separation results can be transferred from the class of grids to the class of (directed or undirected) finite graphs by the encoding technique of [MT97,MST98].

1.1 Historical Remarks

In [Fag75] it was shown that the first level $\underline{\Sigma_1}$ of the monadic hierarchy over graphs is not closed under complement. The question whether the monadic hierarchy is strict was raised in [Fag95].

After that question had been answered affirmatively in [MT97], the authors of [AFS97,AFS98] noted the drawback that the levels of the monadic hierarchy are not closed under first-order quantifications, whereas the levels of the polynomial hierarchy (or, by [Imm88] equivalently, of the full second-order quantifier alternation hierarchy) are. They raised the question whether also the "closed" hierarchy is strict. A formula on the k-th level of this closed hierarchy is obtained from a formula of the k-th level of the ordinary monadic hierarchy by interspersing the second-order quantifiers with (universal or existential) first-order quantifiers. It is observed in [AFS97] that the properties of [MT97] witnessing the infiniteness of the monadic hierarchy are also definable in the first level of the closed hierarchy, thus this level is not captured within any fixed level of the monadic hierarchy.

The same paper [AFS97] introduces the first-order/boolean closure of the levels of the monadic hierarchy. The authors show (using Ehrenfeucht-Fraïssé-like games) that the first level of the closed hierarchy contains a graph property not in the first-order/boolean closure of Σ_1. Besides, they show that if f is a function on the natural numbers and L_f (as defined above) is definable in the first-order/boolean closure of Σ_1 over grids, then f is at most doubly exponential, leaving open whether it can be more than singly exponential.

Acknowledgments. I thank Nicole Schweikardt and Thomas Wilke for pointing out some serious problems in a prior version of this paper, and Wolfgang Thomas for helpful advice concerning the presentation. I thank Thomasz Schoen for his ad-hoc tutorial in number theory and his proof for the first claim of Proposition 41.

2 Definitions and Results

Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the set of positive integers. For positive integers i, j with $i \leq j$ we define $[i, j] = \{i, i+1, \ldots, j\}$ and $[j] = [1, j]$.

2.1 Expansions

Let τ be a relational signature, A be some τ-structure, and $x_1, \ldots, x_r$ constant symbols. Let $u_1, \ldots, u_r \in \mathrm{dom}A$ be elements in the domain of A. Then $(A, u_1, \ldots, u_r)$ denotes the $(\tau \cup \{x_1, \ldots, x_r\})$-expansion of A with $u_1, \ldots, u_r$.

2.2 Graphs and Grids

The signature of graphs $\tau_{\text{Graph}} = \{E\}$ contains a binary relation symbol E. A (finite, directed) *graph* is a finite τ_{Graph}-structure, i.e. a pair $(\mathrm{dom}G, E^G)$ with $E^G \subseteq \mathrm{dom}G \times \mathrm{dom}G$. If E^G is irreflexive and symmetric, we call G *undirected.*

The signature $\tau_{\text{Grid}} = \{S_1, S_2\}$ containing the binary "successor" relation symbols S_1 and S_2 is the signature for *grids*. The universe of a grid is of the form $[m] \times [n]$. The *grid of height m and width n* is the τ_{Grid}-structure

$$\underline{[m] \times [n]} := ([m] \times [n], S_1^{m,n}, S_2^{m,n}),$$

where $S_1^{m,n}$ and $S_2^{m,n}$ are the "vertical" and the "horizontal" successor relations on $[m] \times [n]$, containing all pairs $((i,j),(i+1,j))$, (respectively all pairs $((i,j),(i,j+1))$) from $[m] \times [n]$.

The expressions *row, column, top, bottom* etc. are interpreted as in the terminology of matrices; e.g. the leftmost column contains exactly the positions $(i, 1)$ for $1 \leq i \leq m$, and the top row consists of all positions $(1, j)$ for $1 \leq j \leq n$.

Let $t, r \geq 0$. By $\tau_{t,r\text{-Grid}}$ we denote the signature $\tau_{\text{Grid}} \cup \{X_1, \ldots, X_t\} \cup \{x_1, \ldots, x_r\}$, where $X_1, \ldots, X_t$ are unary relation symbols and $x_1, \ldots, x_r$ are constant symbols. An $\tau_{t,r\text{-Grid}}$-expansion of a grid is called a *(t,r)-grid* for short.

2.3 Formulas

Let τ be a relational signature. We use $x, y, z, \ldots$ as first-order variables (i.e., ranging over elements of the universe) and $C, D, X, Y, \ldots$ as monadic second-order variables (i.e., ranging over subsets of the universe). *Monadic second-order formulas over τ* are built up as usual from atomic formulas $x = y$, $X(y)$, and $r(x_1, \ldots, x_l)$ (for an l-ary relation symbol r of τ) by means of propositional connectives and first- and second-order quantifications. *First-order formulas* over τ are monadic second-order formulas in which no second-order quantifier occurs.

If a formula φ has free variables among $X_1, \ldots, X_t, x_1, \ldots, x_r$, we sometimes write $\varphi(X_1, \ldots, X_t, x_1, \ldots, x_r)$.

If $\mathcal{C}$ is a class of τ-structures and φ is a sentence, i.e., has no free variables, then we write $Mod_{\mathcal{C}}(\varphi)$ for the set $\{M \in \mathcal{C} \mid M \models \varphi\}$ of models of φ. In the context of grids, a formula φ with free variables among $X_1, \ldots, X_t, x_1, \ldots, x_r$ is considered a sentence over the signature $\tau_{t,r\text{-Grid}}$. For such a formula we write $Mod_{t,r}(\varphi)$ for the set of (t,r)-grids that are models of φ.

A class $\mathcal{F}$ of formulas is *closed under first-order quantifications* (respectively *boolean combinations*, respectively *existential monadic second-order quantifications*) iff whenever φ, ψ are in $\mathcal{F}$, then so are $\exists x\varphi$ and $\forall x\varphi$ (respectively $\varphi \vee \psi$ and $\neg\varphi$, respectively $\exists X\varphi$).

Example 1. Consider the signature τ_{Grid} of grids. The atomic formula S_1xy asserts for two positions x, y in a grid that y is the vertical successor of x. The first-order formula $\mathit{topleft}(x) = \neg\exists y(S_1yx) \wedge \neg\exists y(S_2yx)$ asserts for a posistion x in a grid that $x = (1,1)$. The first-order formula $\psi(x) = \neg\exists y(S_1xy) \leftrightarrow \neg\exists y(S_2xy)$ asserts for a position x of a grid that it is at the right border iff it is at the bottom border. The monadic second-order sentence

$$\begin{aligned}\varphi = \exists X \Big(&\forall x\,\forall y\,\forall z\big(S_2xy \wedge S_1yz \rightarrow (Xx \leftrightarrow Xz)\big) \\ &\wedge \big(\forall x(Xx \rightarrow \psi(x))\big) \wedge \exists x\big(Xx \wedge \mathit{topleft}(x)\big)\Big)\end{aligned}$$

is valid in a grid iff it is a square, i.e., $Mod_{0,0}(\varphi) = \{\underline{[m] \times [m]} \mid m \geq 1\}$.

Definition 2 (The (Closed) Monadic Hierarchy). *Let τ be a fixed signature.*

Let Σ_0 and closed(Σ_0) *be the set of all first-order formulas.*

For every $k \geq 0$, we denote by Σ_{k+1} the smallest set of formulas that contains all negations of formulas in Σ_k and is closed under existential monadic second-order quantification. Let Π_k denote the set of all negations of formulas in Σ_k.

For every $k \geq 0$, we denote by closed(Σ_{k+1}) *the smallest set of formulas that contains all negations of formulas in* closed(Σ_k) *and is closed under first-order quantification and existential monadic second-order quantification.*

Definition 3. *Let $\mathcal{C}$ be a set of τ-structures. Two formulas φ, ψ are* equivalent over $\mathcal{C}$ *iff $\varphi \leftrightarrow \psi$ is true in all structures of $\mathcal{C}$ under all assignments. For a set $\mathcal{F}$ of formulas we write $\mathcal{F}^{\mathcal{C}}$ for the set of all formulas that are equivalent (over $\mathcal{C}$) to a formula in $\mathcal{F}$, and $\underline{\mathcal{F}}(\mathcal{C}) := \{Mod_{\mathcal{C}}(\varphi) \mid \varphi$ is a sentence in $\mathcal{F}\}$.*

Let ${\Delta_k}^{\mathcal{C}} = {\Sigma_k}^{\mathcal{C}} \cap {\Pi_k}^{\mathcal{C}}$ for every k and $\underline{\Delta_k}(\mathcal{C}) = \underline{{\Delta_k}^{\mathcal{C}}}(\mathcal{C})$.

In any of the above notations, we omit $\mathcal{C}$ when this causes no confusion. If nothing else is stated, $\mathcal{C}$ is the class of grids. Consequently, we do not distinguish between a class of formulas and its closure under logical equivalence over grids—every formula that is equivalent over grids to a, say, Σ_k-formula will be called "Σ_k-formula" itself.

B$(\mathcal{F})$ (respectively FO$(\mathcal{F})$, respectively FOB$(\mathcal{F})$) denotes, for every class $\mathcal{F}$ of formulas, the smallest superset of $\mathcal{F}$ that is closed under boolean combinations (respectively under first-order quantifications, respectively under both of these operations). These classes are called the *boolean* (respectively *first-order*, respectively *first-order/boolean*) *closure* of $\mathcal{F}$.

Example 4. Let **Grid** be the class of grids. The sentence φ of Example 1 is in Σ_1, and it is also in ${\Pi_1}^{\mathbf{Grid}}$ (and hence in ${\Delta_1}^{\mathbf{Grid}}$) because it is equivalent over grids to

$$\begin{aligned}\varphi' = \forall X \Big(&\forall x\,\forall y\,\forall z\big(S_2xy \wedge S_1yz \rightarrow (Xx \leftrightarrow Xz)\big) \\ &\wedge \big(\forall x(Xx \rightarrow \psi(x))\big) \rightarrow \exists x\big(Xx \wedge \mathit{topleft}(x)\big)\Big).\end{aligned}$$

Remark 5. Let $\mathcal{C}$ be a class of structures. Then ${\Delta_k}^{\mathcal{C}} \subseteq {\Sigma_k}^{\mathcal{C}} \subseteq \mathrm{B}(\Sigma_k)^{\mathcal{C}}$ and these classes are equal iff ${\Sigma_k}^{\mathcal{C}}$ is closed under negation. Furthermore, $\mathrm{FOB}(\mathcal{F})^{\mathcal{C}} = \mathrm{FO}(\mathrm{B}(\mathcal{F}))^{\mathcal{C}}$ and $\mathrm{FO}(\Delta_k)^{\mathcal{C}} \subseteq \mathrm{FO}(\Sigma_k)^{\mathcal{C}} \cap \mathrm{FO}(\Pi_k)^{\mathcal{C}}$ for every class $\mathcal{F}$ of formulas.

Remark 6. The definition of the class $\underline{\mathrm{FO}(\Delta_k)}(\mathcal{C})$ of $\mathcal{C}$-structures required some extra work because we had to introduce the notion of equivalence over $\mathcal{C}$ in order to define the class ${\Delta_k}^{\mathcal{C}}$ of formulas afterwards, which is in turn needed to define the class $\mathrm{FO}(\Delta_k)^{\mathcal{C}}$ of formulas. This was necessary because I want to formulate the separation results for the class of grids in the strongest possible way and thus consider this class $\mathrm{FO}(\Delta_k)^{\mathcal{C}}$, which might be less expressive than $\mathrm{FO}(\Sigma_k)^{\mathcal{C}} \cap \mathrm{FO}(\Pi_k)^{\mathcal{C}}$. All the separation results mentioned in the introduction could have been formulated without the notion of "equivalence over $\mathcal{C}$".

2.4 Results

Our main result is the following:

Theorem 7. *If $\mathcal{C}$ is the class of grids, graphs, or undirected graphs, and $k \geq 1$, then*

1. $\underline{\Delta_{k+2}}(\mathcal{C}) \not\subseteq \underline{\mathrm{FOB}(\Sigma_k)}(\mathcal{C})$,
2. $\underline{\mathrm{closed}}(\Sigma_1)(\mathcal{C}) \not\subseteq \underline{\mathrm{FOB}(\Sigma_k)}(\mathcal{C})$,
3. $\underline{\mathrm{FO}(\Sigma_k)}(\mathcal{C}) \cap \underline{\mathrm{FO}(\Pi_k)}(\mathcal{C}) \not\subseteq \underline{\mathrm{B}(\Sigma_k)}(\mathcal{C})$,
4. $\underline{\mathrm{FOB}(\Sigma_k)}(\mathcal{C}) \subsetneq \underline{\mathrm{FO}(\Sigma_{k+1})}(\mathcal{C}) \cap \underline{\mathrm{FO}(\Pi_{k+1})}(\mathcal{C})$.

If $\mathcal{C}$ is the class of grids, then

5. $\underline{\mathrm{FO}(\Delta_k)}(\mathcal{C}) \not\subseteq \underline{\mathrm{B}(\Sigma_k)}(\mathcal{C})$,
6. $\underline{\mathrm{FOB}(\Sigma_k)}(\mathcal{C}) \subsetneq \underline{\mathrm{FO}(\Delta_{k+1})}(\mathcal{C})$.

This reproves and strengthens [AFS97, Theorem 1.2], which says that there is an undirected graph property definable in $\mathrm{closed}(\Sigma_1)$ but not in $\mathrm{FOB}(\Sigma_1)$, and [AFS97, Theorem 1.1], which says that there is a graph property definable in $\mathrm{FO}(\Sigma_1)$ but not in $\mathrm{B}(\Sigma_1)$.

On the other hand, the following is shown in [MST98].

Theorem 8. *If $\mathcal{C}$ is the class of graphs or undirected graphs and $k \geq 1$,*

$$\underline{\mathrm{FO}(\Sigma_1)}(\mathcal{C}) \not\subseteq \underline{\Sigma_k}(\mathcal{C}).$$

Intuitively, this means that the first-order closure is so powerful that it lifts even the lowest level arbitrarily high in the monadic hierarchy over *graphs*. (I conjecture that this in not the case for *grids*.)

For the proof of Theorem 7 we will, as in [MT97,Sch97], consider the growth rate of functions definable in certain formula classes, where definability of functions is defined as follows.

Definition 9. *For a relation $r \subseteq \mathbb{N}^2$ we denote by L_r the set of grids $\underline{[m] \times [n]}$ with $(m,n) \in r$. We say that a sentence φ over τ_{Grid}* defines *the relation $r \subseteq \mathbb{N}^2$ iff $Mod_{0,0}(\varphi) = L_r$. A relation r is $\mathcal{F}$-definable (for a class of formulas $\mathcal{F}$ over τ_{Grid}), if there is a sentence in $\mathcal{F}$ that defines the set L_r of grids.*

If a relation f is a function, we use the usual notations, for example $f(m)$ for the element n with $(m,n) \in f$.

Example 10. Example 4 shows that the function $m \mapsto m$ is Δ_1-definable.

Definition 11 (k-fold Exponential Functions and Formula Classes). *Let $k \geq 1$. A function $f: \mathbb{N} \to \mathbb{N}$ is called* at most k-fold exponential *(respectively* at least k-fold exponential*, or* k-fold exponential*) if $f(m)$ is $s_k(\mathcal{O}(m))$ (respectively $s_k(\Omega(m))$, or $s_k(\Theta(m))$), where $s_0(m) = m$ and $s_{k+1}(m) = 2^{s_k(m)}$ for every $m \geq 1$, $k \geq 0$.*

Let $\mathcal{F}$ be a class of formulas over τ_{Grid}. We say that

- *$\mathcal{F}$ is* at most k-fold exponential *iff every $\mathcal{F}$-definable function is at most k-fold exponential;*
- *$\mathcal{F}$ is* at least k-fold exponential *iff some $\mathcal{F}$-definable function is at least k-fold exponential;*
- *$\mathcal{F}$ is* k-fold exponential *iff $\mathcal{F}$ is both at least and and most k-fold exponential.*

Example 12. Example 4 shows that Δ_1 is at least 0-fold exponential. In [Gia94], the author shows that Σ_1 is 1-fold exponential. This means that every Σ_1-definable function is at most singly exponential, and there is one singly exponential Σ_1-definable function (namely $m \mapsto 2^m$).

Note that, of course, there are other Σ_1-definable functions which are not at least singly exponential. For example, every constant function is Σ_0-definable and hence Σ_1-definable.

All of the separation results of Theorem 7 are applications of the following simple lemma.

Lemma 13 (Growth Argument). *Let $\mathcal{F}$, $\mathcal{F}'$ be classes of formulas. If $\mathcal{F}$ is at least $(k+1)$-fold exponential and $\mathcal{F}'$ is at most k-fold exponential, then $\underline{\mathcal{F}} \not\subseteq \underline{\mathcal{F}'}$.*

The following is shown in [MST98].

Theorem 14. *For every $k \geq 1$, the classes Δ_k, Σ_k, Π_k, and $\mathrm{B}(\Sigma_k)$ are k-fold exponential.*
closed(Σ_1) *is at least k-fold exponential for every k.*

The first claim says that Δ_k is at least k-fold exponential (see [Sch97]) and that $\mathrm{B}(\Sigma_k)$ is at most k-fold exponential (see [MT97]). It implies (see [MT97,MST98]):

Corollary 15. *If $\mathcal{C}$ is the class of grids or the class of graphs, then $\underline{\mathrm{B}(\Sigma_k)}(\mathcal{C}) \subsetneq \underline{\Delta_{k+1}}(\mathcal{C})$ for every $k \geq 1$.*

The essential contribution of this paper is:

Theorem 16. $\mathrm{FO}(\Delta_k)$, $\mathrm{FO}(\Sigma_k)$, $\mathrm{FO}(\Pi_k)$, *and* $\mathrm{FOB}(\Sigma_k)$ *are $(k+1)$-fold exponential.*

Similarly as in the case of Theorem 14, it suffices to show the following two facts:

Theorem 17. $\mathrm{FOB}(\Sigma_k)$ *is at most $(k+1)$-fold exponential.*

Theorem 18. $\mathrm{FO}(\Delta_k)$ *is at least $(k+1)$-fold exponential.*

The proof of Theorem 17 will be done in Section 3, the proof of Theorem 18 in Section 4.

For the case $k = 1$, Theorem 17 reproves another result of [AFS97], namely that $\mathrm{FOB}(\Sigma_1)$ is at most 2-fold exponential, i.e., every $\mathrm{FOB}(\Sigma_1)$-definable function is at most 2-fold exponential (Theorem 12.4), and Theorem 18 answers an open question of that paper, whether $\mathrm{FOB}(\Sigma_1)$-definable functions can be more than singly exponential.

For the class of grids, the six claims of Theorem 7 are immediate from Lemma 13 and Theorems 14 and 16. (Claims 3 and 4 are immediate from claims 5 and 6, respectively.) The first four claims for the class of graphs or undirected graphs follow by an encoding technique [MST98, Corollary 34, Proposition 35]: the class of grids is "strongly first-order reducible" to the class of graphs, which is a sufficient criterion to transfer non-inclusion results like the first four

claims of Theorem 7. The same technique has been used implicitly in [MT97] and mentioned in [AFS97].

The last two claims of Theorem 13 (cf. Remark 6) can also be transferred to the class of graphs, but this requires some extra work not presented here, especially for the case $k = 1$. The reason for these complications is the dependency on the definition of $\mathrm{FO}(\Delta_k)^{\mathcal{C}}$ of the underlying class $\mathcal{C}$: when one passes, for instance, from the class $\mathcal{C}$ of grids to the larger class $\mathcal{C}'$ of all τ_{Grid}-structures, the induced logical equivalence on the class of formulas gets finer and thus $\mathrm{FO}(\Delta_1)^{\mathcal{C}}$ need not be a subset of $\mathrm{FO}(\Delta_1)^{\mathcal{C}'}$.

3 Adapting the Periodicity Argument to First-Order Closures

The proof of the upper bound of $\mathrm{B}(\Sigma_k)$-definable functions in [MT97] was, roughly speaking, as follows: To a set L of grids and a fixed height $m \geq 1$ we associate the set of words a^n over a singleton alphabet $\{a\}$ with (m, n) is the size of a grid in L. We showed in [MT97] that if L is definable in Σ_k, then $L[m]$ if recognizable by a unary NFA whose state set size is at most k-fold exponential in m. Finally, one exploits periodicity properties of NFAs.

We wish to adapt this technique to the first-order closure of a level of the monadic hierarchy over grids. Suppose we have a $\mathrm{FO}(\Sigma_k)$-formula $\exists x_1 \forall x_2 \varphi(x_1, x_2)$, where φ is in Σ_k. Then we may associate to $\varphi(x_1, x_2)$ the set of "2-pebbled grids" (i.e., grids with two differently marked positions) that are models of φ under the assignment given by the pebbles. If m is, again, a fixed height, we may pass to a word language $L[m]$ as before. But this time, $L[m]$ will not contain words over a *singleton* alphabet, but words in which all but two positions (corresponding to the columns of the two pebbles) carry the same letter. Again, $L[m]$ is recognizable by an NFA whose state set size is at most k-fold exponential in m. Since in input words for this NFA all but two positions carry the same letter, there are similar periodicity properties as for unary NFA. We will exhibit what happens to these periodicity properties when taking the first-order quantifications of x_1 and x_2 into account.

3.1 Preliminaries

Consider the following two functions:

Definition 19. *For every $m \geq 1$, let*

$$F(m) = \max\{\mathrm{lcm}\{x_1, \dots, x_k\} \mid k \geq 1, \sum_i x_i = m\},$$

$$H(m) = 2^{\sqrt{m \log m}}.$$

The function F appears in Chrobak's construction ([Chr86]) for a normal form of non-deterministic finite automata. We will redo this construction in the next subsection.

Recall that $s_0(m) = m$ and $s_{k+1}(m) = 2^{s_k(m)}$ for every $m \geq 1$, $k \geq 0$. What we need to know about F is:

Proposition 20. *$F(m)$ and $H(m)$ are $s_1(m^{\Theta(1)})$.*

Proof. $F(m)$ is $\Theta(H(m))$. (Much more precise approximations are given in [Lan09] and in [Sza80], see also [Chr86]). Besides, $H(m)$ is $\Omega(s_1(m))$ and $\leq s_1(m^2)$. Thus $H(m)$ and $F(m)$ are $s_1(m^{\Theta(1)})$.

3.2 Automata and Periodicity wrt. a Certain Letter

Let alphabet Σ be an alphabet and $a \in \Sigma$ be some letter. The set of all words over Σ is denoted Σ^*.

Let $n, p \geq 1$. A language $L \subseteq \Sigma^*$ is *periodic (with threshold n, period p, wrt. a)* iff for every $u, v \in \Sigma^*$: $ua^n v \in L \iff ua^{n+p}v \in L$. By (p,n)-*periodic* we mean "periodic with threshold n and period p". The explicit mentioning of a will be dropped if it is clear from the context.

A set $N \subseteq \mathbb{N}$ will be called (p,n)-periodic iff $\{a^n \mid n \in N\}$ is (p,n)-periodic, i.e., if $n' \in N \leftrightarrow n' + p \in N$ for every $n' \geq n$.

Remark 21. Every (p,n)-periodic language is (p',n')-periodic for every multiple p' of p and every $n' \geq n$. Every boolean combination of (p,n)-periodic languages is (p,n)-periodic.

If $N \subseteq \mathbb{N}$ is (p,n)-periodic, and $n' \geq n$, then $\{k \mid k + n' \in N\}$ is $(p,0)$-periodic.

Definition 22 ([Chr86]). *Let $\mathfrak{A}$ be a unary NFA, i.e. with singleton input alphabet. By the words "path", "cycle", "strongly connected component" etc. we refer to the graph structure of $\mathfrak{A}$. Let p be the number of states of cycles and n be the number of states not on cycles. Then $S(\mathfrak{A}) = (p,n)$.*

Remark 23. A language $L \subseteq \{a\}^*$ is (p,n)-periodic iff there is a DFA $\mathfrak{A}$ with $S(\mathfrak{A}) = (p,n)$ that recognizes L.

The following theorem has been used in [MT97]. We mention it here though we will have to prove a stronger version of it.

Theorem 24. *([Chr86, Lemma 4.3, Theorem 4.4]) Every unary NFA with n-states is equivalent to an NFA $\mathfrak{A}$ with $S(\mathfrak{A}) \leq (n, \mathcal{O}(n^2))$ such that no state on a cycle has more than one successor.*

Every unary NFA with n-states is equivalent to a DFA with $\mathcal{O}(F(n))$ states.

As explained in the introduction of this section, we have to investigate periodicity properties for NFAs whose input consists of words most of whose letters are identical. Our aim is to show that when such an NFA reads an input, then in the "gaps" between the designated positions it behaves very much like a unary NFA. This is stated in Lemma 26, which is prepared by the following definition.

Definition 25. *Let $a \in \Sigma$. Let $\mathfrak{A} = (Q, \Sigma, q_0, \Delta, F)$ and $\mathfrak{A}' = (Q', \{a\}, q'_0, \Delta', \emptyset)$ be NFAs. Then $\mathfrak{A}'$* simulates *$\mathfrak{A}$* wrt. *a iff there is a family $(F_{q,q'})_{q,q' \in Q}$ of subsets of Q' such that*

$$\forall k \geq 0: \quad \mathfrak{A} : q \xrightarrow{a^k} q' \iff \mathfrak{A}' : q'_0 \xrightarrow{a^k} F_{q,q'}.$$

Lemma 26. *Let $\mathfrak{A}$ be an NFA with n states and a be a letter from its input alphabet. Then there is $p \leq n$ and a unary NFA $\mathfrak{A}'$ with $S(\mathfrak{A}') = (p, (n+1)^2)$ such that $\mathfrak{A}'$ simulates $\mathfrak{A}$ wrt. a and no state of $\mathfrak{A}'$ on a cycle has more than one successor.*

This a slight modification of the first statement of Theorem 24. The proof is almost the same as for [Chr86, Lemma 4.3]. The essential observation is that the transition structure of the NFA constructed there does not depend on the initial and final state, so it can be carried out uniformly for all pairs of initial/final state. The original proof, too, gives these accurate bounds on the numbers of states (rather than the asymptotic ones stated in [Chr86, Lemma 4.3]).

The proof of Lemma 26 depends on the following number theoretic fact, which is stated in [Chr86] and follows from results of [Bra42,Liu64].

Lemma 27. *Let* $a_1, \ldots, a_k \in \{1, \ldots, n\}$. *Let* $X = \{\sum_{i=1}^{k} a_i x_i \mid x_1, \ldots, x_k \geq 0\}$.

Then $X \cap \{x \mid x \geq n^2\}$ *is an arithmetic progression with period* $\gcd\{a_1, \ldots, a_k\}$. *In particular,* X *is* $(\gcd\{a_1, \ldots, a_k\}, n^2)$*-periodic.*

Proof (of Lemma 26, cf. [Chr86]). Since the transitions with labels different from a are of no concern, we may assume that $\mathfrak{A} = (Q, \{a\}, q_0, \Delta, F)$.

Let $T_1, \ldots, T_k$ be the SCCs of $\mathfrak{A}$. For every $i \leq k$ let y_i be the greatest common divisor of the length of all cycles of T_i. Let $m = n^2 + n$.

Choose $\mathfrak{A}' = (Q', \{a\}, 0, \Delta', \emptyset)$, where

- $Q' = \{0, \ldots, m-1\} \cup \bigcup_{i \leq k} (\{i\} \times \{0, \ldots, y_i - 1\})$,
- $\Delta' = \{(i, a, i+1) \mid 0 \leq i < m-1\}$
  $\cup \{((i,j), a, (i, (j+1) \mod y_i)) \mid i \leq k \wedge j < y_i\}$
  $\cup \{(m-1, a, (i,0)) \mid i \leq k\}$

Then on has

- for every $x < m$, $\mathfrak{A}' : 0 \xrightarrow{a^x} q$ iff $q = x$, and
- for every $x \geq m$, $\mathfrak{A}' : 0 \xrightarrow{a^x} (i, s)$ iff $s = ((x - m) \mod y_i)$.

The number of states on cycles is $p := \sum_i y_i \leq \sum_i |T_i| = |\bigcup_i T_i| \leq n$, and the number of states that are not on cycles is $m = n^2 + n < (n+1)^2$.

We wish to show that $\mathfrak{A}'$ simulates $\mathfrak{A}$ wrt. a.

Let $q, q' \in Q$. Choose $F_{q,q'} = \{x < m \mid \mathfrak{A} : q \xrightarrow{a^x} q'\} \cup \{(i, x) \in Q \mid \mathfrak{A} : q \xrightarrow{a^{x+m}} q'\}$.

For $q, q' \in Q$, a (q, q')-superpath is a sequence of the form

$$\alpha = P_1 D_1 P_2 D_2 \ldots P_t D_t P_{t+1}$$

where

- $D_1, \ldots, D_t$ are SCCs of $\mathfrak{A}$,
- $P_1, \ldots, P_{t+1}$ are paths in $\mathfrak{A}$ whose inner points do not belong to SCCs,
- the first vertex of P_1 is q, the last vertex of P_{t+1} is q',
- for every i, the last vertex of P_i belongs to D_i,
- for every i, the first vertex of P_i belongs to D_{i-1}.

For every superpath α, let $L_\alpha = \{x \mid x$ is the length of a path in $\alpha\}$.

For every SCC D of $\mathfrak{A}$ let $\Pi_{q,q'}(D)$ be the set of (q, q')-superpaths α such that D is the last SCC of α.

Let $\alpha \in \Pi_{q,q'}(T_i)$ and let $a_1, \ldots, a_l$ be the length of cycles (in the SCCs) of α. Then $L_\alpha = \{\sum_i a_i x_i \mid x_i \geq 0\}$ is $(\gcd\{a_1, \ldots, a_l\}, n^2)$-periodic and hence (because $\gcd\{a_1, \ldots, a_l\} \mid y_i$) also (y_i, n^2)-periodic. Thus $\{x \geq 0 \mid x + m \in L_\alpha\}$ is $(y_i, 0)$-periodic for every $i \leq k$, $\alpha \in \Pi_{q,q'}(T_i)$.

For every $x \in \{0, \ldots, m-1\}$ obviously $\mathfrak{A}' : 0 \xrightarrow{a^x} F_{q,q'}$ iff $\mathfrak{A} : q \xrightarrow{a^x} q'$.

If $x \geq m$, then for every $i \leq k$, $\alpha \in \Pi_{q,q'}(T_i)$ one has $x \in L_\alpha$ iff $x - m \in \{x' \geq 0 \mid x' + m \in L_\alpha\}$ iff (because that set is $(y_i, 0)$-periodic) $((x - m) \mod y_i) \in L_\alpha$

Thus for every $x \geq m$, one has $\mathfrak{A} : q \xrightarrow{a^x} q'$ iff $\exists i \leq k : x \in \bigcup_{\alpha \in \Pi_{q,q'}(T_i)} L_\alpha$ iff $\exists i \leq k : ((x - m) \mod y_i) \in L_\alpha$ iff $\mathfrak{A}' : 0 \xrightarrow{a^x} F_{q,q'}$.

This completes the proof.

An automaton whose transition structure is like in the above proof is often called “in Chrobak Normal Form”.

Corollary 28. *Let* $\mathfrak{A}$ *be an NFA with* n *states. Then there is* $p \leq F(n)$ *such that for all states* q, q' *of* $\mathfrak{A}$,

$$\mathfrak{A} : q \xrightarrow{a^{(n+1)^2}} q' \iff \mathfrak{A} : q \xrightarrow{a^{(n+1)^2+p}} q'.$$

Proof. Let $\mathfrak{A} = (Q, \Sigma, q_0, \Delta, F)$ be an NFA that recognizes the language L, where $|Q| = n$. By Lemma 26 there is an NFA $\mathfrak{A}' = (Q', \{a\}, q_0', \Delta, \emptyset)$ and $p' \leq n$ with $S(\mathfrak{A}') = (p', (n+1)^2)$ such that no state on a cycle of $\mathfrak{A}'$ has more than one successor and $\mathfrak{A}'$ simulates $\mathfrak{A}$ wrt. a. Choose a family $(F_{q,q'})_{q,q' \in Q}$ of subsets of Q' according to Definition 25, i.e.,

$$\forall k \geq 0 : \quad \mathfrak{A} : q \xrightarrow{a^k} q' \iff \mathfrak{A}' : q_0' \xrightarrow{a^k} F_{q,q'}.$$

Let p be the least common multiple of the length of cycles in $\mathfrak{A}'$. Since the cycles in $\mathfrak{A}'$ have pairwise disjoint state sets, $p \leq F(p') \leq F(n)$. Every path through $\mathfrak{A}'$ of length greater $(n+1)^2$ passes a state on a cycle. This implies the claim.

Theorem 29. *Let $a \in \Sigma$. For every language $L \subseteq \Sigma^*$ that is recognized by an NFA with n states, there is a $p \leq F(n)$ such that L is $(p, (n+1)^2)$-periodic wrt. a.*

Proof. Let $\mathfrak{A}$ be an NFA with n states that recognizes L. Choose $p \leq F(n)$ according to Corollary 28. Then $ua^{(n+1)^2}v \in L$ iff $\exists q_1, q_2 \in Q$ $\exists q_3 \in F : q_0 \xrightarrow{u} q_1 \xrightarrow{a^{(n+1)^2}} q_2 \xrightarrow{v} q_3$ iff $\exists q_1, q_2 \in Q\, \exists q_3 \in F : q_0 \xrightarrow{u} q_1 \xrightarrow{a^{(n+1)^2+p}} q_2 \xrightarrow{v} q_3$ iff $ua^{(n+1)^2+p}v \in L$

The following corollary strengthens the second claim of Theorem 24.

Corollary 30. *Let $\mathfrak{A}$ be a unary NFA with n-states. Then there is a $p \leq F(n)$ and a DFA $\mathfrak{A}'$ with $S(\mathfrak{A}') = (p, (n+1)^2)$ that simulates $\mathfrak{A}$.*

Proof. The claim follows from Theorem 29 and Remark 23 .

The following definition mimics the effect of first-order quantifications on the level of formal languages.

Definition 31. *Let Σ, Σ' be alphabets, $\alpha_0, \alpha_1 \subseteq \Sigma \times \Sigma'$, where α_0 is a total, injective function. The relation $R_{(\alpha_0,\alpha_1)}$ associated to (α_0, α_1) is given by $\{((a_1, \ldots, a_n), (a_1', \ldots, a_n')) \mid \exists i((a_i, a_i') \in \alpha_1 \wedge \forall j \neq i((a_j, a_j') \in \alpha_0))\}$. Let $R = R_{(\alpha_0,\alpha_1)}$.*

$$\langle R \rangle L := \{w \in \Sigma^* \mid \exists w' \in \Sigma'^* : (w, w') \in R \wedge w' \in L\},$$
$$[R]L := \{w \in \Sigma^* \mid \forall w' \in \Sigma'^* : (w, w') \in R \rightarrow w' \in L\}.$$

Lemma 32. *Let Σ, Σ' be alphabets, $\alpha_0, \alpha_1 \subseteq \Sigma \times \Sigma_1$ and α_0 total, functional and injective, $(a, a') \in \alpha_0$. Let $R = R_{(\alpha_0,\alpha_1)}$.*
Let $L \subseteq \Sigma'^*$ be (p, n)-periodic wrt. a'.
Then $\langle R \rangle L$ and $[R]L$ are $(p, 2n + p)$-periodic wrt. a.

Proof. Since $\Sigma^* \setminus ([R]L) = \langle R \rangle (\Sigma'^* \setminus L)$ and a language is (p, n)-periodic iff its complement is, it suffices to show that $\langle R \rangle L$ is $(p, 2n + p)$-periodic.

We have to show that $ua^{2n+p}v \in \langle R \rangle L \iff ua^{2n+2p}v \in \langle R \rangle L$ for every $u, v \in \Sigma^*$.

So let $ua^{2n+p}v \in \langle R \rangle L$. Then there are $u', x', v' \in \Sigma'^*$ such that $|u| = |u'|$, $|v| = |v'|$, and $|x'| = 2n + p$ such that $ua^{2n+p}v \; R \; u'x'v' \in L$.

If $x' = a'^{2n+p}$, then $ua^{2n+2p}v \; R \; u'a'^{2n+2p}v' \in L$ because L is (p, n)-periodic.

If $x' = a'^i a_1 a'^j$ for some a_1 with $(a, a_1) \in \alpha_1$ and $i+j+1 = 2n+p$, then either $i \geq n$ or $j \geq n$. Assume that $i \geq n$ (the other case is analogous). Then $ua^{2n+2p}v \; R \; u'a'^{i+p}a_1a'^jv' \in L$.

So in every case $ua^{2n+2p}v \in \langle R \rangle L$.

Conversely, let $ua^{2n+2p}v \in \langle R \rangle L$. There are $u', v', x' \in \Sigma'$ with $|u| = |u'|$, $|v| = |v'|$, and $|x'| = 2n + 2p$ and $ua^{2n+2p}v \; R \; u'x'v' \in L$.

If $x' = a'^{2n+2p}$, then $ua^{2n+p}v \; R \; u'a'^{2n+p}v' \in L$ as before.

If $x' = a'^i a_1 a'^j$ for some a_1 with $(a, a_1) \in \alpha_1$ and $i + j + 1 = 2n + 2p$, then either $i \geq n + p$ or $j \geq n + p$. Assume that $i \geq n + p$ (the other case is analogous). Then $ua^{2n+p}v \; R \; u'a'^{i-p}a_1a'^jv' \in L$.

So in every case $ua^{2n+p}v \in \langle R \rangle L$. This completes the proof of Lemma 32.

A technicality of the above proof is similar to an idea that is implicit in the proof of [AFS97, Lemma 12.6].

3.3 From Words to Grids

The following definition is, again, similar to [MT97,MST98] and performs a link to automata theory. The idea is to associate a word to every (t,r)-grid such that each letter of the word corresponds to a column of the (t,r)-grid and encodes this column's "colour" and "pebbles".

Definition 33. *Let R be a (t,r)-grid of size (m,n). Then the column word of R is a word of length n over alphabet $(\{0,1\}^{t+r})^m$ defined as*

$$colR = \begin{pmatrix} \overline{a}_{11} \\ \vdots \\ \overline{a}_{m1} \end{pmatrix} \cdots \begin{pmatrix} \overline{a}_{1n} \\ \vdots \\ \overline{a}_{mn} \end{pmatrix},$$

where, for $(i,j) \in [m] \times [n]$, the word $\overline{a}_{ij} = a^1_{ij} \dots a^t_{ij} a^{t+1}_{ij} \dots a^{t+r}_{ij}$ over $\{0,1\}$ is such that

- $a^s_{ij} = 1$ *iff* $(i,j) \in X^R_s$ *for all* $s \le t$,
- $a^s_{ij} = 1$ *iff* $(i,j) = x^R_{s-t}$ *for all* $t+1 \le s \le t+r$.

For a class of (t,r)-grids L and $m \ge 1$, the height-m fragment of L *is given by $L[m] = \{colR \mid R$ is of height m and $R \in L\}$.*

In [MT97] the following is shown:

Theorem 34. *Let $k \ge 1$, $t \ge 0$, and $\varphi(X_1,\dots,X_t)$ be a Σ_k-formula over τ_{Grid}. There exists $c \ge 1$ such that for every $m \ge 1$ there is an NFA with $s_{k-1}(c^m)$ states that recognizes the word language $Mod_{t,0}(\varphi)[m]$ over $(\{0,1\}^t)^m$.*

In order to adapt this result to first-order/boolean closures, we observe that, informally, we may replace free first-order variables by second-order variables, so that we get the following:

Corollary 35. *Let $k \ge 1$, $r \ge 0$, and $\varphi(x_1,\dots,x_r)$ be a Σ_k-formula over τ_{Grid}. There is a $c \ge 1$ such that for every $m \ge 1$ there is an NFA with $s_{k-1}(c^m)$ states that recognizes the word language $Mod_{0,r}(\varphi)[m]$ over $(\{0,1\}^r)^m$.*

The proof of this corollary is immediate when we have the following lemma:

Lemma 36. *For every Σ_k-formula $\varphi(x_1,\dots,x_r)$ there is Σ_k-formula $\varphi'(X_1,\dots,X_r)$ with*

$$Mod_{r,0}(\varphi') = \{(R,\{u_1\},\dots,\{u_r\}) \mid (R,u_1,\dots,u_r) \in Mod_{0,r}(\varphi)\},$$

In particular, $Mod_{r,0}(\varphi')[m] = Mod_{0,r}(\varphi)[m]$ for every m.

Proof. Consider the first-order formula $sing(X) = \exists x \forall y (X(y) \leftrightarrow x = y)$ that says that the set X is a singleton. Using this formula, we can construct a formula $\varphi'(X_1,\dots,X_r)$, namely φ' results from φ by building the conjunction of the formula $\bigwedge_{i=1}^s sing(X_i)$ with a formula that results from φ by replacing atomic formulas according to the following rules:

- $Y(x_s)$ is replaced with $\forall x (X_s(x) \to Y(x))$ for every monadic second-order variable Y;
- $x_2 = y$ is replaced with $\forall x (X_s(x) \leftrightarrow x = y)$ for every $s \le r$ and every first-order variable y not among $\{x_1,\dots,x_r\}$.
- $x_s = x_{s'}$ is replaced with $\forall x (X_s(x) \leftrightarrow X_{s'}(x))$ for every $s, s' \le r$;
- $S_i(x_s, x_{s'})$ is replaced with $\forall x \forall y (X_s(x) \wedge X_{s'}(y) \to S_i(x,y))$ for every $s, s' \le r$ and $i \in \{1,2\}$.

The next lemma states that, informally, the operators $\langle R \rangle$ and $[R]$ of Definition 31 relate to the height-m fragment the same way that existential and universal first-order quantifications relate to the set of models.

Lemma 37. *Let $\Sigma = (\{0,1\}^r)^m$ and $\Sigma' = (\{0,1\}^{r+1})^m$ and*

$$\alpha_0 = \left\{ \left(\begin{pmatrix} \overline{a_1} \\ \vdots \\ \overline{a_m} \end{pmatrix}, \begin{pmatrix} \overline{a_1}0 \\ \vdots \\ \overline{a_m}0 \end{pmatrix} \right) \middle| \; \overline{a_i} \in \{0,1\}^r \right\}$$

$$\alpha_1 = \left\{ \left(\begin{pmatrix} \overline{a_1} \\ \vdots \\ \overline{a_m} \end{pmatrix}, \begin{pmatrix} \overline{a_1}b_1 \\ \vdots \\ \overline{a_m}b_m \end{pmatrix} \right) \middle| \; \overline{a_i} \in \{0,1\}^r,\ b_1 \dots b_m \in 0^*10^* \right\}$$

Let $R = R_{(\alpha_0,\alpha_1)}$, $\varphi(x_1,\dots,x_r)$ be a first-order formula and $m \geq 1$. Then

$$Mod_{0,r}(\exists x_{r+1}\varphi)[m] = \langle R\rangle(Mod_{0,r+1}(\varphi)[m])$$
$$Mod_{0,r}(\forall x_{r+1}\varphi)[m] = [R](Mod_{0,r+1}(\varphi)[m])$$

Proof. Let

$$w = \begin{pmatrix} \overline{a_{11}} \\ \vdots \\ \overline{a_{m1}} \end{pmatrix} \cdots \begin{pmatrix} \overline{a_{1n}} \\ \vdots \\ \overline{a_{mn}} \end{pmatrix}$$

(where $\overline{a_i j} \in \{0,1\}^r$) be in $Mod_{0,r}(\exists x_{r+1}\varphi)[m]$.

Then there exists $(0,r)$-grid $A = (\underline{[m]\times[n]}, x_1,\dots,x_r)$ such that $A \models \exists x_{r+1}\varphi$. Thus there is $x_{r+1} = (\hat{\imath},\hat{\jmath}) \in [m]\times[n]$ with $A,(i,j) \models \varphi$. Choose

$$\overline{a'_{ij}} = \begin{cases} \overline{a_{ij}}0 \text{ if } (i,j) \in [m]\times[n] \setminus \{(\hat{\imath},\hat{\jmath})\} \\ \overline{a_{ij}}1 \text{ if } (i,j) = (\hat{\imath},\hat{\jmath}) \end{cases}$$

Then

$$w = \begin{pmatrix} \overline{a_{11}} \\ \vdots \\ \overline{a_{m1}} \end{pmatrix} \cdots \begin{pmatrix} \overline{a_{1n}} \\ \vdots \\ \overline{a_{mn}} \end{pmatrix} R \begin{pmatrix} \overline{a'_{11}} \\ \vdots \\ \overline{a'_{m1}} \end{pmatrix} \cdots \begin{pmatrix} \overline{a'_{1n}} \\ \vdots \\ \overline{a'_{mn}} \end{pmatrix} \in Mod_{0,r}(\varphi)[m],$$

thus $w \in \langle R\rangle(Mod_{0,r+1}(\varphi)[m])$. The converse direction is shown similarly, and the second claim is valid by duality. This completes the proof.

Now we can combine Corollary 35, Corollary 28, and the preceding lemma to get the following.

Lemma 38. *Let $m \geq 1$. For every $r \geq 0$ and every $Q_1,\dots,Q_r \in \{\exists,\forall\}$, if $\varphi(x_1,\dots,x_r)$ is a formula such that $Mod_{0,r}(\varphi)[m]$ is (p,n)-periodic, then $Mod_{0,0}(Q_1x_1\dots Q_rx_r\varphi)[m]$ is $(p, 2^r(n+p)-p)$-periodic.*

Proof. Induction on r. The case $r=0$ is trivial because $2^0(n+p)-p = n$. So assume the claimed implication is true for some r. Let $Q_1,\dots,Q_{r+1} \in \{\exists,\forall\}$ and $\varphi(x_1,\dots,x_{r+1})$ be a formula such that $Mod_{0,r+1}(\varphi)[m]$ is (p,n)-periodic.

Assume that $Q_{r+1} = \exists$. (The other case is dual.) Let R be as in the preceding lemma. By induction $Mod_{0,r}(\exists x_{r+1}\varphi)[m])$ is $(p, 2^r(n+p)-p)$-periodic. By Lemma 32 , $Mod_{0,r+1}(\varphi)[m] = \langle R\rangle(Mod_{0,r}(\exists x_{r+1}\varphi)[m])$ is $(p, 2\cdot(2^r(n+p)-p)+p)$-periodic, i.e. $(p, 2^{r+1}(n+p)-p)$-periodic. This completes the induction.

Lemma 39. *Let $k,r \geq 0$ and $\varphi(x_1,\dots,x_r)$ be a $\mathrm{B}(\Sigma_k)$-formula, and $Q_1,\dots,Q_r \in \{\exists,\forall\}$. Then there is some $d \geq 0$ and some $p \leq F(s_k(dm))$ such that for every $m \geq 1$,*

$$Mod_{0,0}(Q_1x_1\dots Q_rx_r\varphi)[m]$$

is $(p, s_{k+1}(dm))$-periodic.

Proof. Assume that $\varphi(x_1,\dots,x_r)$ is a boolean combination of the Σ_k-formulas $\varphi_1,\dots,\varphi_j$. By Corollary 35, for every $i \leq j$ there is some constant c_i such that for every $m \geq 1$ there is a $s_k(c_i^m)$-state NFA that recognizes language $Mod_{0,r}(\varphi_i)[m]$.

Choose $c := \max\{c_1,\dots,c_j\}$ and d such that $2^r((s_k(cm))^{2j} + F(s_k(cm))) \leq s_{k+1}(dm)$ for every m.

Let $m \geq 1$. For every i, let $\mathfrak{A}_i = (Q_i,\Sigma,q_{0,i},\Delta_i,F_i)$ be an NFA with $s_k(c^m)$ states that recognizes $Mod_{0,r}(\varphi_i)[m]$. Build the usual "cross-product automaton" $\mathfrak{A} = (Q_1\times\dots\times Q_j, \Sigma, (q_{01},\dots,q_{0j}), \Delta, \emptyset)$ with state set of size $(s_k(c^m))^j =: n$. Choose $p \leq F(n)$ according to Corollary 28, i.e. such that for every $q,q' \in Q$:

$$\mathfrak{A} : q \xrightarrow{a^{(n+1)^2}} q' \iff \mathfrak{A} : q \xrightarrow{a^{(n+1)^2+p}} q'$$

Then $Mod_{0,r}(\varphi)[m]$ is $(p,(n+1)^2)$-periodic. By Lemma 38, the set $Mod_{0,0}(Q_1x_1\dots Q_rx_r\varphi)[m]$ is $(p, (2^r((n+1)^2+p)))$-periodic. The claim follows by choice of d.

Now we can prove Theorem 17.

Proof (of Theorem 17). Let $f : \mathbb{N} \to \mathbb{N}$ be definable by a FOB(Σ_k)-sentence ψ. By Remark 5, we may assume that ψ is of the form $Q_1 x_1 \ldots Q_r x_r \varphi$, where $r \geq 0$ and φ is a B(Σ_k)-formula.

For every $m \geq 1$, we have $\{a^n \mid f(m) = n\} = L_f[m] = Mod_{0,0}(\psi)[m]$, where a is the element of $(\{0\}^r)^m$.

Choose p, d as in the above lemma, i.e., such that for all m the set $\{f(m)\}$ is $(p, s_{k+1}(dm))$)-periodic. Then this implies that $f(m) \leq s_{k+1}(dm)$ for all m.

Thus we have shown that FOB(Σ_k) is $s_{k+1}(\mathcal{O}(m))$. This completes the proof of Theorem 17.

4 Definability of Fast Growing Functions

In this section we will prove Theorem 18, i.e., we will construct a FO(Δ_k)-formula that defines a $(k+1)$-fold exponential function.

Unfortunately, we have to redo much of Schweikardt's proof ([Sch97]) of the fact that Δ_k is at least k-fold exponential, but with with very slight modifications. In order to keep this as readable as possible, we keep closely to Schweikardt's notation.[1]

Some definitions and ideas of Schweikardt's proof can themselves be found similarly in [MT97]. However, we do need Schweikardt's tricks that allow to verify the tight bounds on the number of monadic quantifier alternations.

4.1 The Fast Growing Functions

We will be concerned with the following functions.

Definition 40. *For every $m \geq 1$, let $G(m) = \mathrm{lcm}\{m, m+1, \ldots, 2m-1\}$.*

Let $f_1(m) := 2^m$, $f_{k+1}(m) := f_k(m) 2^{f_k(m)}$ for $m, k \geq 1$. Let $g_k(m) = G(f_k(m))$ for every $k, m \geq 1$.

Recall that $s_0(m) = m$ and $s_{k+1}(m) = 2^{s_k(m)}$ for every $m \geq 1$, $k \geq 0$.

[1] I even copied some of Schweikardt's LaTeX source code.

Proposition 41. *1. $G(m)$ is $s_1(\Omega(m))$.*
2. $G(m)$ is $s_1(m^{\Theta(1)})$.
3. f_k is $s_k(\Omega(m))$.
4. g_k is $s_{k+1}(\Omega(m))$.

Proof. For the first claim, let $m \geq 1$. Let $p_1, \ldots, p_k$ be the distinct prime numbers $< m$, and $p = \prod_i p_i$, and $l_i = \max\{l \mid p_i^l < m\}$ for every $i \leq k$. Then $\mathrm{lcm}\{1, \ldots, m-1\} = \prod_i p_i^{l_i} = \frac{1}{p} \prod_i p_i^{l_i+1} \geq \frac{m^k}{p}$. On the other hand, $\mathrm{lcm}\{1, \ldots, m-1\} \geq \prod_i p_i = p$. Thus $(\mathrm{lcm}\{1, \ldots, m-1\})^2 \geq p \cdot \frac{m^k}{p} = m^k$.

k is $\Omega\,(m/\log m)$, thus $G(m) = \mathrm{lcm}\{m, \ldots, 2m-1\} \geq \mathrm{lcm}\{1, \ldots, m-1\} \geq m^{k/2}$ is $m^{\Omega(m/\log m)} = 2^{\Omega(m)}$.

Concerning the second claim, we have that $G(m) = \mathrm{lcm}\{1, \ldots, 2m-1\} \leq F(\sum_{i=1}^{2m-1} i)$ is (by Proposition 20) $F(\mathcal{O}(m^2)) \leq s_1(m^{\mathcal{O}(1)})$, which, together with the first claim, implies the second.

The third claim is immediate from the definitions of f_k and s_k.

By the first claim, $g_k(m) = G(f_k(m))$ is $G(s_k(\Omega(m)) = s_{k+1}(\Omega(m))$, i.e. the last claim.

In [Sch97], Schweikardt shows that f_k is Δ_k-definable. We will show that g_k is FO(Δ_k)-definable.

4.2 Some Preparation for the Formulas

Let $m, n \geq 1$. A *rectangle section* of $[m] \times [n]$ is a set of the form $[m] \times [i, j]$ for some $1 \leq i \leq j \leq n$. The width of that rectangle section is $j - i + 1$.

We may consider the restriction of the (t, r)-grid with domain $[m] \times [n]$ to a rectangle section $Z \subseteq [m] \times [n]$, which is defined as usual and is again a grid[2].

[2] Strictly speaking, we run into troubles here because we required the domain of a grid to be of the form $[m] \times [n]$ for some n, in particular $(1, 1)$ is in the domain of any grid. However, this requirement was only made for notational convenience. The restriction of a grid to a rectangle section is at least isomorphic to a grid.

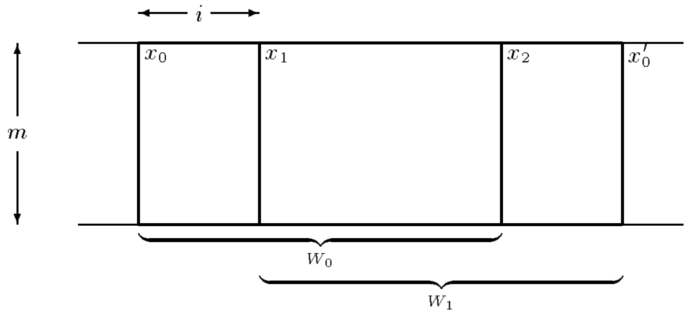


Fig. 1. A Twin Block with Offset i

Let $X \subseteq [1]\times[n]$. A position $(1,j') \in [1]\times[n]$ is the X-*successor* of a position $(1,j) \in [1]\times[n]$ iff $j \leq j'$ and $(\{1\}\times[j+1,j'-1])\cap X = \emptyset$ and $(1,j),(1,j') \in X$.

A triple (Z_0, Z_1, Z_2) of subsets of $\{1\} \times [n]$ is a *shuffle triple (in* $\underline{[m] \times [n]}$*)* iff $(1,1) \in Z_0$ and whenever x' is the $(Z_0 \cup Z_1 \cup Z_2)$-successor of x we have $x \in Z_l \iff x' \in Z_{l+1}$, (index modulo 3). Note that e.g. $Z_0 = Z_1 = Z_2$ is not forbidden.

A shuffle triple (Z_0, Z_1, Z_2) is complete iff the right-most element of $Z_0 \cup Z_1 \cup Z_2$ is in Z_2.

In the following, let (Z_0, Z_1, Z_2) be a shuffle triple.

A *twin (of* (Z_0, Z_1, Z_2)*)* is a pair (x_0, x_1), where $x_0 \in Z_0$, $x_1 \in Z_1$, and x_1 is the $(Z_0 \cup Z_1)$-successor of x_0. The *offset* of a pair $((1,i),(1,j))$ is the difference $j-i$.

A shuffle triple is *equidistant* if every twin has the same offset.

If $x = (1,i) \in Z_l$ (where $l \in \{0,1,2\}$) and $j \in [i+1,n]$ is maximal with $[i+1,j] \cap Z_{l+2} = \emptyset$ (index modulo 3), then $[m] \times [i,j]$ is the rectangle section *spanned* by x (wrt. (Z_0, Z_1, Z_2)).

A *first block* (respectively *second block*) of (Z_0, Z_1, Z_2) is a rectangle section that is spanned by some element of Z_0 (respectively of Z_1).

Figure 1 shows a rectangle section of some grid and illustrates these definitions: If (Z_0, Z_1, Z_2) is a shuffle triple and $x_0, x'_0 \in Z_0$,

$x_1 \in Z_1$, and $x_2 \in Z_2$, then (x_0, x_1) is a twin with offset i. The rectangle section W_0 (respectively W_1) is a first (respectively second) block.

How to Read Formulas. All formulas will be in the fixed signature $\tau_{\text{Grid}} = \{S_1, S_2\}$.

Let $X_1, X_2, \ldots$ be monadic second-order variables. We write $\overline{X}_k$ for the variable tuple $(X_1, \ldots, X_k)$. We allow ourselves to mix syntax and semantics when we talk about variables and their values, as we already did in Example 1. For example, we may say "the formula $\varphi(x,y)$ asserts for a position x in the first column that y is in the second column" and mean that for every $m,n \geq 1$ and every $u,v \in [m] \times [n]$ with $u \in [m] \times \{1\}$ we have $\underline{[m] \times [n]} \models \varphi[u,v]$ iff $v \in [m] \times \{2\}$.

When $\varphi(\overline{X}_t, \overline{x}_r)$ is some formula we often need the formula *rel*-$\varphi(\overline{X}_t, \overline{x}_r, Z)$ that results from φ by relativization of first-order quantifications to the (fresh) variable Z, i.e., by replacing every existential first-order quantification $\exists y(\ldots)$ with $\exists y(Zx \wedge \ldots)$ and every universal first-order quantification $\forall y(\ldots)$ with $\forall y(Zx \rightarrow \ldots)$. When R is some (t,r)-grid and Z is some rectangle section of dom$R$, then *rel*-$\varphi(\overline{X}_t, \overline{x}_r, Z)$ asserts for the restriction of the R to Z the same as $\varphi(\overline{X}_t, \overline{x}_r)$ asserts for R.

4.3 Counting in Grids

The key idea for the definition of fast growing function is to establish some counting mechanism is grids. This idea has first been presented in [MT97], then improved in [Sch97], and will now be developed further. We start with some definitions that clarify this mechanism.

The following definitions are from [Sch97,MST98].

Definition 42 (f-Marking and f-Numbering). *Let $m,n \geq 1$. Let $Y \subseteq [m] \times [n]$ and let the characteristic $m\times n$-matrix (c_{ij}) of Y be defined as follows.*

$$c_{ij} = \begin{cases} 1 & \text{if } (i,j) \in Y \\ 0 & \text{if } (i,j) \notin Y \end{cases} \qquad \text{for every } (i,j) \in [m] \times [n].$$

We call Y a column numbering *iff the string $c_{1j}\dots c_{mj}$ is the binary representation of $j-1$ for every $j \leq m$. A column numbering Y is* complete *iff $2^m \mid n$, otherwise incomplete.*

If $j, l \leq n$ with $j+l-1 \leq n$, and $x = (1, j) \in [m] \times [n]$, then let us write $Y(x, l)$ for $c_{1,j}\dots c_{1,j+l-1}$, i.e. the sequence of length l starting in x in the characteristic matrix of Y.

*We call Y a f-*numbering *iff n is a multiple of $f(m)2^{f(m)}$ and $Y \subseteq \{1\} \times [n]$ and $Y((1, if(m)+1), f(m))$ $(= c_{1,if(m)+1}\dots c_{1,(i+1)f(m)})$ is the binary representation of i for every $i \leq \frac{n}{f(m)}$.*

An f-numbering Y is complete *iff n is a multiple of $f(m)2^{f(m)}$.*

*We call Y an f-*marking *iff n is a multiple of $f(m)$, $Y \subseteq \{1\} \times [n]$, and $Y((1, if(m)+1), f(m)) \in 10^*$ for every $i \leq \mathfrak{n}f(m)$.*

In contrast to [Sch97,MST98] we need notions of f-marking and f-numbering that are relativized to some rectangle section, so we proceed with the following definition.

Definition 43 (Relativized f-Marking and f-Numbering). *Let $m, n \geq 1$. Let $Y \subseteq [m] \times [n]$ and let $Z = [m] \times [i, j]$ be a rectangle section. We call Y a* complete column numbering *(a* complete f-numbering, *or an* f-marking, *respectively)* on Z *iff Y is a complete column numbering (a complete f-numbering, or an f-marking, respectively) for the restriction of $\underline{[m] \times [n]}$ to Z.*

Let $k \geq 0$. Let $C, X_1, \dots, X_k, Y_1, \dots, Y_{k-1} \subseteq [m] \times [n]$. Then $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a complete k-marking/numbering on Z *iff*

- *C is a complete column numbering on Z,*
- *X_l is an f_l-marking on Z for every $l \leq k$,*
- *Y_l is a complete f_l-numbering on Z for every $l \leq k-1$, and*
- *$X_k \cap Z = \{(1, i)\}$.*

Let $j \geq 0$. Then $(C, \overline{X}_j, \overline{Y}_j)$ is an incomplete j-marking/numbering on Z *iff*

- *C is a column-numbering, and C is complete iff $j \geq 1$.*
- *X_i is an f_i-marking for all $i \in \{1, \dots, j\}$,*
- *Y_i is a complete f_i-numbering for all $i \in \{1, \dots, j-1\}$,*
- *if $j \geq 1$, then Y_j is an incomplete f_j-numbering.*

Definition 44. *Let (X_k, X'_k, Z_2) be a shuffle triple in a grid $\underline{[m] \times [n]}$ and (x, x') be a twin.*

Let $T = (C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X'}_k, \overline{Y'}_{k-1}, D, D')$ be a tuple of subsets of $[m] \times [n]$. T is a complete k-twin numbering *iff*

1. *(X_k, X'_k, Z_2) is a complete shuffle triple,*
2. *$(C, \overline{X}_k, \overline{Y}_{k-1})$ is a complete k-marking/numbering on every first block of (X_k, X'_k, Z_2),*
3. *$(C', \overline{X'}_k, \overline{Y'}_{k-1})$ is a complete k-marking/numbering on every second block of (X_k, X'_k, Z_2),*
4. *D (respectively D') is the union of all first (respectively second) blocks of (X_k, X'_k, Z_2).*

T is a twin numbering fragment *iff X_k contains only one element (which is hence $(1,1)$), D is the rectangle spanned by $(1,1)$, and either*

1. *there is an $x' \in X'_k$, the rectangle section spanned by x' is in D', and $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a complete k-marking/numbering on D, and $(C', \overline{X'}_k, \overline{Y'}_{k-1})$ is an incomplete k-marking/numbering on D'; or*
2. *$X'_k = \emptyset$ (and hence also $Z_2 = \emptyset$), and $(C, \overline{X}_k, \overline{Y}_{k-1})$ is an incomplete k-marking/numbering on D.*

T is an incomplete k-twin numbering *iff there is an $n' < n$ such that T is a complete k-twin numbering* on *$[m] \times [n']$ and T is a twin numbering fragment on $[m] \times [n'+1, n]$.*

A (complete or incomplete) k-twin numbering T as above is called equidistant with offset i *iff all twins of (X_k, X'_k, Z_2) have the same offset i.*

The following proposition is important.

Proposition 45. *An equidistant complete (respectively incomplete) k-twin numbering with offset i exists iff $f_k(m) + i \mid n$ (respectively $(f_{k-1}(m) \mid n) \wedge (f_k(m) \nmid n)$).*

Here and in the following we extend the family (f_k) of functions by $f_0(m) := 1$ for all $m \geq 1$.

Proof (of Proposition 45). For the proof we note that for an equidistant k-twin numbering with offset i, the distance between successive X_k-positions is $i + f_k(m)$. Besides, there is a k-twin numbering with offset i for every grid $\underline{[m] \times [n]}$ and every $i < f_k(m)$ (in fact it is unique) and it is complete iff $f_k(m) + i \mid n$.

This k-twin numbering for a grid $\underline{[m] \times [n]}$ can be found like this: Let $t = \lfloor n/(i + f_k(m)) \rfloor$. Let

$$X_k = \{j(f_k(m) + i) + 1 \mid j \leq t\}$$

$$X'_k = \begin{cases} \{j(f_k(m) + i) + i + 1 \mid j \leq t\} \\ \qquad \text{if } (t-1)(f_k(m) + i) + f_k(m) \leq n, \\ \{j(f_k(m) + i) + i + 1 \mid j < t\} \\ \qquad \text{if } (t-1)(f_k(m) + i) + f_k(m) > n, \end{cases}$$

$$Z_2 = \begin{cases} \{j(f_k(m) + i) + f_k(m) + 1 \mid j \leq t\} \\ \qquad \text{if } (t-1)(f_k(m) + i) + f_k(m) + 1 \leq n, \\ \{j(f_k(m) + i) + f_k(m) + 1 \mid j < t\} \\ \qquad \text{if } (t-1)(f_k(m) + i) + f_k(m) + 1 > n. \end{cases}$$

Then (X_k, X'_k, Z_2) is a shuffle triple. The remaining components of the k-twin numbering are obvious: D (respectively D') becomes the union of all first (respectively second) blocks of that shuffle triple, and $(C, \overline{X}_k, \overline{Y}_{k-1})$ (respectively $(C', \overline{X'}_k, \overline{Y'}_{k-1})$) becomes the complete k-marking/numbering of all first (respectively second) blocks of (X_k, X'_k, Z_2), except maybe for the very last one, where the j-marking/numbering is complete for the highest possible $j \leq k$.

The above proposition provides a way how to express with an existential formula that the length of a grid is or is not a multiple of $f_k(m) + i$. We are going to put that formal in Lemma 56. For the start, we need some simple Σ_1-formulas.

Lemma 46. *There are Σ_1-formulas compl-shuffle(Z_0, Z_1, Z_2), rect(D), rectangle(Z, x, y), twin(Z_0, Z_1, x_0, x_1), Block(X, x, y), first-blocks(Z_0, Z_1, Z_2, D), second-blocks(Z_0, Z_1, Z_2, D) such that:*

- *shuffle(Z_0, Z_1, Z_2) asserts that (Z_0, Z_1, Z_2) is a shuffle triple.*
- *compl-shuffle(Z_0, Z_1, Z_2) asserts that (Z_0, Z_1, Z_2) is a complete shuffle triple.*

- *rect(D) asserts that D is a rectangle section.*
- *rectangle(Z, x, y) asserts that there are i, j such that $x = (1, i)$ and $y = (1, j)$ with $i \leq j$ and $Z = [1, n] \times [i, j]$, i.e., that Z is the smallest rectangle section that contains both of the top row positions x and y.*
- *twin(Z_0, Z_1, x_0, x_1) asserts for a shuffle triple (Z_0, Z_1, Z_2) that (x_0, x_1) is a twin.*
- *Block(X, x, y) asserts for two positions x, y in the top row that $y \in$ Block(X, x).*
- *first-blocks(Z_0, Z_1, Z_2, D) asserts for a shuffle triple (Z_0, Z_1, Z_2) that D is the union of all first blocks of (Z_0, Z_1, Z_2).*
- *second-blocks(Z_0, Z_1, Z_2, D) asserts for a shuffle triple (Z_0, Z_1, Z_2) that D is the union of all second blocks of (Z_0, Z_1, Z_2).*

For the proof we note that it is not difficult to write down such formulas. However, I encourage the reader familiar with the notion of "recognizability of picture languages" to deduce this from the recognizability to certain picture (or even string-) languages.

For example, the existence of a Σ_1-formula *compl-shuffle* can be deduced from the fact that $L = (a^*ba^*ca^*d)^*a^* \cup (a^*e)^*a^*$ (with $a = (0,0,0)^\top$, $b = (1,0,0)^\top$, $c = (0,1,0)^\top$, $d = (0,0,1)^\top$, and $e = (1,1,1)^\top$) is a recognizable string language. Therefore, the "picture language" containing all matrices whose top row is in L is recognizable. Hence (by [GRST96]) the corresponding set of $(3,0)$-grids, which is exactly $Mod_{3,0}(\textit{compl-shuffle})$, is definable in Σ_1.

Lemma 47. *There is a Δ_1-formula equal-col-num(C, x, y) that asserts for two top row positions x, y that the C-colouring in the two columns of x and y are equal.*

Proof. Let *col-closed*$(X) = \forall x, x'(S_1xx' \to (Xx \leftrightarrow Xx'))$ be a first-order formula that asserts that X is a union of columns. Let *row-closed*(X) be a similar first-order formula that asserts that X is a union of rows.

We write "$X \cap top = \{x\}$" as a short-hand for a first-order formula that asserts that x is the only element of X in the top row, and "$Z \cap X = C \cap X$" as a shorthand for the formula $\forall x(Zx \wedge Xx \leftrightarrow$

$Cx \wedge Xx$). Then the required Σ_1- and Π_1-formula can be defined as follows.

$$\begin{aligned}
&\textit{equal-col-num}'(C,x,y) = \\
&\quad \exists X,Y,Z\big(\textit{col-closed}(X) \wedge \textit{col-closed}(Y) \wedge \textit{row-closed}(Z) \\
&\qquad \wedge (X \cap \textit{top} = \{x\}) \wedge (Y \cap \textit{top} = \{y\}) \\
&\qquad \wedge (Z \cap X = C \cap X) \wedge (Z \cap Y = C \cap Y)\big)
\end{aligned}$$

$$\begin{aligned}
&\textit{equal-col-num}''(C,x,y) = \\
&\quad \forall X,Y,Z\Big(\big(\textit{col-closed}(X) \wedge \textit{col-closed}(Y) \wedge \textit{row-closed}(Z) \\
&\qquad \wedge (X \cap \textit{top} = \{x\}) \wedge (Y \cap \textit{top} = \{y\})\big) \\
&\qquad \to ((Z \cap X = C \cap X) \wedge (Z \cap Y = C \cap Y))\Big)
\end{aligned}$$

4.4 Formulas Borrowed from Schweikardt

In the following, we sometimes write $\boldsymbol{CXY}_k$ (for $k \geq 1$) to abbreviate the variable tuple $(C, X_1, \ldots, X_k, Y_1, \ldots, Y_{k-1})$. Remember that $\overline{X}_k$ is an abbreviation for the tuple $(X_1, \ldots, X_k)$.

The following is shown implicitly in [Sch97,MST98]

Lemma 48. *For every $k \geq 1$, there is a Π_{k-1}-formula*

$$\textit{compl-mark-num}_k(\boldsymbol{CXY}_k)$$

that asserts that $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a complete k-marking/numbering.

For every $k \geq 0$, there is a Π_k-formula

$$\textit{incompl-mark-num}_k(C, \overline{X}_k, \overline{Y}_k)$$

that asserts that $(C, \overline{X}_k, \overline{Y}_k)$ is an incomplete k-marking/numbering.

Proof. In Schweikardt's notation ([Sch97,MST98]), these formulas are:

$$\begin{aligned}
&\textit{compl-mark-num}_k := \\
&\quad \textit{compl-col-num}(C) \wedge f_1\textit{-marking}(X_1, C) \wedge \\
&\quad \textstyle\bigwedge_{i=1}^{k-1}(\textit{compl-}f_i\textit{-num}(Y_i, \overline{X}_i, \overline{Y}_{i-1}) \wedge \textit{marking}(X_{i+1}, X_i, Y_i)) \\
&\quad \wedge \textit{Singleton}(X_k)),
\end{aligned}$$

$$\begin{aligned}
&\textit{incompl-mark-num}_0 := \\
&\quad \textit{col-num}(C) \wedge \forall x\big(\textit{left}(x) \to \neg Cx \wedge \exists x(\textit{right}(x) \wedge \neg Cx)\big),
\end{aligned}$$

$$\begin{aligned}
&\textit{incompl-mark-num}_k := \\
&\quad \textit{compl-col-num}(C) \wedge f_1\textit{-marking}(X_1, C) \wedge \\
&\quad \textstyle\bigwedge_{i=1}^{k-1}(\textit{compl-}f_i\textit{-num}(Y_i, \overline{X}_i, \overline{Y}_{i-1}) \wedge \textit{marking}(X_{i+1}, X_i, Y_i)) \wedge \\
&\quad f_k\textit{-num}(Y_k, \overline{X}_k, \overline{Y}_{k-1}) \wedge \textit{leftmin}(Y_k, X_k) \wedge \neg\textit{rightmax}(Y_k, X_k).
\end{aligned}$$

Remark 49. Let $k \geq 1$. For the grid $[m] \times [n]$, the sentence

- $\exists \boldsymbol{CXY}_k(\textit{compl-mark-num}_k(\boldsymbol{CXY}_k))$ asserts that $f_k(m) \mid n$,
- $\exists C, X_1, \ldots X_j, Y_1, \ldots, Y_j(\textit{incompl-mark-num}_j(C, \overline{X}_j, \overline{Y}_j))$ asserts that $f_{j-1}(m) \mid n$ and $f_j(m) \nmid n$.
- $\exists \boldsymbol{CXY}_k\left(\bigvee_{j=0}^{k} \textit{incompl-mark-num}_j(C, \overline{X}_j, \overline{Y}_j)\right)$ asserts that $f_k \nmid n$.

Note that

$$\exists \boldsymbol{CXY}_k(\textit{Singleton}(X_k) \wedge \textit{compl-mark-num}_k(\boldsymbol{CXY}_k))$$

is a Σ_k-formula that defines f_k, which is Part 1 of the proof of the "lower bound part" of Theorem 14, and

$$\begin{aligned}
&\exists \boldsymbol{CXY}_k(\\
&\quad (\textit{compl-mark-num}_k(\boldsymbol{CXY}_k) \wedge \neg\textit{Singleton}(X_k)) \\
&\quad \vee \textstyle\bigvee_{j=0}^{k-1} \textit{incompl-mark-num}_j(C, \overline{X}_j, \overline{Y}_j))
\end{aligned}$$

is a Σ_k-formula whose negation defines f_k, which is Part 2 of the proof of "lower bound part" of Theorem 14. (See [Sch97,MST98]). However, we need to redo parts of the proof and to reuse some of Schweikardt's auxiliary formulas, sometimes relativized to rectangle sections, in order to prove Theorem 18. The crucial point is that in the presence of a k-marking/numbering, there are Π_k-formulas that allow to compare sections of length $f_k(m)$.

4.5 More Formulas

Lemma 50. *Let $k \geq 1$. There is a Σ_k-formula*

$$\textit{rect-compl-mark-num}_k(\boldsymbol{CXY}_k, D)$$

such that:

If D is a union of columns, then the formula rect-compl-mark-num$_k(\boldsymbol{CXY}_k, D)$ asserts that $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a k-marking/-numbering on every maximal rectangle section $D' \subseteq D$.

Proof (of Lemma 50, first version). Consider the case $k = 1$ first. Schweikardt's formula *compl-mark-num*$_1(C)$, which is named *compl-col-num*(C) in [Sch97,MST98], says that every column that has no predecessor column contains no elements in C, every column that has no successor column contains no elements not in C, and every two successive columns code successive binary numbers. (Since binary counting is a local process, *compl-col-num*(C) is first-order.)

Let *rect-compl-mark-num*$_1(C, D)$ result from the first-order formula *compl-mark-num*$_1(C)$ by relativizing first-order quantifications to D. An investigation of Schweikardt's proof for the correctness of *compl-mark-num*$_0(C)$ shows that this will indeed suffice.

Now let $k > 1$. Let *rel-compl-mark-num*$_k(\boldsymbol{CXY}_k, D)$ result from *compl-mark-num*$_k(\boldsymbol{CXY}_k, D)$ by relativizing first-order quantifications to D. If Schweikardt had used our version of the formula *equal-col-num* (see Lemmas 47 and 53), the formula *rect-compl-mark-num*$_k$:= *rel-compl-mark-num*$_k(\boldsymbol{CXY}_k, D)$ would indeed be a Π_{k-1}-formula (and thus a Σ_k-formula) with the required property.

The above proof even gives Π_{k-1}-formulas, but it requires an investigation of Schweikardt's proofs of [MST98, Lemmas 24,25]. We avoid this here and give a different proof for Lemma 50 that yields Σ_k-formulas only.

Proof (of Lemma 50, second version). The case $k = 1$ is treated as in the first version of the proof, which even gives a first-order formula. So let $k > 1$.

Let *maxrect*(D', D) be a Σ_1-formula that asserts for a union of columns D that $D' \subseteq D$ is a maximal rectangle section, i.e.,

$$\begin{aligned}
&\textit{maxrect}(D', D) = \\
&\quad \textit{rect}(D') \wedge (D' \subseteq D) \\
&\quad \wedge \forall x, x' (S_2 x x' \rightarrow ((D'x \wedge \neg D'x \rightarrow Dx \wedge \neg Dx) \\
&\qquad\qquad \wedge \ (\neg D'x \wedge D'x \rightarrow \neg Dx \wedge Dx))).
\end{aligned}$$

Let, as before, the Π_{k-1}-formula *rel-compl-mark-num*$_k(\boldsymbol{CXY}_k, D)$ result from *compl-mark-num*$_k(\boldsymbol{CXY}_k)$ by relativizing first-order quantifications to D. Then let

$$\begin{aligned}
&\textit{rect-compl-mark-num}_k = \\
&\quad \forall D' (\textit{maxrect}(D', D) \rightarrow \textit{rel-compl-mark-num}_k(\boldsymbol{CXY}_k, D')).
\end{aligned}$$

Obviously, *rect-compl-mark-num*$_k$ asserts the desired property and is in $\Pi_{k-1} \subseteq \Sigma_k$.

Lemma 51. *There is a Δ_k-formulas smaller$_k(x)$ that asserts that $x \in \{1\} \times [f_k(m)]$*

Proof. Let

$$\begin{aligned}
\textit{smaller}_k(x) :=\ & \exists \boldsymbol{CXY}_k \exists Z (\\
& \exists y (\textit{left}(y) \wedge \textit{rectangle}(Z, y, x) \wedge \textit{Singleton}(X_k)) \\
& \wedge \textstyle\bigvee_{j=0}^{k} (\textit{rect-incompl-mark-num}_j(\boldsymbol{CXY}_j, Z))), \\
\textit{smaller}'_k(x) :=\ & \neg \textit{top}(x) \vee \exists \boldsymbol{CXY}_k \exists Z (\\
& \exists y (\textit{left}(y) \wedge \textit{rect}(Z) \wedge Zy \wedge \neg Zx) \\
& \wedge \textit{rect-compl-mark-num}_k(\boldsymbol{CXY}_k, Z)\big).
\end{aligned}$$

Then $\textit{smaller}_k(x)$ (respectively $\textit{smaller}'(x)$) is a Σ_k-formula that asserts that $x \in \{1\} \times [f_k(m)]$ (respectively $x \notin \{1\} \times [f_k(m)]$).

The following is a slight modification of [MST98, Lemma 24].

Lemma 52 (How to Compare f_k-Numbers). *For all $k \geq 1$ there are Π_k-formulas equal$_k(x, y, Y, \boldsymbol{CXY}_k)$ and inc$_k(x, y, Y, \boldsymbol{CXY}_k)$ such that over grids the following holds:*

If there are rectangle sections V, W of $[m] \times [n]$ such that $(C, Y, \overline{X}_k, \overline{Y}_{k-1})$ is a k-marking/numbering on V as well as on W, and if $x \in X_k \cap V$ and $y \in X_k \cap W$, then

1. *$equal_k(x, y, Y, \boldsymbol{CXY}_k)$ asserts that the blocks* $\mathrm{Block}(X_k, x)$ *and* $\mathrm{Block}(X_k, y)$ *have the same Y-number, i.e., $Y(x, f_k(m)) = Y(y, f_k(m))$, and*
2. *$inc_k(x, y, Y, \boldsymbol{CXY}_k)$ asserts that the Y-number of* $\mathrm{Block}(X_k, y)$ *is one larger than the Y-number of* $\mathrm{Block}(X_k, x)$*, i.e., that $Y(y, f_k(m)) \equiv Y(x, f_k(m)) + 1 \pmod{2^{f_k(m)}}$.*

Proof. The proof is by induction on k. The induction step is the same as in [Sch97,MST98]. The induction base, however, is a bit different.

Let $k = 1$. Let $m, n \geq 1$ and V, W rectangle sections of $[m] \times [n]$. Let $X_1 \subseteq [m] \times [n]$ such that X is an f_1-marking on V and on W. Furthermore let $x \in X_1 \cap V$ and $y \in X_1 \cap W$; and let C be a complete column-numbering on V and on W.

Then the Π_1-formula[3]

$$\begin{aligned} &equal_1(x, y, Y, C, X_1) := \\ &\quad \forall x_0, y_0 \Big(\big((Block(X_1, x, x_0) \wedge Block(X_1, y, y_0) \\ &\quad \wedge equal\text{-}col\text{-}num(C, x_0, y_0) \big) \rightarrow (Y x_0 \leftrightarrow Y y_0) \Big) \end{aligned}$$

holds iff $Y(x, 2^m) = Y(y, 2^m)$.

The formula $inc_1(x, y, C, X_1, Y)$ can be obtained from Schweikardt's version similarly. This completes the induction base and thus the proof.

We will not use the formulas inc_k here any more. They are just needed for the induction step carried out in [Sch97,MST98].

Now we state a lemma that says that in the presence of a k-marking/numbering of two rectangle sections V and W, we are able to determine corresponding positions inside V and W by a Π_{k-1}-formula.

This lemma is the only result that made it necessary to formulate Definition 43 and Lemma 52.

Lemma 53. *Let $k \geq 2$. There is a Π_{k-1}-formula $equal\text{-}pos_k(\boldsymbol{CXY}_k, x', y')$ such that the following holds:*

[3] Here we exploit that *equal-col-num* and *Block* are in Σ_1.

If $x' = (1, i')$, $y' = (1, j')$ are in the top row and there are rectangle sections V, W and $x = (1, i) \in V$, $y = (1, j) \in W$ such that $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a complete k-marking/numbering on V and on W and $X_k \cap V = \{x\}$ and $X_k \cap W = \{y\}$, then $i - i' = j - j'$ iff $equal\text{-}pos_k(C, \overline{X}_k, \overline{Y}_{k-1}, x', y')$ holds.

Proof. We will use the following fact: Let $1 = m_0, m_1, \ldots, m_k \in \mathbb{N}$ with $m_i \mid m_{i+1}$ for every $i < k$. Then $\lfloor \frac{i}{m_l} \rfloor \equiv \lfloor \frac{j}{m_l} \rfloor \pmod{m_{l+1}}$ for every $l < k$ iff $i \equiv j \pmod{m_k}$.

Let

$$\begin{aligned} &equal\text{-}pos_k(\boldsymbol{CXY}_k, x', y') := \\ &\quad equal\text{-}col\text{-}num(C, x', y') \wedge \\ &\quad \textstyle\bigwedge_{l<k} \forall x, y (Block(X_l, x, x') \wedge Block(X_l, y, y') \\ &\qquad \rightarrow equal_l(x, y, Y_l, \boldsymbol{CXY}_l)) \end{aligned}$$

Exploiting that *equal-col-num* is a Π_1-formula, we observe that $equal\text{-}pos_k$ is a Π_{k-1}-formula.

Let $m, n \geq 1$, $V, W \subseteq [m] \times [n]$, and $x', y', C, \overline{X}_k, \overline{Y}_{k-1}$ be chosen such that all the assumptions hold.

Choose $i', j', i, j \leq n$ such that $(1, i') = x'$, $(1, j') = y'$, $(1, i) = x$, and $(1, j) = y$.

Since C is a complete column-numbering on V and on W and x and y' are the leftmost positions in the top row of V and W, the C-colourings of the columns i' and j' are equal iff $i - i' \equiv j - j' \pmod{2^m}$.

Furthermore, if $1 \leq l < k$, and $x_l = (1, i_l)$ and $y_l = (1, j_l)$ are the nearest positions in X_l left to x' and y', respectively, then x_l and y_l are (the only) positions for which $Block(X_l, x_l, x')$ and $Block(X_l, y_l, y')$ become true, and $equal_l(x_l, y_l, Y_l, \boldsymbol{CXY}_l)$ asserts that $Y_l(x_l, f_l(m)) = Y_l(x_l, f_l(m))$, which is, because Y_l is an f_l-numbering, in turn equivalent to $\lfloor \frac{i-i'}{f_l(m)} \rfloor \equiv \lfloor \frac{j-j'}{f_l(m)} \rfloor \pmod{f_{l+1}(m)}$.

Summing up, we have that $equal\text{-}pos_k(C, \overline{X}_k, \overline{Y}_{k-1}, x', y')$ asserts that

$$\left\lfloor \frac{i - i'}{f_l(m)} \right\rfloor \equiv \left\lfloor \frac{j - j'}{f_l(m)} \right\rfloor \pmod{f_{l+1}(m)}$$

for all $l \in \{0,\ldots,k-1\}$, which is equivalent to $i-i' \equiv j-j' \pmod{f_k}$. Since $V \cap X_k$ and $W \cap X_k$ are singletons and X_k is an f_k-marking, the length of V and W are $f_k(m)$ and thus $i-i', j-j' \le f_k(m)$. Thus *equal-pos*$_k(\boldsymbol{CXY}_k, x', y')$ asserts that $i-i' = j-j'$.

Lemma 54. *Let $k \ge 1$. There is a Σ_k-formula equal-offset$_k(\boldsymbol{CXY}_k, X'_k)$ such that the following holds: If (X_k, X'_k, Z_2) is a shuffle triple, $(C, \overline{X}_k, \overline{Y}_{k-1})$ is a k-marking/numbering on every first block of (X_k, X'_k, Z_2), then it asserts that $x-x' = y-y'$ for all twins (x,x'), (y,y').*

Proof. Let us consider the case $k=1$ first. Let

$$\begin{aligned} \textit{equal-offset}_1(C, X_1, X'_1) := \exists X, Z\big(&\textit{col-closed}(X) \wedge X \cap \textit{top} = X'_1 \\ &\wedge \textit{row-closed}(Z) \wedge Z \cap X = C \cap X\big). \end{aligned}$$

equal-offset$_1$ is a Σ_1-formula.

Now let $k > 1$.

$$\begin{aligned} &\textit{equal-offset}_k(C, \overline{X}_k, \overline{Y}_{k-1}, X'_k) := \\ &\quad \forall x, x', y, y'\Big(\big(\textit{twin}(X_k, X'_k, x, x') \wedge \textit{twin}(X_k, X'_k, y, y')\big) \\ &\qquad \to \textit{equal-pos}_k(C, \overline{X}_k, \overline{Y}_{k-1}, x', y')\Big). \end{aligned}$$

Since *equal-pos*$_k$ is a Π_{k-1}-formula and *twin* is a Σ_1-formula, *equal-offset*$_k$ is a Π_{k-1}-formula for every $k > 1$.

Summing up, *equal-offset*$_k$ it is a Σ_k-formula for every k.

Lemma 55. *Let $k \ge 1$. There are Σ_k-formulas*

$$\textit{compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D'),$$

$$\textit{incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D')$$

that assert that $(C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X}'_k, \overline{Y}'_{k-1}, Z_2, D, D')$ is a complete (respectively incomplete) k-twin numbering.

Proof.

$$\begin{aligned} &\textit{compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') := \\ &\quad \textit{compl-shuffle}(X_k, X'_k, Z_2) \\ &\quad \wedge \textit{first-blocks}(X_k, X'_k, Z_2, D) \\ &\quad \wedge \textit{rect-compl-mark-num}_k(C, \overline{X}_k, \overline{Y}_{k-1}, D) \\ &\quad \wedge \textit{second-blocks}(X_k, X'_k, Z_2, D') \\ &\quad \wedge \textit{rect-compl-mark-num}_k(C', \overline{X}'_k, \overline{Y}'_{k-1}, D'). \end{aligned}$$

compl-twin-numb$_k$ is a Σ_k-formula.

$$\begin{aligned} &\textit{fragment-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') := \\ &\quad \textit{shuffle}(X_k, X'_k, Z_2) \wedge \textit{Singleton}(X_k) \wedge \\ &\qquad (((X'_k \ne \emptyset) \wedge \textit{first-blocks}(X_k, X'_k, Z_2, D) \\ &\qquad \wedge \textit{rect-compl-mark-num}_k(\boldsymbol{CXY}_k, D) \\ &\qquad \wedge \textit{second-blocks}(X_k, X'_k, Z_2, D') \\ &\qquad \wedge \textit{rect-incompl-mark-num}_k(\boldsymbol{CXY}'_k, D')) \\ &\quad \vee ((X'_k = \emptyset) \wedge \textit{incompl-mark-num}_k(\boldsymbol{CXY}_k, D))). \end{aligned}$$

fragment-twin-numb$_k$ is a Σ_k-formula.

Let *rel-fragment-twin-numb*$_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z)$ result from *fragment-twin-numb*$_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D')$ by relativizing first-order quantifications to Z. Let Z be a rectangle section of $[m] \times [n]$, then *rel-fragment-twin-numb*$_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z)$ asserts that $(C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X}'_k, \overline{Y}'_{k-1}, Z_2, D, D')$ is a k-twin numbering fragment on Z.

$$\begin{aligned} &\textit{incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', E, F) := \\ &\quad \textit{rect}(E) \wedge \exists x(\textit{topleft}(x) \wedge Ex) \\ &\quad \wedge \forall x(Ex \leftrightarrow \neg F(x)) \\ &\quad \wedge \textit{rel-compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', E) \\ &\quad \wedge \textit{rel-fragment-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', F). \end{aligned}$$

Note that *rel-compl-twin-sequence*$(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', E) \wedge$ *rel-fragment-twin-numb*$_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', F)$ implies *rel-compl-shuffle*$(X_k, X'_k, Z_2, E) \wedge$ *shuffle*(X_k, X'_k, Z_2), which, if E is the complement of F, implies that *shuffle*(X_k, X'_k, Z_2) holds.

incompl-twin-numb$_k$ is a Σ_k-formula as desired.

Lemma 56. *Let $k \geq 1$. There are Σ_k-formulas*

$$\begin{gathered} \textit{equi-compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D'), \\ \textit{equi-incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') \end{gathered}$$

that assert that $(C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X}'_k, \overline{Y}'_{k-1}, Z_2, D, D')$ is an equidistant complete (respectively incomplete) k-twin numbering.

Proof. Let

$$\begin{aligned} &\textit{equi-compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') := \\ &\qquad (\textit{compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') \\ &\qquad \wedge \textit{equal-offset}_k(\boldsymbol{CXY}_k, X'_k)). \end{aligned}$$

$\textit{equi-compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D')$ is a Σ_k-formula. It asserts that $(C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X}'_k, \overline{Y}'_{k-1})$ is a complete k-twin numbering on each twin block of the complete shuffle triple (X_k, X'_k, Z_2), D (respectively D') is the union of all first (respectively second) blocks of each shuffle triple, and all twin blocks have the same offset. Note that if that offset is j, then the existence of those sets $C, \overline{X}_k, \overline{Y}_{k-1}, C', \overline{X}'_k, \overline{Y}'_{k-1}, D, D'$ is equivalent to $f(m) + j \mid n$.

The construction of the formula $\textit{equi-incompl-twin-numb}_k$ is similar:

$$\begin{aligned} &\textit{equi-incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') := \\ &\qquad \textit{incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D') \\ &\qquad \wedge \textit{equal-offset}_k(\boldsymbol{CXY}_k, X'_k). \end{aligned}$$

$\textit{equi-incompl-twin-numb}_k(z)$ is a Σ_k-formula. This completes the proof of Lemma 56.

Proof (of Theorem 18). Let $\textit{rel-equi-compl-twin-numb}_k$ (respectively $\textit{rel-equi-incompl-twin-numb}_k$ result from $\textit{equi-compl-twin-numb}_k$ (respectively $\textit{equi-incompl-twin-numb}_k$) by relativizing first-order quan-

tifications to the fresh variable Z. Let

$$\begin{aligned} &\rho_k(z, y) := \\ &\quad \exists \boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z(\\ &\quad \exists x(\textit{left}(x) \wedge \textit{rectangle}(Z, x, y) \wedge X'_k z) \\ &\quad \wedge \textit{rel-equi-compl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z)), \\ &\rho'_k(z, y) := \\ &\quad \exists \boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z(\\ &\quad \exists x(\textit{topleft}(x) \wedge \textit{rectangle}(Z, x, y) \wedge X'_k z) \\ &\quad \wedge \textit{rel-equi-incompl-twin-numb}_k(\boldsymbol{CXY}_k, \boldsymbol{CXY}'_k, Z_2, D, D', Z)). \end{aligned}$$

$\rho_k(z, y)$ (respectively $\rho'_k(z, y)$) is a Σ_k-formula that asserts, for $z = (1, j)$ with $j \leq f_k(m)$ and $y = (1, i)$, that $(j-1) + f_k(m) \mid i$ (respectively that $(j-1) + f_k(m) \nmid i$).

Let

$$\begin{aligned} \sigma_k(z, y) &:= (\textit{top}(y) \wedge \textit{smaller}_k(z)) \rightarrow \rho_k(z, y), \\ \sigma'_k(z, y) &:= (\neg \textit{top}(y) \vee \neg \textit{smaller}_k(z)) \wedge \rho'_k(z, y). \end{aligned}$$

σ_k is a Σ_k-formula that asserts that if y, z are positions in the top row, say $y = (1, i)$ and $z = (1, j)$, and if $j \leq f_k(m)$, then $(j-1) + f_k(m) \mid i$.

σ'_k is a Σ_k-formula equivalent to $\neg \sigma_k$, thus σ_k is in fact a Δ_k-formula.

Let

$$\psi(y) = \textit{top}(y) \wedge \forall z(\sigma_k(z, y)).$$

Then $\psi(y)$ is a $\mathrm{FO}(\Delta_k)$-formula that asserts that y is in the top row, say $y = (1, i)$, and that $(g_k(m) =)\ \mathrm{lcm}\{f_k(m), f_k(m)+1, \ldots, 2f_k(m)-1\}$ divides i.

Let

$$\varphi = \forall y\Big(\textit{top}(y) \rightarrow \big(\textit{right}(y) \leftrightarrow \psi(y)\big)\Big).$$

φ is in $\mathrm{FO}(\Delta_k)$ and asserts for a grid $[m] \times [n]$ that n is the only $i \leq n$ with $g_k(m) \mid i$, which is equivalent to $g_k(m) = n$. Thus the $\mathrm{FO}(\Delta_k)$-formula φ defines g_k, and $g_k(m)$ is $s_{k+1}(\Theta(m))$. This completes the proof of Theorem 18.

5 Conclusion

We have improved the periodicity argument of [MT97] in order to show that in level $k+2$ of the monadic hierarchy over grids there is a property that is not in the first-order/boolean closure of level k, and that the monadic second-order hierarchy remains strict when each level is closed under first-order quantifications.

As in [MT97,Sch97,MST98] we have considered sets of grids where the length is a function of the height. Our separation results are derived from sharp asymptotic growth rate bounds for these functions. These separations can be transferred from the class of grids to the class of graphs by encoding techniques presented in [MT97,MST98].

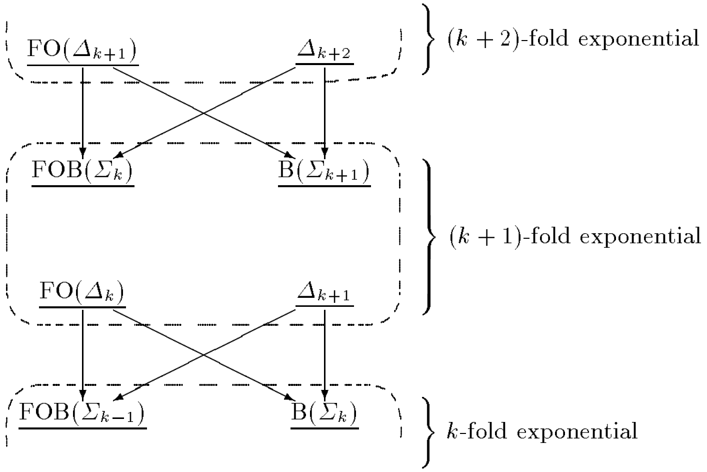
The separation results for the class of grids we have proved in this paper are displayed in Figure 2.

However, some separations are not (yet?) possible using (non-coloured) grids. For example, it is still open whether $\underline{\Sigma_1} = \underline{\Pi_1}$ for the class of grids (I conjecture this)—at least, these classes do not differ with respect to the rate of growth of functions. In this situation, coloured grids are useful and allow to separate the graph classes $\underline{\Sigma_k}$ from $\underline{\Pi_k}$ (and thus $\underline{\mathrm{B}(\Sigma_k)}$ from $\underline{\Delta_k}$) for every $k \geq 1$.

In the context of first-order closure, we encounter again the phenomenon that—with respect to the rate of growth of functions!—Δ_k is already as powerful as the boolean closure of $\mathrm{B}(\Sigma_k)$, so the "growth argument" cannot be used to separate $\underline{\mathrm{FO}(\Delta_k)}$ from $\underline{\mathrm{FOB}(\Sigma_k)}$. I do not know how to separate these classes for coloured grids.

The sets of functions f for which L_f is in $\underline{\mathrm{FO}(\Delta_k)}$, $\underline{\mathrm{FOB}(\Sigma_k)}$, $\underline{\Delta_{k+1}}$, or $\underline{\mathrm{B}(\Sigma_{k+1})}$, respectively, do not differ *wrt. their asymptotic rate of growth*. Except for the trivial inclusions $\underline{\mathrm{FO}(\Delta_k)} \subseteq \underline{\mathrm{FOB}(\Sigma_k)}$ and $\underline{\Delta_{k+1}} \subseteq \underline{\mathrm{B}(\Sigma_{k+1})}$, we do not know anything about the (non-)-inclusions between these classes. We do not even know whether every *function* definable in, say, $\mathrm{FO}(\Delta_k)$ is already definable in $\mathrm{B}(\Sigma_{k+1})$.

It would be nice to have some inclusion results for the class of grids rather than only separation results. For example, I conjecture that $\underline{\mathrm{FO}(\Delta_k)} = \underline{\mathrm{FOB}(\Sigma_k)}$ and $\underline{\Delta_k} = \underline{\mathrm{B}(\Sigma_k)}$ for the class of grids. Furthermore, I conjecture that for the class of grids, $\underline{\mathrm{FOB}(\Sigma_k)} \subseteq \underline{\Sigma_{k+2}}$ for every k.



$X \to Y$ means $X \not\subseteq Y$. If X is displayed exactly above Y, then $X \supseteq Y$ is valid for trivial reasons. Classes encircled by dashed lines allow for the same rate of growth of functions — nothing is claimed about mutual inclusions!

Fig. 2. Separation Results for the Class of Grids

References

[AFS97] M. Ajtai, R. Fagin, and L.J. Stockmeyer. The closure of monadic NP. Research Report RJ 10092, IBM, Dec 1997.

[AFS98] M. Ajtai, R. Fagin, and L.J. Stockmeyer. The closure of monadic NP. In *The Thirtieth Annual ACM Symposium on Theory of Computing*. SIGACT, 1998.

[Bra42] A. Brauer. On a problem of partitions. *Amer. J. Math.*, 64:299–312, 1942.

[Chr86] M. Chrobak. Finite automata and unary languages. *Theoretical Computer Science*, 47, 1986.

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

[Fag95] R. Fagin. Problem proposed in Oberwolfach Meeting 1995; published in: Collection of open problems in finite model theory, see `http://www.informatik.rwth-aachen.de/WWW-math/FMT.html`, 1995.

[Gia94] D. Giammarresi. Two-dimensional languages and recognizable functions. In G. Rozenberg and A. Salomaa, editors, *Developments in Language Theory, Proceedings of the conference, Turku (Finnland) '93*, pages 290–301. world scientific, Singapore, 1994.

[GRST96] D. Giammarresi, A. Restivo, S. Seibert, and W. Thomas. Monadic second-order logic and recognizability by tiling systems. *Information and Computation*, 125:32–45, 1996.

[Imm88] N. Immermann. Languages that capture complexity classes. *SIAM J. Comput.*, 16(4), 1988.

[Lan09] E. Landau. *Handbuch über die Verteilung der Primzahlen*, pages 222–229. Teubner, 1909.

[Liu64] U. I. Liubicz. Bounds for the optimal determinization of nondeterministic autonomic automata. *Sibirskii Mathemat. Journal*, pages 337–355, 1964. (in Russian).

[MST98] O. Matz, N. Schweikardt, and W. Thomas. The monadic quantifier alternation hierarchy over grids and graphs. *submitted to Information and Computation, LICS special issue*, 1998.

[MT97] O. Matz and W. Thomas. The monadic quantifier alternation hierarchy over graphs is infinite. In *Twelfth Annual IEEE Symposium on Logic in Computer Science*, pages 236–244, Warsaw, Poland, 1997. IEEE.

[Sch97] N. Schweikardt. The monadic quantifier alternation hierarchy over grids and pictures. In Mogens Nielson and Wolfgang Thomas, editors, *Proceedings of the Annual Conference of the European Association for Computer Science Logic*, pages 383–398, Aarhus, Denmark, 1997. BRICS.

[Sza80] M. Szalay. On the maximal order in s_n and s_n^*. *Acta Arithm.*, 37:321–331, 1980.