Homogeneity and Homogenizability: Hard Problems for the Logic SNP

Jakub Rydval
Technische Universität Wien, Austria

Abstract

The infinite-domain CSP dichotomy conjecture extends the finite-domain CSP dichotomy theorem to reducts of finitely bounded homogeneous structures. Every countable finitely bounded homogeneous structure is uniquely described by a universal first-order sentence up to isomorphism, and every reduct of such a structure by a sentence of the logic SNP. By Fraïssé's Theorem, testing the existence of a finitely bounded homogeneous structure for a given universal first-order sentence is equivalent to testing the amalgamation property for the class of its finite models. The present paper motivates a complexity-theoretic view on the classification problem for finitely bounded homogeneous structures. We show that this meta-problem is EXPSPACE-hard or PSPACE-hard, depending on whether the input is specified by a universal sentence or a set of forbidden substructures. By relaxing the input to SNP sentences and the question to the existence of a structure with a finitely bounded homogeneous expansion, we obtain a different meta-problem, closely related to the question of homogenizability. We show that this second meta-problem is already undecidable, even if the input SNP sentence comes from the Datalog fragment and uses at most binary relation symbols. As a byproduct of our proof, we also get the undecidability of some other properties for Datalog programs, e.g., whether they can be rewritten in the logic MMSNP, whether they solve some finite-domain CSP, or whether they define a structure with a homogeneous Ramsey expansion in a finite relational signature.

2012 ACM Subject Classification Theory of computation → Logic; Theory of computation → Computational complexity and cryptography

Keywords and phrases constraint satisfaction problems, finitely bounded, homogeneous, amalgamation property, universal, SNP, homogenizable

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.150

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version *Full Version*: https://arxiv.org/abs/2108.00452 [47]

Funding *Jakub Rydval*: This research was funded in whole or in part by the Austrian Science Fund (FWF) [I 5948]. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Acknowledgements The author thanks Manuel Bodirsky, Simon Knäuer, and Jakub Opršal for many inspiring discussions on the topic, and the anonymous reviewers for many helpful suggestions.

1 Introduction

Strict NP (SNP) is an expressive fragment of existential second-order logic and thus, by Fagin's Theorem, of the complexity class NP. If one only considers structures over a finite relational signature, then SNP can be obtained from the universal fragment of first-order logic simply by allowing existential quantification over relation symbols at the beginning of the quantifier prefix. In particular, universal first-order formulas themselves are SNP formulas. Despite the name, SNP already has the full power of NP, in the sense that every problem in NP is equivalent to a problem in SNP under polynomial-time reductions [30]. In addition, this logic class has many connections to *Constraint Satisfaction Problems* (CSPs), which we use as the primary source of motivation for the present article. The CSP of a relational structure $\mathfrak{B}$, denoted by $\mathrm{CSP}(\mathfrak{B})$, is (the membership problem for) the class of all finite structures which homomorphically map to $\mathfrak{B}$. Every computational decision problem is polynomial-time equivalent to a CSP [12]. Many practically relevant problems like Boolean satisfiability or graph colorability can even be formulated as a *finite-domain* CSP, i.e., where the *template* $\mathfrak{B}$ can be chosen finite. The basic link from SNP to CSP is that every sentence of the monotone fragment of this logic defines a finite disjoint union of CSPs of (possibly infinite) relational structures [8]. There are, however, some more nuanced connections, such as the one that led to the formulation of the *Feder-Vardi conjecture*, now known as the finite-domain CSP dichotomy theorem [52]. In their seminal work [30], Feder and Vardi showed that the *Monotone Monadic* fragment of SNP (MMSNP) exhibits a dichotomy between P and NP-completeness if and only if the seemingly less complicated class of all finite-domain CSPs exhibits such a dichotomy,[1] they also conjectured the latter to be true. The logic class MMSNP contains all finite-domain CSPs, and many other interesting combinatorial problems, e.g., the problem of deciding whether the vertices of a given graph can be 2-coloured without obtaining any monochromatic triangle [42]. The Feder-Vardi conjecture was confirmed in 2017 independently by Bulatov and Zhuk [23, 51].

There is a yet unconfirmed generalization of the Feder-Vardi conjecture, to CSPs of *reducts of finitely bounded homogeneous structures*, formulated by Bodirsky and Pinsker in 2011 [18]. Here we refer to it as the *Bodirsky-Pinsker conjecture.* A structure is *finitely bounded* if it has a finite relational signature and the class of all finite structures embeddable into it is definable by a universal first-order sentence, and *homogeneous* if every isomorphism between two of its finite substructures extends to an automorphism. *Reducts* of such structures are obtained simply by removing some of the original relations. A prototypical example of a structure that satisfies both finite boundedness and homogeneity is $(\mathbb{Q};<)$, the set of rational numbers equipped with the usual countable dense linear order without endpoints. It is a folklore fact that the class of reducts of finitely bounded homogeneous structures is closed under taking expansions of structures by first-order definable relations [8]. Roughly said, the condition imposed on the structures within the scope of the Bodirsky-Pinsker conjecture ensures that the CSP is in NP and that its template enjoys some of the universal-algebraic properties that have played an essential role in the proofs of the Feder-Vardi conjecture [6]. At the same time, it covers CSP-reformulations of many natural problems in qualitative reasoning, as well as all problems definable in MMSNP.

Every countable finitely bounded homogeneous structure is uniquely described by a universal first-order sentence up to isomorphism, and every reduct of such a structure by a sentence of the logic SNP. The CSPs of both kinds of structures are always definable in the monotone fragment of SNP. By Fraïssé's theorem, asking whether a given universal first-order sentence describes a finitely bounded homogeneous structure is equivalent to asking whether the class of its finite models has the *Amalgamation Property* (AP). This question has been considered many times in the context of the Lachlan-Cherlin classification programme for homogeneous structures [3, 36, 26], and is known to be decidable in the case of binary signatures [38, 15]. It also appears as an open problem in Bodirsky's book on infinite-domain constraint satisfaction [8]. Whether a given SNP sentence describes a reduct of a finitely bounded homogeneous structure is a different question, closely related to *homogenizability* [2, 4, 28, 35]. To the best of our knowledge, neither of the two questions is known to be decidable in general. Hence, it is unclear which CSPs actually fall within

[1] The correspondence between MMSNP and finite-domain CSP was initially only achieved up to randomized reductions, but it was later derandomized by Kun [37].

the scope of the Bodirsky-Pinsker conjecture. Besides CSPs, they are also relevant to other areas of theoretical computer science such as verification of database-driven systems [21] or description logics with concrete domains [39, 5]. Below, we state the two questions explicitly.

1. **The amalgamation meta-problem.** Given a universal sentence Φ over a finite relational signature, does there exist a finitely bounded homogeneous structure $\mathfrak{B}$ such that the finite models of Φ are precisely the finite substructures of $\mathfrak{B}$ up to isomorphism?
2. **The homogenizability meta-problem.** Given an SNP sentence Φ over a finite relational signature, does there exist a reduct $\mathfrak{B}$ of a finitely bounded homogeneous structure such that the finite models of Φ are precisely the finite substructures of $\mathfrak{B}$ up to isomorphism?

Contributions. In the present paper, we prove the intractability of the two meta-problems. More specifically, we show that the amalgamation meta-problem is EXPSPACE-hard (Theorem 6) or PSPACE-hard (Theorem 7), depending on the encoding of the input, and that the homogenizability meta-problem is undecidable (Theorem 13). Theorem 6 and Theorem 7 are proved in Section 3.1 by taking a proof-theoretic perspective on the AP for classes defined by universal Horn sentences. We show that, for some of these classes, the failures of the AP are in a 1:1 correspondence with the rejecting runs of certain Datalog programs verifying instances of the rectangle tiling problem. Here, by Datalog we mean the monotone Horn fragment of SNP. Theorem 13 is proved in Section 4, by analyzing model-theoretic properties of a very natural encoding of context-free grammars into Datalog sentences. As a byproduct of the proof, we also get the undecidability of some other properties for Datalog programs, e.g., whether they can be rewritten in the logic MMSNP, whether they solve some finite-domain CSP, or whether they define a structure with a homogeneous Ramsey expansion in a finite relational signature.

It is known that, from every finite structure $\mathfrak{A}$ over a finite relational signature one can construct in polynomial time a finite structure $\mathfrak{B}$ over a finite binary relational signature such that CSP($\mathfrak{A}$) and CSP($\mathfrak{B}$) are polynomial-time equivalent [24, 30]. By our results, such a reduction is unlikely to exist for universal sentences representing finitely bounded homogeneous structures, unless it avoids the amalgamation meta-problem. The reason is that, for binary relational signatures, the amalgamation meta-problem can be decided in coNEXPTIME (Proposition 3). Our results provide evidence for the need for a fundamentally new language-independent approach to the Bodirsky-Pinsker conjecture. First steps in this direction were taken in the recent works of Mottet and Pinsker [44] and Bodirsky and Bodor [9], but they do not fully address the issues stemming from the two meta-problems. We elaborate on this claim below. To keep our results as general as possible, we formulate them for some reasonable promise relaxations of the two meta-problems, i.e., where a subclass and a superclass of the positive instances are being separated from each other with the promise that the input never belongs to the complement of the subclass within the superclass.

The subtleties of the Bodirsky-Pinsker conjecture. In 2016, Bodirsky and Mottet presented an elegant tool for lifting tractability from finite-domain constraint satisfaction to the infinite [17], hereby establishing the first general link between the Feder-Vardi and the Bodirsky-Pinsker conjecture. Since then, their method has been used numerous times to prove new or reprove old complexity classification results for infinite-domain CSPs. One prominent such example is the universal-algebraic proof of the complexity dichotomy for MMSNP [16]. Conveniently enough, every MMSNP sentence defines a finite union of CSPs of structures within the scope of the Bodirky-Pinsker conjecture, so the two meta-problems were not relevant in this context. There is a prospect that the methods from [17] will also prove useful in proving a dichotomy for the even more general logic class *Guarded Monotone* SNP (GMSNP) introduced in [7]. Also GMSNP enjoys the above mentioned property of MMSNP [15], and hence avoids the two meta-problems.

However, outside of GMSNP there exists a regime where the methods from [17] definitely fall short, and where the two meta-problems become relevant. Consider for instance the dichotomy for *temporal CSPs*, i.e., for CSPs of structures with domain $\mathbb{Q}$ and whose relations are definable by a Boolean combination of formulas of the form $(x = y)$ or $(x < y)$, obtained by Bodirsky and Kára in 2010 [14]. At the present time, these problems are already very well understood; tractable temporal CSPs can always be solved by an algorithm that repeatedly searches for a set of potential minimal elements among the input variables, where each instance of the search is performed using an oracle for a tractable finite-domain CSP. The latter is generally determined by the shape of the Boolean combinations. E.g., in the case of $\mathrm{CSP}(\mathbb{Q}; \{(x, y, z) \in \mathbb{Q}^3 \mid (x = y < z) \vee (y = z < x) \vee (z = x < y)\})$, solving the finite-domain CSP in question amounts to solving linear equations modulo 2 [14, 19]. It is known that the tractability results from [14] cannot be obtained using the reduction from [17].

In 2022, Mottet and Pinsker introduced the machinery of *smooth approximations* [44], which vastly generalizes the methods in [17]. The last section of their paper is devoted to temporal CSPs, and the authors manage to reprove a significant part of the dichotomy on just a few pages. They achieve this by applying some of their general results to first-order expansions of $(\mathbb{Q}; <)$ and obtaining either NP-hardness for the CSP, or one of the two types of symmetry that played a fundamental role in the original proof from [14]. This symmetry can then be used to prove correctness of the reduction to a finite-domain CSP described above, but only under an explicit usage of the homogeneity of $(\mathbb{Q}; <)$ (see Proposition 3.1 in [19] and the last section of [44]). In contrast to the methods in [17] which only use homogeneity as a blackbox, this approach can be described as *language-dependent*.

A similar situation occurs in the case of phylogeny CSPs [13], which capture decision problems concerning the existence of a binary tree satisfying certain constraints imposed on its leaves. Tractable phylogeny CSPs are strikingly similar to tractable temporal CSPs; they can always be solved by an algorithm that repeatedly searches for a subdivision of the input variables into two parts, representing the two different branches below the root of a binary tree, where each instance of this search is performed using an oracle for a tractable finite-domain CSP. However, for tractable phylogeny CSPs, already the homogeneity of the infinite-domain CSP template is both sufficient and necessary for proving the correctness of the reduction to the finite-domain CSP (Theorem 6.13 and Lemma 6.12 in [13]). We can therefore speak of a case of extreme language-dependency. Temporal and phylogeny CSPs are special cases of CSPs of structures obtainable from the universal homogeneous binary tree [10] by specifying relations using first-order formulas. Achieving a complexity dichotomy in this context will require a non-trivial combination of the methods from [14] and [13].

An optimal way of approaching the conjecture would be gaining a very good understanding of the class of reducts of finitely bounded homogeneous structures, e.g., through some sort of a classification. However, it is unclear how realistic this prospect is as model-theoretic properties often tend to be undecidable [25]. We remark that homogeneity is a vital part of the Bodirsky-Pinsker conjecture; this assumption can be weakened or strengthened but not dropped entirely, as otherwise we get a class that provably does not have a dichotomy [30, 8].

2 Preliminaries

Relational structures. The set $\{1, \ldots, n\}$ is denoted by $[n]$, and we use the bar notation $\bar{t}$ for tuples. A (*relational*) *signature* τ is a set of *relation symbols*, each $R \in \tau$ with an associated natural number called *arity*. We say that τ is *binary* if it consists of symbols of arity ≤ 2. A (*relational*) *τ-structure* $\mathfrak{A}$ consists of a set A (the *domain*) together with the relations $R^{\mathfrak{A}} \subseteq A^k$ for each $R \in \tau$ with arity k. An *expansion* of $\mathfrak{A}$ is a σ-structure $\mathfrak{B}$ with $A = B$ such that $\tau \subseteq \sigma$, $R^{\mathfrak{B}} = R^{\mathfrak{A}}$ for each relation symbol $R \in \tau$. Conversely, we call $\mathfrak{A}$ a *reduct* of $\mathfrak{B}$. The *union* of two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ is the τ-structure $\mathfrak{A} \cup \mathfrak{B}$ with domain $A \cup B$ and relations of the form $R^{\mathfrak{A}\cup\mathfrak{B}} := R^{\mathfrak{A}} \cup R^{\mathfrak{B}}$ for every $R \in \tau$.

A *homomorphism* $h\colon \mathfrak{A} \to \mathfrak{B}$ for τ-structures $\mathfrak{A}, \mathfrak{B}$ is a mapping $h\colon A \to B$ that *preserves* each relation of $\mathfrak{A}$, i.e., if $\bar{t} \in R^{\mathfrak{A}}$ for some k-ary relation symbol $R \in \tau$, then $h(\bar{t}) \in R^{\mathfrak{B}}$. We write $\mathfrak{A} \to \mathfrak{B}$ if $\mathfrak{A}$ maps homomorphically to $\mathfrak{B}$. The *Constraint Satisfaction Problem* (CSP) of $\mathfrak{A}$, denoted by $\mathrm{CSP}(\mathfrak{A})$, is defined as the class of all finite structures which homomorphically map to $\mathfrak{A}$. An *embedding* is a homomorphism $h\colon \mathfrak{A} \to \mathfrak{B}$ that additionally satisfies the following condition: for every k-ary relation symbol $R \in \tau$ and $\bar{t} \in A^k$ we have $h(\bar{t}) \in R^{\mathfrak{B}}$ only if $\bar{t} \in R^{\mathfrak{A}}$. We write $\mathfrak{A} \hookrightarrow \mathfrak{B}$ if $\mathfrak{A}$ embeds to $\mathfrak{B}$. The *age* of $\mathfrak{A}$, denoted by $\mathrm{age}(\mathfrak{A})$, is the class of all finite structures which embed to $\mathfrak{A}$. A *substructure* of $\mathfrak{A}$ is a structure $\mathfrak{B}$ over $B \subseteq A$ such that the inclusion map $i\colon B \to A$ is an embedding. An *isomorphism* is a surjective embedding. Two structures $\mathfrak{A}$ and $\mathfrak{B}$ are *isomorphic* if there exists an isomorphism from $\mathfrak{A}$ to $\mathfrak{B}$. An *automorphism* is an isomorphism from $\mathfrak{A}$ to $\mathfrak{A}$. The *orbit* of a tuple $\bar{t} \in A^k$ in $\mathfrak{A}$ is the set $\{g(\bar{t}) \mid g \text{ is an automorphism of } \mathfrak{A}\}$. A countable structure $\mathfrak{A}$ is *ω-categorical* if, for every $k \geq 1$, there are only finitely many orbits of k-tuples in $\mathfrak{A}$.

First-order logic. We assume that the reader is familiar with classical *first-order* logic as well as with basic preservation properties of first-order formulas, e.g., that every first-order formula ϕ is preserved by isomorphisms; by embeddings if ϕ is existential, and by homomorphisms if ϕ is existential positive. For a first-order sentence Φ, we denote the class of all its finite models by $\mathrm{fm}(\Phi)$. We say that a first-order formula is *k-ary* if it has k free variables. For a first-order formula ϕ, we use the notation $\phi(\bar{x})$ to indicate that the free variables of ϕ are among $\bar{x}$. This does not mean that the truth value of ϕ depends on each entry in $\bar{x}$. We assume that equality $=$ as well as the nullary predicate symbol $\bot$ for falsity are always available when building first-order formulas. Thus, *atomic τ-formulas*, or *τ-atoms* for short, over a relational signature τ are of the form $\bot$, $(x = y)$, and $R(\bar{x})$ for some $R \in \tau$. We say that a formula is *equality-free* if it does not contain any occurrence of the default equality predicate. If ϕ is a disjunction of possibly negated τ-atoms, then we define the *Gaifman graph* of ϕ as the undirected graph whose vertex set consists of all free variables of ϕ and where two distinct variables x, y form an edge if and only if they appear jointly in a negative atom of ϕ. Let Φ be a universal τ-sentence in prenex normal form whose quantifier-free part ϕ is in CNF. We call Φ *Horn* if every clause of ϕ is Horn, i.e., contains at most one positive disjunct. We call Φ *complete* if the Gaifman graph of each clause of ϕ is complete. It is a folklore fact that, if Φ is complete, then $\mathrm{fm}(\Phi)$ is preserved by unions.

SNP and its fragments. An *SNP τ-sentence* is a second-order sentence Φ of the form $\exists X_1, \dots, X_n \forall \bar{x}.\, \phi$ where ϕ is a quantifier-free formula in CNF over $\tau \cup \{X_1, \dots, X_n\}$. We call Φ *monadic* if X_i is unary for every $i \in [n]$; *monotone* if ϕ does not contain any positive τ-atoms (in particular no positive equality atoms); and *guarded* if, for every positive atom β there exists a negative atom α containing all variables of β. Note that all notions from the previous paragraph easily transfer to SNP sentences viewed as universal sentences in an extended signature. The monadic monotone and the guarded monotone fragments of SNP are denoted by MMSNP and GMSNP, respectively. The monotone Horn fragment of SNP is commonly known as the logic programming language *Datalog*. When we say that a Datalog program Φ *solves* the CSP of a structure $\mathfrak{B}$, we simply mean that $\mathrm{fm}(\Phi) = \mathrm{CSP}(\mathfrak{B})$.

Homogeneity, homogenizability, and finite boundedness. A countable structure $\mathfrak{S}$ is *homogeneous* if every isomorphism between two finite substructures of $\mathfrak{S}$ extends to an automorphism of $\mathfrak{S}$. Clearly, every homogeneous structure in a finite relational signature is ω-categorical, and so are the reducts of such structures. Homogeneous structures arise as limit objects of well-behaved classes of finite structures in the sense of Theorem 1.

Let $\mathcal{K}$ be a class of finite structures in a finite relational signature τ closed under isomorphisms and substructures. We say that $\mathcal{K}$ has the *amalgamation property* (AP) if, for all $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathcal{K}$ whose substructures on $B_1 \cap B_2$ are identical, there exists $\mathfrak{C} \in \mathcal{K}$ together with embeddings $f_1 \colon \mathfrak{B}_1 \hookrightarrow \mathfrak{C}$ and $f_2 \colon \mathfrak{B}_2 \hookrightarrow \mathfrak{C}$ such that $f_1|_{B_1 \cap B_2} = f_2|_{B_1 \cap B_2}$. We refer to $\mathfrak{C}$ as an *amalgam* of $\mathfrak{B}_1$ and $\mathfrak{B}_2$ in $\mathcal{K}$. Note that, for a class closed under isomorphisms and substructures, the AP is implied by the property of being closed under unions $\mathfrak{B}_1 \cup \mathfrak{B}_2$, also called *free amalgams*.

▶ **Theorem 1** (Fraïssé). *For a class $\mathcal{K}$ of finite structures in a finite relational signature τ, the following are equivalent:*

- *$\mathcal{K}$ is the age of an up to isomorphism unique countable homogeneous τ-structure;*
- *$\mathcal{K}$ is closed under isomorphisms, substructures, and has the AP.*

As already mentioned in the introduction, the structure $(\mathbb{Q}; <)$ is homogeneous because every local isomorphism can be extended to an automorphism using a piecewise affine transformation. Its age is the class of all finite strict linear orders.

A countable structure $\mathfrak{S}$ is *homogenizable* if it is a reduct of a homogeneous structure $\mathfrak{H}$ over a finite relational signature such that $\mathfrak{S}$ and $\mathfrak{H}$ have the same sets of automorphisms [28]. Whenever this happens, by the theorem of Ryll-Nardzewski, all relations of $\mathfrak{H}$ are first-order definable in $\mathfrak{S}$ [33]. One might say that $\mathfrak{S}$ already has all the relations necessary for homogeneity but they perhaps do not all have names. A prototypical example of this phenomenon is the universal "homogeneous" binary tree, which is homogenizable but not homogeneous, see, e.g., Proposition 3.2 in [10]. We call a class $\mathcal{K}$ of finite structures in a finite relational signature τ *homogenizable* if it forms the age of a homogenizable structure.

For a class $\mathcal{N}$ of finite structures in a finite relational signature τ, the class $\mathrm{Forb}_e(\mathcal{N})$ consists of all finite τ-structures which do not embed any member of $\mathcal{N}$. Following the terminology in [40], we say that a class $\mathcal{K}$ of finite structures in a finite relational signature is *finitely bounded* if there exists a finite $\mathcal{N}$ such that $\mathcal{K} = \mathrm{Forb}_e(\mathcal{N})$. We refer to $\mathcal{N}$ as a set of *bounds* for $\mathcal{K}$, and define the *size* of $\mathcal{N}$ as the sum of the cardinalities of the domain and the relations of all structures in $\mathcal{N}$. A structure $\mathfrak{S}$ is *finitely bounded* if its age is finitely bounded. We say that a class $\mathcal{K}$ is *finitely bounded homogenizable* if it forms the age of a reduct $\mathfrak{R}$ of a finitely bounded homogeneous structure $\mathfrak{H}$ such that $\mathfrak{R}$ and $\mathfrak{H}$ have the same sets of automorphisms. Sufficient conditions for finitely bounded homogenizability were provided by Hubička and Nešetřil [34], generalizing previous work of Cherlin, Shelah, and Shi [27].

3 The Amalgamation Meta-Problem

By Theorem 1, every homogeneous structure is uniquely described by its age (up to isomorphism). Consequently, every finitely bounded homogeneous structure is uniquely described by a finite set of bounds. It is known that the question whether $\mathrm{Forb}_e(\mathcal{N})$ has the AP for a given finite set of bounds $\mathcal{N}$ can be tested algorithmically in the case where the signature is binary [38]. This decidability result is based on the following observation. A *one-point amalgamation diagram* is an input $\mathfrak{B}_1, \mathfrak{B}_2$ to the AP where $|B_1 \setminus B_2| = |B_2 \setminus B_1| = 1$.

▶ **Proposition 2** ([38]). *A class of finite relational τ-structures that is closed under isomorphisms and substructures has the AP if and only if it has the AP restricted to one-point amalgamation diagrams.*

As a consequence of Proposition 2, if τ is binary and $\mathrm{Forb}_e(\mathcal{N})$ does not have the AP, then the size of a smallest counterexample to the AP is polynomial in the size of $\mathcal{N}$ [15]. Such a counterexample can be non-deterministically guessed and verified using a coNP-oracle, which places the problem at the second level of the polynomial hierarchy (Theorem 15 in [5]).

There is a second, arguably more practical, equivalent definition of finite boundedness. Namely, a class $\mathcal{K}$ of finite structures in a finite relational signature is finitely bounded if and only if there exists a universal sentence Φ such that $\mathcal{K} = \mathrm{fm}(\Phi)$. Using this definition, it is easy to see that $(\mathbb{Q}; <)$ is finitely bounded because its age, the class of all finite strict linear orders, admits a finite universal axiomatization (irreflexivity, transitivity, and totality). From a complexity-theoretical perspective, the two definitions are equivalent only up to a single-exponential blow-up in one direction. Given a finite set of bounds $\mathcal{N}$, we can obtain a universal sentence Φ of size polynomial in the size of $\mathcal{N}$ satisfying $\mathrm{fm}(\Phi) = \mathrm{Forb}_e(\mathcal{N})$ by describing each structure in $\mathcal{N}$ up to isomorphism using a quantifier-free formula. However, given a universal sentence Φ, it can be the case that a smallest $\mathcal{N}$ satisfying $\mathrm{fm}(\Phi) = \mathrm{Forb}_e(\mathcal{N})$ is of size single-exponential in the size of Φ. The reason is that obtaining $\mathcal{N}$ from Φ is comparable to rewriting Φ in DNF. Consequently, the algorithm from [15] only gives us a relatively weak upper bound for the case where the inputs are specified by universal sentences.

▶ **Proposition 3.** *Let Φ be a universal sentence over a finite binary relational signature τ. If $\mathrm{fm}(\Phi)$ does not have the AP, then the size of a smallest counterexample to the AP is at most single-exponential in the size of Φ. Consequently, the question whether $\mathrm{fm}(\Phi)$ has the AP is decidable in* coNEXPTIME*.*

The upper bound provided by Proposition 3 is not unreasonable since a smallest counterexample to the AP might be of size exponential in the size of the input sentence even if the signature is binary. This is demonstrated in Example 4.

▶ **Example 4.** Let τ be the signature consisting of the unary symbols $\{L, R\} \cup \{X_i \mid i \in [n]\}$ and the binary symbols $\{E\} \cup \{Y_i \mid i \in [n]\}$ for some $n \in \mathbb{N}$. Consider the universal sentence

$$\Phi := \forall x, y_1, y_2 \Big(L(y_1) \wedge R(y_2) \wedge E(x, y_1) \wedge E(x, y_2) \wedge \big(\bigwedge\nolimits_{i \in [n]} Y_i(y_1, y_2) \Leftrightarrow X_i(x) \big) \Rightarrow \bot \Big).$$

Our first claim is that $\mathrm{fm}(\Phi)$ does not have the AP. We define the one-point amalgamation diagram $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathrm{fm}(\Phi)$ as follows. The domains are $B_i := \{b_i\} \cup \{b_S \mid S \subseteq [n]\}$, $i \in \{1, 2\}$, and the relations are given by the following conjunction of atomic formulas:

$$L(b_1) \wedge R(b_2) \wedge \bigwedge\nolimits_{S \subseteq [n]} E(b_S, b_1) \wedge E(b_S, b_2) \wedge \bigwedge\nolimits_{i \in S} X_i(b_S).$$

We have that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ satisfy Φ because $R^{\mathfrak{B}_1} = \emptyset$ and $L^{\mathfrak{B}_2} = \emptyset$. Clearly, no amalgam for $\mathfrak{B}_1$ and $\mathfrak{B}_2$ can be obtained by identifying b_1 and b_2 because $L^{\mathfrak{B}_1} = \{b_1\}$ and $R^{\mathfrak{B}_2} = \{b_2\}$. The free amalgam $\mathfrak{B}_1 \cup \mathfrak{B}_2$ does not satisfy Φ because of the assignment $y_1 := b_1$, $y_2 := b_2$, and $x := b_\emptyset$. But since we can assign $x := b_S$ for any $S \subseteq [n]$, also no amalgam satisfying Φ can be obtained by adding the pair (b_1, b_2) to any subset of the relations $Y_1^{\mathfrak{B}_1 \cup \mathfrak{B}_2}, \ldots, Y_n^{\mathfrak{B}_1 \cup \mathfrak{B}_2}$ of the free amalgam. We conclude that $\mathrm{fm}(\Phi)$ does not have the AP.

Our second claim is that every one-point amalgamation diagram $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathrm{fm}(\Phi)$ satisfying $|B_1 \cap B_2| < 2^n$ has an amalgam in $\mathrm{fm}(\Phi)$. Let b_1 and b_2 be the unique elements contained in $B_1 \setminus B_2$ and $B_2 \setminus B_1$, respectively. If $\mathfrak{B}_1 \cup \mathfrak{B}_2 \models \Phi$, then we are done because $\mathfrak{B}_1 \cup \mathfrak{B}_2$ is an amalgam for $\mathfrak{B}_1$ and $\mathfrak{B}_2$. So suppose that $\mathfrak{B}_1 \cup \mathfrak{B}_2 \not\models \Phi$. Consider any evaluation of the quantifier-free part of Φ witnessing the fact that $\mathfrak{B}_1 \cup \mathfrak{B}_2 \not\models \Phi$. Since b_1 and b_2 do not appear together in any relation of $\mathfrak{B}_1 \cup \mathfrak{B}_2$ and $\mathfrak{B}_1, \mathfrak{B}_2 \models \Phi$, by the shape of Φ, it must be the case that x is assigned some element $b \in B_1 \cap B_2$, y_1 is assigned b_1, and y_2 is assigned b_2 (or vice versa). Since $|B_1 \cap B_2| < 2^n$, there must exist $S \subseteq [n]$ such that, for every $b \in B_1 \cap B_2$, it is not the case that $b \in X_i^{\mathfrak{B}_1 \cup \mathfrak{B}_2}$ if and only if $i \in S$. Consequently, we can obtain an amalgam $\mathfrak{C} \in \mathrm{fm}(\Phi)$ for $\mathfrak{B}_1$ and $\mathfrak{B}_2$ by adding the pairs (b_1, b_2) and (b_2, b_1) to $Y_i^{\mathfrak{B}_1 \cup \mathfrak{B}_2}$ for every $i \in S$. We conclude that a smallest counterexample to the AP for $\mathrm{fm}(\Phi)$ is of size $> 2^n$.

Very little progress has been done on signatures containing symbols of arities larger than 2. In particular, it is not even known whether the AP is decidable for finitely bounded classes in general. The scenario where this is not the case is not unrealistic since the closely related *joint embedding property* (JEP) is undecidable already for finitely bounded classes of graphs [22]. The JEP determines whether a finitely bounded class forms the age of any structure, without the requirement of homogeneity [33]. Note that the undecidability of the JEP does not necessarily have any consequences for the Bodirsky-Pinsker conjecture, similarly as it did not have any for the Feder-Vardi conjecture.

If the AP turns out to be undecidable as well (for finitely bounded classes), then the Bodirsky-Pinsker conjecture addresses a class of structures with an undecidable membership problem, at least under the currently best known input to the meta-problem. It seems that, to some extent, the decidability issue can be ignored by only using homogeneity as a blackbox. This was demonstrated in the recent work [45] on the complexity of CSPs of homogeneous uniform hypergraphs, whose classification remains an open problem [3]. We remark that the complexity of the amalgamation meta-problem is already open in the following case, where we can only prove PSPACE-hardness.

▶ **Theorem 5.** *Given a universal Horn sentence Φ over a finite relational signature that is binary except for one ternary symbol, the question whether $\mathrm{fm}(\Phi)$ has the AP is PSPACE-hard.*

The next theorem states that testing the AP becomes properly harder than in the binary case if we do not impose any restrictions on the input (unless coNEXPTIME = EXPSPACE). The fact that this is also true for the strong version of the AP might be of independent interest to model-theorists. The *strong* version of the AP is when $\mathfrak{C} \in \mathcal{K}$ and $f_i \colon \mathfrak{B}_i \hookrightarrow \mathfrak{C}$ for $i \in \{1, 2\}$ can always be chosen so that $f_1(B_1) \cap f_2(B_2) = f_1(B_1 \cap B_2)$. Note that the theorem is formulated as a statement of the form "the question whether X or not even Y is hard." This is a compact way for writing that both X and Y (and every property in between) are hard, the formulation tacitly assumes that the inputs never satisfy "Y and not X."

▶ **Theorem 6.** *Given a universal sentence Φ over a finite signature, the question whether $\mathrm{fm}(\Phi)$ has the strong AP or not even the AP is EXPSPACE-hard.*

Theorem 7 is a variant of Theorem 6 in the setting where the input is specified by a set of bounds instead of a universal sentence. This setting can be compared to the situation where, in Theorem 6, the quantifier-free part of Φ is required to be in DNF. As a consequence, we cannot profit from succinctness of general universal sentences, which leads to a weaker PSPACE lower bound on the complexity. On the other hand, it turns out that PSPACE-hardness is witnessed even by instances whose domain size remains constant.

▶ **Theorem 7.** *Given a finite set $\mathcal{N}$ of finite structures over a finite signature, the question whether $\mathrm{Forb}_e(\mathcal{N})$ has the strong AP or not even the AP is PSPACE-hard. The statement is true even when the domain size for the structures in $\mathcal{N}$ is bounded by a constant.*

3.1 Proofs of Theorem 6 and Theorem 7

Our proofs of Theorem 6 and Theorem 7 are based on the fact that, if Φ is a universal Horn sentence such that $\mathrm{fm}(\Phi)$ does not have the AP, then every counterexample to the AP has the form of a particular Horn clause which can be derived from Φ in a syntactical manner. By padding each Horn clause in Φ with auxiliary negative atoms and hereby increasing its "degree of completeness," we gain some control over the form of the counterexamples to the AP. When this is performed in a careful and systematic way, the counterexamples to the AP can be brought into a 1:1 correspondence with the rejecting runs of certain Datalog programs. In our case, such programs verify the validity of tilings w.r.t. given input parameters to a bounded tiling problem. This is the main technical contribution of the present article.

▶ **Definition 8.** *Let Φ be an equality-free universal Horn sentence over a relational signature τ. Additionally, let $\phi(\bar{x})$ and $\psi(\bar{x},\bar{y})$ be equality-free conjunctions of atomic τ-formulas. We write $\psi(\bar{x},\bar{y}) \leq_\Phi \phi(\bar{x})$ if, for every atomic τ-formula $\chi(\bar{x})$ other than equality,*

$$\Phi \models \forall \bar{x},\bar{y}\big(\psi(\bar{x},\bar{y}) \Rightarrow \chi(\bar{x})\big) \quad \textit{implies} \quad \Phi \models \forall \bar{x}\big(\phi(\bar{x}) \Rightarrow \chi(\bar{x})\big).$$

In the next lemma, we reformulate the (strong) AP using Definition 8.

▶ **Lemma 9.** *Let Φ be an equality-free universal Horn sentence over a relational signature τ. Then the following are equivalent:*

1. $\mathrm{fm}(\Phi)$ *has the strong AP.*
2. $\mathrm{fm}(\Phi)$ *has the AP.*
3. *If $\phi(\bar{x})$, $\phi_1(\bar{x},y_1)$, and $\phi_2(\bar{x},y_2)$ are equality-free conjunctions of atomic formulas, where y_1 and y_2 are distinct variables not contained in $\bar{x}$, such that, for both $i \in \{1,2\}$, every atom in $\phi_i(\bar{x},y_i)$ contains the variable y_i and $\phi(\bar{x}) \wedge \phi_i(\bar{x},y_i) \leq_\Phi \phi(\bar{x})$, then*

$$\phi(\bar{x}) \wedge \phi_1(\bar{x},y_1) \wedge \phi_2(\bar{x},y_2) \leq_\Phi \phi(\bar{x}) \wedge \phi_1(\bar{x},y_1).$$

Let Φ be an equality-free universal Horn sentence and ψ a Horn clause over a relational signature τ. An *SLD-derivation of ψ from Φ* is a finite sequence of Horn clauses $\psi_0,\dots,\psi_s = \psi$ such that ψ_0 is a conjunct in Φ and, for every $i \in [s]$, there exists a Horn clause ϕ_i which is, up to renaming of variables, a conjunct in Φ, and such that ψ_i is obtained from ψ_{i-1} by replacing a negative atom of ψ_{i-1} that appears positively in ϕ_i with all negative atoms of ϕ_i. We say that ψ_i is a *resolvent* of ψ_{i-1} and ϕ_i. We call ψ a *weakening* of a clause ψ' if ψ' can be obtained from ψ by removing any amount of atoms. In particular, ψ is a weakening of itself. There exists an *SLD-deduction* of ψ from Φ, written as $\Phi \vdash \psi$, if ψ is a tautology or a weakening of a Horn clause ψ' that has an SLD-derivation from Φ. The following theorem presents a fundamental property of equality-free universal Horn sentences.

▶ **Theorem 10** (Theorem 7.10 in [46])**.** *Let Φ be an equality-free universal Horn sentence and ψ an equality-free Horn clause, both in a fixed signature τ. Then $\Phi \models \psi$ if and only if $\Phi \vdash \psi$.*

Our hardness proofs are by polynomial-time reductions from the complements of two well-known bounded versions of the tiling problem. Consider the signature σ consisting of the two binary symbols P_h, P_v, as well as the four unary symbols P_ℓ, P_r, P_t, P_b. For natural numbers $m,n \geq 1$, the σ-structure $\mathfrak{R}_{m,n}$ has the domain $[m] \times [n]$ and the relations

$$P_h^{\mathfrak{R}_{m,n}} := \{((i,j),(i+1,j)) \mid i \in [n-1], j \in [m]\},$$
$$P_v^{\mathfrak{R}_{m,n}} := \{((i,j),(i,j+1)) \mid i \in [n], j \in [m-1]\},$$
$$P_\ell^{\mathfrak{R}_{m,n}} := [m] \times \{1\},\ P_r^{\mathfrak{R}_{m,n}} := [m] \times \{n\},$$
$$P_b^{\mathfrak{R}_{m,n}} := \{1\} \times [n],\ P_t^{\mathfrak{R}_{m,n}} := \{m\} \times [n].$$

The *rectangle tiling problem* asks whether, given a natural number n and a finite σ-structure $\mathfrak{T}$, there exists a natural number m such that $\mathfrak{R}_{m,n} \to \mathfrak{T}$. Note that this is just a reformulation of the usual statement using the language of homomorphisms.[2] In contrast to the better-known NP-complete square tiling problem, one dimension of the tiling grid is not part of the input and is existentially quantified instead. As a result, the problem becomes PSPACE-complete [50].[3] One can further increase the complexity by allowing a succinct encoding of the space bound. The input remains the same but now we ask for a a rectangle tiling with 2^n columns. Analogously to the natural complete problems based on Turing machines, this yields a decision problem that is complete for the complexity class EXPSPACE [48].

Inputs specified by universal sentences. Theorem 6 is proved by polynomial-time reduction from the complement of the exponential rectangle tiling problem. From every input, we construct a universal sentence Φ of polynomial size such that $\mathrm{fm}(\Phi)$ has the AP if and only if there exists no exponential rectangle tiling satisfying the given parameters. The sentence Φ is almost Horn but disjunctions of non-negated atoms are used in premises of implications to represent exponentially many Horn clauses in a universal sentence of polynomial size. In the text that follows, we allow ourselves to still call such sentences Horn. Our encoding is very compact; each row, i.e., an ordered sequence of 2^n-many tiles, is represented using a constant amount of variables. This is achieved by storing the information about each individual row in binary using $(n+1)$-ary atoms whose entries always contain at most three variables. We refer to the variables representing rows of the tiling as *path* nodes. In order to check the tiling from bottom to top, i.e., parse a chain of path nodes, we require each pair of subsequent path nodes to be verified by a set of 2^n-many *verifier* nodes. This process ensures the vertical consistency of the tiling as well as the presence of 2^n-many tiles in every row. The precise number of verifier nodes is achieved using combinations of n pairs of unary atoms.

To control the occurrence of amalgamation failures, we first introduce a binary symbol E and two unary symbols L, R. Atoms with these symbols serve no other purpose than to ensure that *almost* each conjunct in Φ is complete, i.e., defines a class of structures that is preserved by taking unions and hence has the AP. More concretely, the premise of almost every Horn clause in Φ has a subformula of the form

$$L(y_1) \wedge R(y_2) \wedge \bigwedge\nolimits_{i\in[k]} E(y_1, x_i) \wedge E(y_2, x_i) \wedge \bigwedge\nolimits_{j\in[k]} E(x_i, x_j)$$

making the Horn clause almost complete, with the exception of one potentially missing edge in the Gaifman graph between y_1 and y_2. Our intention is to make this missing edge the only place at which potential faulty one-point amalgamation diagrams can be built (see Figure 1). The sentence Φ is defined as $\Phi_1 \wedge \Phi_2$, where the two parts are described below.

The first part Φ_1 does not yet explain how our reduction works, but ensures that it does not fall apart, e.g., due to ill-behaved identifications of variables. For every $\alpha \in T$, the signature τ contains an $(n+1)$-ary symbol T_α. The first n arguments in a T_α-atom serve as binary counters, and the last argument carries a given path node p. Suppose that the variables $\mathtt{0}$ and $\mathtt{1}$ represent the bits 0 and 1, respectively. Then each atomic formula $T_\alpha(c_1, \dots, c_n, p)$ with $c_1, \dots, c_n \in \{\mathtt{0}, \mathtt{1}\}$ represents the situation in which a tile α is present in the p-th row and in the $\big(1 + \sum_{k\in[n]}(c_k = \mathtt{1}) \cdot 2^{n-k}\big)$-th column. First, we want to ensure the horizontal consistency of the tiling. To this end, for every pair $(\alpha, \beta) \in T^2 \setminus P_h^{\mathfrak{T}}$, we

[2] For comparison, see, e.g., Section 4 in [32].
[3] In [50], the rectangle tiling problem is called the corridor tiling problem.

include in Φ_1 a complete Horn sentence without positive atoms ensuring that two horizontally adjacent positions in the p-th row cannot be tiled with α and β. Here, we encode the successor relation w.r.t. binary addition using a combination of equalities: $(c_{n+1}, \dots, c_{2n})$ is the successor of $(c_1, \dots, c_n)$ if and only if there exists $j \in [n]$ such that $c_i = c_{i+n} \in \{\mathtt{0}, \mathtt{1}\}$ for every $i \in [j-1]$, $c_j = \mathtt{0}$ and $c_{n+j} = \mathtt{1}$, and $c_i = \mathtt{1}$ and $c_{i+n} = \mathtt{0}$ for every $i \in [n] \setminus [j]$. This encoding only makes sense if $\mathtt{0}$ and $\mathtt{1}$ truly represent the bits 0 and 1, so we introduce a simple mechanism (in terms of a Horn sentence) for distinguishing between the variables $\mathtt{0}$ and $\mathtt{1}$. Next, we want to ensure that every position in the p-th row is occupied by at most one tile. To this end, for every $(\alpha, \beta) \in T^2$ with $\alpha \neq \beta$, we include

$$\forall p, \mathtt{0}, \mathtt{1}, c_1, \dots, c_n \big(T_\alpha(c_1, \dots, c_n, p) \wedge T_\beta(c_1, \dots, c_n, p) \wedge \bigwedge\nolimits_{i \in [n]} (c_i = \mathtt{0} \vee c_i = \mathtt{1}) \Rightarrow \bot\big)$$

as a conjunct in Φ_1. Finally, we want to ensure that each verifier node v represents at most one number from $[2^n]$. For every $i \in [n]$, τ contains two unary symbols $\mathbf{0}_i$ and $\mathbf{1}_i$ which will be used to encode numbers in binary. We include $\forall v \big(\bigvee_{i \in [n]} \mathbf{0}_i(v) \wedge \mathbf{1}_i(v) \Rightarrow \bot \big)$ as the last conjunct in Φ_1. Now the idea is that the combinations of atomic formulas $\mathbf{0}_i(v)$ and $\mathbf{1}_i(v)$ at a verifier node v will be compared with the combinations of $\mathtt{0}$ and $\mathtt{1}$ in atomic formulas $T_\alpha(c_1, \dots, c_n, p)$. The Horn sentences in the second part of Φ will be formulated so that verifying the presence of all 2^n atoms of the form $T_\alpha(c_1, \dots, c_n, p_1)$ at a path node p_1 using verifier nodes $v_1, \dots, v_{2^n}$ is the only possible way to progress to a next path node p_2. Consequently, we do not need to add an explicit requirement for rows, represented by path nodes, to be completely tiled from left to right. For the same reason, we also do not need to add an explicit requirement for verifier nodes to represent at least one number from $[2^n]$.

We now proceed with the sentence Φ_2, which explains how the parsing of a tiling actually works. The parsing of a tiling starts from a path node p representing a row whose leftmost position contains a tile that can be present in the bottom left corner of a tiling grid. This must be confirmed by a verifier node, in which case a 6-ary Q_b-atom is derived, representing the fact that the leftmost column of the p-th row has been checked. To this end, we include in Φ_2 suitable Horn sentences for every $\alpha \in P_\ell^{\mathfrak{T}} \cap P_b^{\mathfrak{T}}$. These sentences form *the* non-complete part of Φ; we intentionally leave a missing edge between y_1 and y_2 in the Gaifman graph to enable the formation of potential AP-counterexamples. Moreover, the variables y_1 and y_2 appear in the 1st and the 2nd entry of the derived atom, respectively, and this invariant is maintained throughout the whole construction of Φ_2.

Using 2^n-many verifier nodes and propagation of Q_b-atoms, the whole bottom row is checked for the presence of tiles. Their horizontal consistency already follows from the conditions imposed on path nodes by Φ_2 and needs not to be checked during this step. To this end, we include in Φ_2 suitable Horn sentences for every $\alpha \in P_b^{\mathfrak{T}}$. After the p-th row has been checked by a 2^n-th verifier node, we mark p with a Q-atom indicating that the parsing can progress to a successor path node. To this end, we include in Φ_2 suitable Horn sentences for every $\alpha \in P_b^{\mathfrak{T}} \cap P_r^{\mathfrak{T}}$. The successor relation for path nodes is represented by the binary symbol S, and the certificate of vertical verification for pairs (p_1, p_2) of successive path nodes is represented by the 7-ary symbol Q_v. For every $(\alpha, \beta) \in P_v^{\mathfrak{T}} \cap (P_\ell^{\mathfrak{T}})^2$, we include in Φ_2 a Horn sentence verifying the vertical consistency of the leftmost positions in the rows p_1 and p_2 and deriving the first Q_v-atom containing p_1 and p_2, but only if there is a Q-atom containing p_1. Next, for every $(\alpha, \beta) \in P_v^{\mathfrak{T}}$, we include in Φ_2 a Horn sentence verifying the vertical consistency for the intermediate positions by deriving further Q_v-atoms containing p_1 and p_2. And finally, for every $(\alpha, \beta) \in P_v^{\mathfrak{T}} \cap (P_r^{\mathfrak{T}})^2$, we include in Φ_2 a Horn sentence verifying the vertical consistency of the leftmost positions in the rows p_1 and p_2 and deriving the first Q-atom containing p_2 only. The top row is verified using a 6-ary symbol Q_t similarly as the bottom row; however, the verification of the rightmost position in the top row results in the derivation of $\bot$.

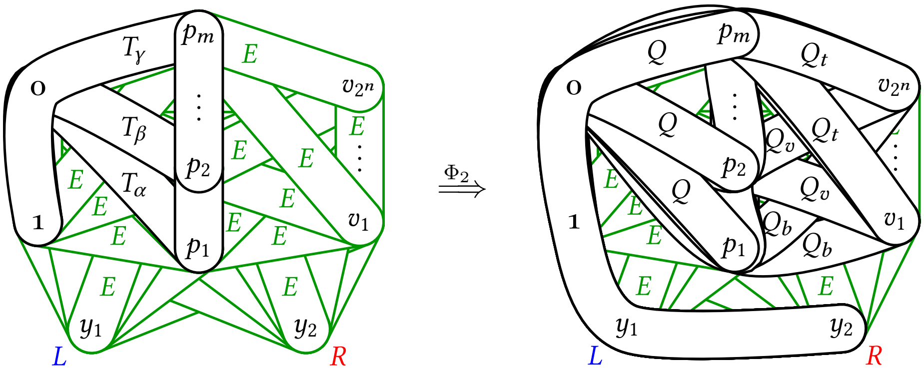

Figure 1 An illustration of an AP-counterexample representing a verification of a valid tiling of an exponential rectangle with m rows and 2^n columns.

Proof of Theorem 6. We first argue that Φ is equivalent to a particular Horn sentence, to which we can apply Lemma 9. By construction, the quantifier-free part of each conjunct in Φ has the form of an implication where the premise possibly also contains instances of disjunction, which are not allowed in Horn clauses, but no instances of negation. Therefore, it can be rewritten as a conjunction of Horn clauses by converting the premise into positive DNF and then considering each disjunct as a separate premise. As a result, the size of the sentence increases exponentially, but this does not matter for the purpose of the proof. Subsequently, all equality atoms can be eliminated by replacing each variable c_i with either $\mathtt{0}$ or $\mathtt{1}$. We denote the resulting Horn sentence by $\overline{\Phi}$ and the two parts stemming from Φ_1 and Φ_2 by $\overline{\Phi}_1$ and $\overline{\Phi}_2$, respectively.

"$\Rightarrow$" Suppose that there exists a tiling $f\colon \mathfrak{R}_{m,2^n} \to \mathfrak{T}$. Guided by f, we define a one-point amalgamation diagram $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathrm{fm}(\overline{\Phi})$ which has no amalgam in $\mathrm{fm}(\overline{\Phi})$ (see Figure 1). The domains are $B_i \coloneqq \{y_i, p_1, \dots, p_m, v_1, \dots, v_{2^n}, \mathtt{0}, \mathtt{1}\}$ for $i \in \{1, 2\}$, and the relations are given by the following conjunctions of atomic formulas. We require $T_\alpha(c_1, \dots, c_n, p_j)$ for $c_1, \dots, c_n \in \{\mathtt{0}, \mathtt{1}\}$ if and only if $f(1 + \sum_{k\in[n]}(c_k = \mathtt{1}) \cdot 2^{n-k}, j) = i$. Next, we require all of the L, R, and E-atoms necessary for enabling the Horn clauses in $\mathfrak{B}_1 \cup \mathfrak{B}_2$. Finally, we require $\bigwedge_{i\in[m-1]} S(p_i, p_{i+1})$ to define a successor chain through path nodes, and $\mathtt{0}_i(v_j)$ or $\mathtt{1}_i(v_j)$ if and only if $j = 1 + \sum_{k\in[n]} \lambda_k \cdot 2^{n-k}$ for $\lambda_1, \dots, \lambda_n \in \{0, 1\}$ and $\lambda_i = 0$ or $\lambda_i = 1$, respectively. Clearly, the tiling atoms are placed correctly and the verifier nodes correctly represent values in $[2^n]$. Since $R^{\mathfrak{B}_1} = \emptyset$ and $L^{\mathfrak{B}_2} = \emptyset$, we have $\mathfrak{B}_1, \mathfrak{B}_2 \models \overline{\Phi}_2$. Since f is horizontally consistent, we have $\mathfrak{B}_1, \mathfrak{B}_2 \models \overline{\Phi}_1$, i.e., $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathrm{fm}(\overline{\Phi})$. But since f is also vertically consistent and $\overline{\Phi}_2$ is a universal Horn sentence, we have $\mathfrak{C} \not\models \overline{\Phi}_2$ for every τ-structure $\mathfrak{C}$ with a homomorphism from $\mathfrak{B}_1 \cup \mathfrak{B}_2$. Hence, $\mathrm{fm}(\overline{\Phi})$ does not have the AP.

"$\Leftarrow$" Suppose that $\mathrm{fm}(\overline{\Phi})$ does not have the AP. Then there exists a counterexample to item (3) in Lemma 9, i.e., there exists a Horn clause ψ of the form $\phi(\bar{x}) \wedge \phi_1(\bar{x}, y_1) \wedge \phi_2(\bar{x}, y_2) \Rightarrow \chi$, where ϕ, ϕ_1, and ϕ_2 satisfy the prerequisites of item (3) in Lemma 9 and $\chi(\bar{x}, y_1)$ is an atomic τ-formula other than equality, such that

$$\overline{\Phi} \models \forall \bar{x}, y_1, y_2 \big(\phi \wedge \phi_1 \wedge \phi_2 \Rightarrow \chi\big), \tag{1}$$
$$\overline{\Phi} \not\models \forall \bar{x}, y_1 \big(\phi \wedge \phi_1 \Rightarrow \chi\big). \tag{2}$$

We choose ψ minimal with respect to the number of its atomic subformulas. By Theorem 10, ψ has an SLD-deduction from $\overline{\Phi}$. Note that, by (2), $\chi(\bar{x}, y_1)$ cannot be a subformula of $\phi(\bar{x}) \wedge \phi_1(\bar{x}, y_1)$. Also, $\chi(\bar{x}, y_1)$ cannot be a subformula of $\phi_2(\bar{x}, y_2)$ because every atom in

$\phi_2(\bar{x}, y_2)$ contains the variable y_2 which does not appear in $\chi(\bar{x}, y_1)$. Hence, $\chi(\bar{x}, y_1)$ is not a subformula of $\phi(\bar{x}) \wedge \phi_1(\bar{x}, y_1) \wedge \phi_2(\bar{x}, y_2)$, i.e., ψ is not a tautology. Consequently, ψ is a weakening of a Horn clause ψ' which has an SLD-derivation $\psi'_0, \dots, \psi'_s = \psi'$ from $\overline{\Phi}$. Recall that every atom from ψ' appears in ψ.

▷ Claim 11. $\overline{\Phi}_2 \vdash \psi'$.

Proof. We start by showing that ψ'_0 is a conjunct of $\overline{\Phi}_2$. Suppose, on the contrary, that ψ'_0 is a conjunct of $\overline{\Phi}_1$. Then ψ' does not contain any positive atom because, by construction, Horn sentences in $\overline{\Phi}_1$ do not contain any positive atoms. By construction, in $\overline{\Phi}$ there is no Horn clause containing a positive atom that occurs negatively in a Horn clause from $\overline{\Phi}_1$, i.e., it is impossible to take resolvents of ψ'_0 and Horn clauses from $\overline{\Phi}$. It follows that $s = 0$. Also, every Horn clause from $\overline{\Phi}_1$ is complete. Since there is no edge between y_1 and y_2 in the Gaifman graph of ψ, and, for $i \in \{1, 2\}$, each atom in ϕ_i contains the variable y_i, either ϕ_1 or ϕ_2 must be empty. Since ψ' does not contain any positive atom and $\phi(\bar{x}) \wedge \phi_i(\bar{x}, y_i) \leq_{\overline{\Phi}} \phi(\bar{x})$ for both $i \in \{1, 2\}$, we get a contradiction to (2). Thus, ψ'_0 must be a conjunct of $\overline{\Phi}_2$. Since no Horn clause in $\overline{\Phi}_1$ contains a positive atom, no Horn clause from $\overline{\Phi}_1$ can be used as a resolvent. We conclude that ψ' has an SLD-derivation from $\overline{\Phi}_2$. ◁

▷ Claim 12. ψ' contains no positive atoms, and no atoms with a symbol from $\{Q_b, Q_v, Q_t, Q\}$.

Proof. By the construction of $\overline{\Phi}_2$, for every $i \in [s]$, if ψ'_{i-1} contains variables z_1, z_2 such that

$$\begin{array}{c}\text{every atomic subformula with a symbol from } \{Q_b, Q_v, Q_t, Q\} \\ \text{contains } z_1 \text{ in its 1st and } z_2 \text{ in its 2nd argument, respectively,}\end{array} \tag{3}$$

then this is also the case for ψ'_i, for the same variables z_1, z_2 up to renaming. Since every possible choice of ψ'_0 from $\overline{\Phi}_2$ initially satisfies (3), it follows via induction that (3) holds for $\psi' = \psi'_s$ for some variables z_1, z_2. Next, we show that $\{z_1, z_2\} = \{y_1, y_2\}$ holds for the pair z_1, z_2 satisfying (3) for ψ'. Suppose, on the contrary, that both z_1 and z_2 are among $\bar{x}, y_1$ or $\bar{x}, y_2$. By construction, the only Horn clauses in $\overline{\Phi}_2$ that are not complete have the property that the incompleteness is only due to one missing edge in the Gaifman graph between two distinguished variables satisfying (3). Therefore, for every $i \in [s]$, ψ'_i is a resolvent of ψ'_{i-1} and a Horn clause from $\overline{\Phi}_2$ which is almost complete except possibly for one missing edge in the Gaifman graph between a pair of variables which must be substituted for the pair (z_1, z_2) satisfying (3) for ψ'_{i-1}. Since the variables y_1 and y_2 do not appear together in any atom in ψ' and $\{z_1, z_2\} \neq \{y_1, y_2\}$, they also do not appear together in any atom during the SLD-derivation. Then it follows from the fact that ϕ, ϕ_1, and ϕ_2 satisfy the prerequisites of item (3) in Lemma 9 that we already have $\overline{\Phi}_2 \vdash \forall \bar{x}, y_1 \big(\phi \wedge \phi_1 \Rightarrow \chi\big)$, a contradiction to (2). Since $\{z_1, z_2\} = \{y_1, y_2\}$ holds for the pair z_1, z_2 satisfying (3) for ψ', ψ' cannot contain any negative atoms with a symbol from $\{Q_b, Q_v, Q_t, Q\}$. Suppose that the conclusion of ψ'_0 is not $\perp$. Then, by construction, ψ' contains a positive atom with a symbol from $\{Q_b, Q_v, Q_t, Q\}$. But then, since z_1, z_2 with $\{z_1, z_2\} = \{y_1, y_2\}$ satisfy (3) for ψ', the said positive atom in ψ' contains both variables y_1 and y_2. This leads to a contradiction to (1) where we assume that the positive atom in ψ may may only contain variables from $\bar{x}, y_1$. Thus the conclusion of ψ'_0 is $\perp$, which means that χ equals $\perp$ due to the minimality assumption. ◁

It remains to show that the existence of such ψ' implies the existence of a tiling $f : \mathfrak{R}_{m,2^n} \to \mathfrak{T}$. By the first and the second claim, the conclusion of ψ'_0 is $\perp$. Since ψ' does not contain any atoms with a symbol from $\{Q_b, Q_v, Q_t, Q\}$, the last such atom must have been eliminated from ψ'_{s-1} by taking a resolvent with one of the incomplete Horn clauses in $\overline{\Phi}_2$. By the construction of $\overline{\Phi}_2$, to obtain ψ' through an SLD-derivation $\psi'_0, \dots, \psi'_s = \psi'$ from $\overline{\Phi}_2$, all Horn clauses introduced in the definition of $\overline{\Phi}_2$ must have been used in the intended order. Recall that we have replaced each variable c_i in Φ_2 with either $\mathtt{0}$ or $\mathtt{1}$ while rewriting Φ_2 as an Horn sentence. Every Horn clause from $\overline{\Phi}_2$ has the property that every positive atomic subformula has a symbol from $\{Q_b, Q_v, Q_t, Q\}$ and contains all variables that appear in a negative atomic subformula, with the following two exceptions. First, verifier nodes are not carried over in any atoms because their only contribution is the encoding of a unique number. Second, after a pair of successive rows has been checked by deriving a Q_v-atom containing a 2^n-th verifier node, the variable representing the lower row is not carried over in any atom because it is no longer needed. Since ϕ, ϕ_1, and ϕ_2 satisfy the prerequisites of item (3) in Lemma 9, no ill-behaved variable identifications might have occurred during the SLD-derivation above as otherwise, we would have $\overline{\Phi}_1 \models \forall \bar{x}, y_1 \big(\phi \wedge \phi_1 \Rightarrow \chi\big)$, a contradiction to (2). Consequently, the SLD-derivation must have the full intended length $(2^n + 1) \cdot m$ for some $m \geq 1$, because every intermediate stage starts and ends with verifier nodes encoding the numbers 2^n and 1, respectively, and one can only progress in steps which decrement the encoded number by one. Clearly, the SLD-derivation witnesses the existence of $f \colon \mathfrak{R}_{m,2^n} \to \mathfrak{T}$. ◀

Inputs specified by sets of bounds. We continue with the proof of Theorem 7. This time, we reduce from the complement of the basic rectangle tiling problem. The proof strategy is similar. In particular, we include in τ the two auxiliary unary symbols L, R and the binary auxiliary symbol E. However, the encoding of the tiles is different. We include in τ a symbol I of arity $\lceil \log_2 |T| \cdot n \rceil + 1$. The first $\lceil \log_2 |T| \cdot n \rceil$ entries serve as binary counters to represent the pairs $(i, \alpha) \in [n] \times T$ in binary, and the last entry carries a path node representing a row of the tiling grid. Each pair $(i, \alpha) \in [n] \times T$ is to be interpreted as the fact that the i-th column in the row represented by a particular path node contains the tile α. The reason for this choice of encoding is that we aim to construct a universal sentence Φ which is equivalent to a set of forbidden substructures $\mathcal{N}$ of size polynomial in $|T| \cdot n$. To achieve this, we use a constant number of symbols whose arity is logarithmic in the size of the input.

Suppose that $\mathtt{0}$ and $\mathtt{1}$ are two variables representing the bits 0 and 1, respectively. For each pair $(i, \alpha) \in [n] \times T$, the ternary formula $\mathrm{TILE}_{i,\alpha}(\mathtt{0}, \mathtt{1}, p)$ is the I-atom whose last entry contains the variable p and the first $\lceil \log_2 |T| \cdot n \rceil$ entries contain the variables $\mathtt{0}$ and $\mathtt{1}$ in the unique way that represents the number $i \cdot \alpha$ in binary when read from left to right. Note that the number of such formulas is polynomial in the size of the input to the tiling problem. In contrast to the proof of Theorem 6, it is not necessary to introduce any verifier nodes as the number of columns in the tiling grid is polynomial in the size of the input. The sentence Φ is defined similarly as in the proof of Theorem 6, so we only provide a general overview and highlight the main differences. We want each row to be horizontally consistent. For all $i \in [n-1]$ and $(\alpha, \beta) \in T^2 \setminus P_h^{\mathfrak{T}}$, we include the following sentence as a conjunct in Φ_1:

$$\forall p, \mathtt{0}, \mathtt{1} \big(\psi(\mathtt{0}, \mathtt{1}) \wedge \mathrm{TILE}_{i,\alpha}(\mathtt{0}, \mathtt{1}, p) \wedge \mathrm{TILE}_{i+1,\beta}(\mathtt{0}, \mathtt{1}, p) \Rightarrow \bot\big),$$

where $\psi(\mathtt{0}, \mathtt{1})$ represents a simple mechanism for distinguishing between $\mathtt{0}$ and $\mathtt{1}$. We also want each position in a given row to be occupied by at most one tile. For all $i \in [n], \alpha, \beta \in T$, we include the following sentence as a conjunct in Φ_1:

$$\forall p, \mathtt{0}, \mathtt{1} \big(\mathrm{TILE}_{i,\alpha}(\mathtt{0}, \mathtt{1}, p) \wedge \mathrm{TILE}_{i,\beta}(\mathtt{0}, \mathtt{1}, p) \Rightarrow \bot\big).$$

As in the proof of Theorem 7, we do *not* need to include an explicit condition stating that each row must be completely tiled from left to right. For the purpose of verifying the validity of a tiling, we include in τ a 5-ary symbol Q, a $(\lceil \log_2 |T| \cdot n \rceil + 4)$-ary symbol Q_v, and two $(\lceil \log_2 |T| \cdot n \rceil + 3)$-ary symbols Q_b, Q_t. For each pair $(i, \alpha) \in [n] \times T$, the formulas

$$\mathrm{BOT}_{i,\alpha}(0, 1, p, y_1, y_2), \quad \mathrm{TOP}_{i,\alpha}(0, 1, p, y_1, y_2), \quad \text{and} \quad \mathrm{VERT}_{i,\alpha}(0, 1, p_1, p_2, y_1, y_2)$$

are defined analogously to $\mathrm{TILE}_{i,\alpha}(0, 1, p)$ but using the Q_b, Q_t and Q_v-atoms instead. We use them to verify the bottom row, top row, and the vertical consistency of a given tiling. In contrast to $\mathrm{TILE}_{i,\alpha}$, the parameter α is not important in $\mathrm{BOT}_{i,\alpha}$, $\mathrm{TOP}_{i,\alpha}$, and $\mathrm{VERT}_{i,\alpha}$.

We now explain how to convert Φ into a set of forbidden substructures. Let $\mathcal{N}$ be the class of all τ-structures with the domain $[i]$ for some $i \in [6]$ that do not satisfy Φ. Since Φ only uses six variables, we have $\mathrm{fm}(\Phi) = \mathrm{Forb}_e(\mathcal{N})$. It remains to show that there exists a polynomial that bounds the size of $\mathcal{N}$. Since there is a constant number of domains of structures in $\mathcal{N}$ and their sizes are also constant, it is enough to show that there exists a polynomial that bounds the number of structures in $\mathcal{N}$. The only non-constant parameters in the construction are the four symbols I, Q_b, Q_t, and Q_v whose arity grows logarithmically with $|T| \cdot n$. Thus, there exists a constant c such that the number of structures in $\mathcal{N}$ is bounded by $c \cdot \left(2^{\lceil \log_2 |T| \cdot n \rceil}\right)^4 \leq c \cdot (2 \cdot |T| \cdot n)^4$. The rest is analogous to the proof of Theorem 6.

4 The Homogenizability Meta-Problem

Every reduct $\mathfrak{R}$ of a finitely bounded homogeneous structure $\mathfrak{H}$ is uniquely described by an SNP sentence, which can be obtained from a universal sentence for $\mathrm{age}(\mathfrak{H})$ by existentially quantifying all the surplus predicates upfront. This is (arguably) the most natural representation for such structures. The homogenizability meta-problem asks whether a given SNP τ-sentence Φ is logically equivalent to an SNP τ-sentence $\Psi = \exists Y_1, \dots, Y_m \forall \bar{y}.\, \psi$ such that $\mathrm{fm}(\forall \bar{y}.\, \psi)$ has the AP in the signature $\tau \cup \{Y_1, \dots, Y_m\}$. We are additionally interested in the refinement of the question where we require the homogeneous structure from Theorem 1 associated to $\mathrm{fm}(\forall \bar{y}.\, \psi)$ to have the same set of automorphisms as its reduct to the original signature τ. This amounts to asking whether $\mathrm{fm}(\Phi)$ is finitely bounded homogenizable.

MMSNP was presented in [30] as a large subclass of SNP which has a dichotomy between P and NP-completeness if and only if the class of all finite-domain CSPs has one. The latter has been confirmed, and the dichotomy for MMSNP has received a new universal-algebraic proof within the programme attacking the Bodirsky-Pinsker conjecture [16]. The new proof relies on the observation that every MMSNP sentence Φ is equivalent to a finite disjunction $\Phi_1 \vee \dots \vee \Phi_n$ of MMSNP sentences such that, for every $i \in [n]$, there exists a reduct $\mathfrak{R}_i$ of a finitely bounded homogeneous structure $\mathfrak{H}_i$ such that $\mathrm{fm}(\Phi_i) = \mathrm{CSP}(\mathfrak{R}_i)$.[4] Moreover, the structure $\mathfrak{H}$ can be chosen so that its age has the *Ramsey property*, which plays an essential role in an argument in [16] showing that the authors correctly identified all of the tractable cases. The exact definition of this property is not essential to the present article and is therefore omitted. GMSNP was first introduced in [41] in its seemingly weaker form MMSNP_2, as a generalization of MMSNP where "monadic" second-order variables may also range over atomic formulas. It was later shown that relaxing the above requirement for monotone SNP to guardedness does not result in a more expressive logic [7]. There is a prospect that GMSNP will also have dichotomy between P and NP-completeness since it enjoys similar model-theoretic properties as MMSNP [15].

Theorem 13 is the most general version of our undecidability result. It applies not only to the original formulation of the homogenizability meta-problem, but also to its generalization to ω-categorical structures. The second item of the theorem might give the impression that

[4] This observation was first made in the the proof of Theorem 7 in [11].

one cannot effectively distinguish between CSPs of reducts of finitely bounded structures and CSPs of reducts of finitely bounded homogeneous structures. Recall that the former class does not have a dichotomy [30, 8]. However, as indicated by the formulation of the second item, all CSPs of reducts of finitely bounded homogeneous structures in the proof of Theorem 13 are in fact finite-domain CSPs, for which there is a dichotomy. Therefore, Theorem 13 merely shows that SNP sentences are an exceptionally bad choice of an input to the question, albeit one that is often used [4, 43].

▶ **Theorem 13.** *For a given a Datalog sentence Φ using at most binary relation symbols, it is undecidable whether:*

1. Φ *is logically equivalent to a monadic Datalog sentence,*
   or Φ *is not even logically equivalent to any GMSNP sentence;*
2. Φ *simultaneously satisfies the following three conditions:*
   - $\mathrm{fm}(\Phi)$ *is the CSP of a finite structure,*
   - $\mathrm{fm}(\Phi)$ *is a finitely bounded homogenizable class,*
   - $\mathrm{fm}(\Phi)$ *is the age of a reduct of a finitely bounded homogeneous Ramsey structure,*

   or $\mathrm{fm}(\Phi)$ *is not even the CSP or the age of any ω-categorical structure.*

The following corollary extracts the statement originally announced in the introduction.

▶ **Corollary 14.** *It is undecidable whether a given SNP sentence defines the age of a reduct of a finitely bounded homogeneous structure. The statement is true even if the SNP sentence comes from the Datalog fragment and uses at most binary relation symbols.*

4.1 A proof of Theorem 13

As usual, the *Kleene plus* and the *Kleene star* of a finite set of symbols Σ, denoted by Σ^+ and Σ^*, are the sets of all finite words over Σ of lengths ≥ 1 and ≥ 0, respectively.

A *context-free grammar* (CFG) is a 4-tuple $\mathscr{G} = (N, \Sigma, P, S)$ where N is a finite set of *non-terminal symbols*, Σ is a finite set of *terminal symbols*, P is a finite set of *production rules* of the form $A \to w$ where $A \in N$ and $w \in (N \cup \Sigma)^+$, $S \in N$ is the *start symbol.* For $u, v \in (N \cup \Sigma)^+$ we write $u \to_G v$ if there are $x, y \in (N \cup \Sigma)^+$ and $(A \to w) \in P$ such that $u = xAy$ and $v = xwy$. The *language of* $\mathscr{G}$ is $L(\mathscr{G}) := \{w \in \Sigma^+ \mid S \to^*_{\mathscr{G}} w\}$, where $\to^*_{\mathscr{G}}$ denotes the transitive closure of $\to_{\mathscr{G}}$. Note that with this definition the *empty word* ϵ can never be an element of $L(\mathscr{G})$; some authors use a modified definition that also allows rules that derive ϵ, but for our purposes the difference is not essential. A context-free grammar is called *(left-)regular* if its production rules are always of the form $A \to a$ or $A \to Ba$ for non-terminal symbols A, B and a terminal symbol a. For a finite set Σ, we call a set $L \subseteq \Sigma^+$ *regular* if it is the language of a regular grammar with terminal symbols Σ.

▶ **Example 15.** Consider the CFG $\mathscr{G}$ with a single terminal symbol a, non-terminal symbols S, A, B, C, and production rules $S \to a$, $S \to aa$, $S \to aaa$, $S \to Aa$, $A \to Ba$, $B \to Ca$, $C \to Ca$, and $C \to a$. Clearly, $\mathscr{G}$ is not regular. However, $L(\mathscr{G}) = \{a\}^+$ is regular.

Let $\mathscr{G} = (N, \Sigma, P, S)$ be a CFG. The signature τ_Σ consists of the unary symbols I, T and the binary symbols R_a for every $a \in \Sigma$, and the signature τ_N consists of a binary symbol R_a for every element $a \in N$. For $a_1 \dots a_n \in \Sigma^+$, we set $\phi_{a_1 \dots a_n}(x_1, \dots, x_{n+1}) := \bigwedge_{i \in [n]} R_{a_i}(x_i, x_{i+1})$. Let $\overline{\Phi}_{\mathscr{G}}$ be the universal Horn sentence over the signature $\tau_\Sigma \cup \tau_N$ whose quantifier-free part contains, for every $(A, w) \in P$, the Horn clause $\phi_w(x_1, \dots, x_{|w|+1}) \Rightarrow R_A(x_1, x_{|w|+1})$, and additionally the Horn clause $I(x_1) \wedge R_S(x_1, x_2) \wedge T(x_2) \Rightarrow \bot$. Then $\Phi_{\mathscr{G}}$ is the Datalog sentence obtained from $\overline{\Phi}_{\mathscr{G}}$ by existentially quantifying all symbols from τ_N

upfront. This encoding of CFGs into Datalog programs is standard (Exercise 12.26 in [1]), and the correspondence provided by the next lemma can be shown via a straightforward induction. For a proof, we refer the reader to [20].

▶ **Lemma 16.** *For a τ_Σ-structure $\mathfrak{A}$, we have $\mathfrak{A} \models \Phi_{\mathfrak{G}}$ if and only if, for every $w \in L(\mathfrak{G})$,*

$$\mathfrak{A} \models \forall x_1, \dots, x_{|w|+1} \big(I(x_1) \wedge \phi_w(x_1, \dots, x_{|w|+1}) \wedge T(x_{|w|+1}) \Rightarrow \bot \big).$$

The following lemma is proved by establishing a connection between the well-known Myhill-Nerode correspondence and ω-categoricity, under the addition of several auxiliary results from [29, 15].

▶ **Lemma 17.** *Let $\mathfrak{G}$ be a context-free grammar. Then the following are equivalent:*

1. *$L(\mathfrak{G})$ is regular.*
2. *$\Phi_{\mathfrak{G}}$ is equivalent to a monadic Datalog sentence.*
3. *$\Phi_{\mathfrak{G}}$ is equivalent to a GMSNP sentence.*
4. *$\mathrm{fm}(\Phi_{\mathfrak{G}})$ is the CSP of a finite structure.*
5. *$\mathrm{fm}(\Phi_{\mathfrak{G}})$ is the CSP of an ω-categorical structure.*
6. *$\mathrm{fm}(\Phi_{\mathfrak{G}})$ is the age of a reduct of a finitely bounded homogeneous Ramsey structure.*
7. *$\mathrm{fm}(\Phi_{\mathfrak{G}})$ is the age of an ω-categorical structure.*
8. *$\mathrm{fm}(\Phi_{\mathfrak{G}})$ is finitely bounded homogenizable.*

Proof of Theorem 13. It is well-known that the questions whether $L(\mathfrak{G})$ is regular for a given context-free grammar $\mathfrak{G}$ is undecidable, see, e.g., Theorem 6.6.6 in [49]. Hence, all eight equivalent conditions in Lemma 17 are undecidable for $\mathfrak{G}$. ◀

5 Open Questions

We proved the EXPSPACE-hardness of the amalgamation meta-problem. However, our methods rely heavily on the following three facts. First, symbols of arity > 2 allow us to simulate a restricted form of Datalog computation within one-point amalgamation diagrams. Second, Boolean combinations of atoms enable succinct representations of the Datalog rules. And third, symbols of unbounded arity enable storing exponential amount of information on a constant number of variables. We do not know how to extend our hardness result beyond EXPSPACE. In particular, it does not seem to be possible to reduce from any of the standard undecidable problems, which can be done for the closely related joint embedding property [22, 20]. Intuitively, the reason is that every representation of a run of a Turing machine in a finitely bounded class requires some sort of a successor predicate (see, e.g., [31]), and the successor predicate is never definable in any ω-categorical structure [8].

Open question: Is the amalgamation meta-problem decidable in EXPSPACE?

References

1 Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of databases*, volume 8. Addison-Wesley, 1995. URL: `http://webdam.inria.fr/Alice/`.
2 Ove Ahlman. Homogenizable structures and model completeness. *Arch. Math. Log.*, 55(7-8):977–995, 2016. `doi:10.1007/s00153-016-0507-6`.
3 Reza Akhtar and Alistair H. Lachlan. On countable homogeneous 3-hypergraphs. *Arch. Math. Log.*, 34(5):331–344, 1995. `doi:10.1007/BF01387512`.

4 Albert Atserias and Szymon Torunczyk. Non-homogenizable classes of finite structures. In Jean-Marc Talbot and Laurent Regnier, editors, *25th EACSL Annual Conference on Computer Science Logic, CSL 2016, August 29 – September 1, 2016, Marseille, France*, volume 62 of *LIPIcs*, pages 16:1–16:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.CSL.2016.16.

5 Franz Baader and Jakub Rydval. Using model theory to find decidable and tractable description logics with concrete domains. *J. Autom. Reason.*, 66(3):357–407, 2022. doi:10.1007/s10817-022-09626-2.

6 Libor Barto, Michael Kompatscher, Miroslav Olšák, Trung Van Pham, and Michael Pinsker. The equivalence of two dichotomy conjectures for infinite domain constraint satisfaction problems. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12. IEEE Computer Society, 2017. available at arXiv:1612.07551. doi:10.1109/LICS.2017.8005128.

7 Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, csp, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014. doi:10.1145/2661643.

8 Manuel Bodirsky. *Complexity of Infinite-Domain Constraint Satisfaction*. Cambridge University Press, 2021. doi:10.1017/9781107337534.

9 Manuel Bodirsky and Bertalan Bodor. Canonical polymorphisms of ramsey structures and the unique interpolation property. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 – July 2, 2021*, pages 1–13. IEEE, IEEE, 2021. doi:10.1109/LICS52264.2021.9470683.

10 Manuel Bodirsky, David Bradley-Williams, Michael Pinsker, and András Pongrácz. The universal homogeneous binary tree. *J. Log. Comput.*, 28(1):133–163, 2018. doi:10.1093/logcom/exx043.

11 Manuel Bodirsky and Víctor Dalmau. Datalog and constraint satisfaction with infinite templates. *J. Comput. Syst. Sci.*, 79(1):79–100, 2013. A preliminary version appeared in the proceedings of the Symposium on Theoretical Aspects of Computer Science (STACS'05). doi:10.1016/j.jcss.2012.05.012.

12 Manuel Bodirsky and Martin Grohe. Non-dichotomies in constraint satisfaction complexity. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part II – Track B: Logic, Semantics, and Theory of Programming & Track C: Security and Cryptography Foundations*, volume 5126 of *Lecture Notes in Computer Science*, pages 184–196. Springer, July 2008. doi:10.1007/978-3-540-70583-3_16.

13 Manuel Bodirsky, Peter Jonsson, and Van Trung Pham. The complexity of phylogeny constraint satisfaction problems. *ACM Trans. Comput. Log.*, 18(3):23:1–23:42, 2017. doi:10.1145/3105907.

14 Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *J. ACM*, 57(2):9:1–9:41, 2010. doi:10.1145/1667053.1667058.

15 Manuel Bodirsky, Simon Knäuer, and Florian Starke. ASNP: A tame fragment of existential second-order logic. In Marcella Anselmo, Gianluca Della Vedova, Florin Manea, and Arno Pauly, editors, *Beyond the Horizon of Computability – 16th Conference on Computability in Europe, CiE 2020, Fisciano, Italy, June 29 – July 3, 2020, Proceedings*, volume 12098 of *Lecture Notes in Computer Science*, pages 149–162. Springer, Springer, 2020. doi:10.1007/978-3-030-51466-2_13.

16 Manuel Bodirsky, Florent R. Madelaine, and Antoine Mottet. A proof of the algebraic tractability conjecture for monotone monadic SNP. *SIAM J. Comput.*, 50(4):1359–1409, 2021. doi:10.1137/19M128466X.

17 Manuel Bodirsky and Antoine Mottet. Reducts of finitely bounded homogeneous structures, and lifting tractability from finite-domain constraint satisfaction. In Martin Grohe, Eric Koskinen, and Natarajan Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 623–632. ACM, 2016. available at arXiv:1601.04520. doi:10.1145/2933575.2934515.

18 Manuel Bodirsky, Michael Pinsker, and András Pongrácz. Projective clone homomorphisms. *J. Symb. Log.*, 86(1):148–161, 2021. doi:10.1017/jsl.2019.23.

19 Manuel Bodirsky and Jakub Rydval. On the descriptive complexity of temporal constraint satisfaction problems. *J. ACM*, 70(1):2:1–2:58, 2023. doi:10.1145/3566051.

20 Manuel Bodirsky, Jakub Rydval, and André Schrottenloher. Universal horn sentences and the joint embedding property. *Discret. Math. Theor. Comput. Sci.*, 23(2), 2021. doi:10.46298/dmtcs.7435.

21 Mikolaj Bojanczyk, Luc Segoufin, and Szymon Torunczyk. Verification of database-driven systems via amalgamation. In Richard Hull and Wenfei Fan, editors, *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA – June 22–27, 2013*, pages 63–74. ACM, 2013. doi:10.1145/2463664.2465228.

22 Samuel Braunfeld. The undecidability of joint embedding and joint homomorphism for hereditary graph classes. *Discret. Math. Theor. Comput. Sci.*, 21(2), 2019. available at arXiv:1903.11932. doi:10.23638/DMTCS-21-2-9.

23 Andrei A. Bulatov. A dichotomy theorem for nonuniform csps. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 319–330. IEEE, IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.37.

24 Jakub Bulin, Dejan Delic, Marcel Jackson, and Todd Niven. A finer reduction of constraint problems to digraphs. *Log. Methods Comput. Sci.*, 11(4), 2015. doi:10.2168/LMCS-11(4:18)2015.

25 Gregory L. Cherlin. Forbidden substructures and combinatorial dichotomies: WQO and universality. *Discret. Math.*, 311(15):1543–1584, 2011. doi:10.1016/j.disc.2011.03.014.

26 Gregory L. Cherlin. *Homogeneous ordered graphs, metrically homogeneous graphs, and beyond*, volume 1. Cambridge University Press, 2022. doi:10.1017/9781009229661.

27 Gregory L. Cherlin, Saharon Shelah, and Niandong Shi. Universal graphs with forbidden subgraphs and algebraic closure. *Advances in Applied Mathematics*, 22(4):454–491, 1999. doi:10.1006/aama.1998.0641.

28 Jacinta Covington. Homogenizable relational structures. *Illinois Journal of Mathematics*, 34(4):731–743, 1990. doi:10.1215/ijm/1255988065.

29 Péter L. Erdös, Claude Tardif, and Gábor Tardos. Caterpillar dualities and regular languages. *SIAM J. Discret. Math.*, 27(3):1287–1294, 2013. doi:10.1137/120879270.

30 Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998. doi:10.1137/S0097539794266766.

31 Haim Gaifman, Harry G. Mairson, Yehoshua Sagiv, and Moshe Y. Vardi. Undecidable optimization problems for database logic programs. *J. ACM*, 40(3):683–713, 1993. doi:10.1145/174130.174142.

32 Nicola Gigante, Andrea Micheli, Angelo Montanari, and Enrico Scala. Decidability and complexity of action-based temporal planning over dense time. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9859–9866. AAAI Press, 2020. doi:10.1609/aaai.v34i06.6539.

33 Wilfrid Hodges. *A shorter model theory.* Cambridge University Press, Cambridge, 1997.

34 Jan Hubička and Jaroslav Nešetřil. Homomorphism and embedding universal structures for restricted classes. *J. Multiple Valued Log. Soft Comput.*, 27(2-3):229–253, 2016. available at arXiv:0909.4939. URL: http://www.oldcitypublishing.com/journals/mvlsc-home/mvlsc-issue-contents/mvlsc-volume-27-number-2-3-2016/mvlsc-27-2-3-p-229-253/.

35 Jan Hubička and Jaroslav Nešetřil. All those ramsey classes (ramsey classes with closures and forbidden homomorphisms). *Advances in Mathematics*, 356:1–89, 2019. doi:10.1016/j.aim.2019.106791.

36 Julia F. Knight and Alistair H. Lachlan. Shrinking, stretching, and codes for homogeneous structures. In *Classification Theory: Proceedings of the US-Israel Workshop on Model Theory in Mathematical Logic held in Chicago, Dec. 15–19, 1985*, volume 1292, pages 192–229. Springer, Springer, Berlin, Heidelberg, 2006. doi:10.1007/BFb0082239.

37 Gábor Kun. Constraints, MMSNP and expander relational structures. *Comb.*, 33(3):335–347, 2013. doi:10.1007/s00493-013-2405-4.

38 Alistair H. Lachlan. Homogeneous structures. In *Proceedings of the International Congress of Mathematicians*, volume 1, pages 314–321, Berkeley, 1986. AMS.

39 Carsten Lutz and Maja Milicic. A tableau algorithm for description logics with concrete domains and general tboxes. *J. Autom. Reason.*, 38(1-3):227–259, 2007. doi:10.1007/s10817-006-9049-7.

40 Dugald Macpherson. A survey of homogeneous structures. *Discret. Math.*, 311(15):1599–1634, 2011. doi:10.1016/j.disc.2011.01.024.

41 Florent R. Madelaine. Universal structures and the logic of forbidden patterns. *Log. Methods Comput. Sci.*, 5(2), 2009. doi:10.2168/LMCS-5(2:13)2009.

42 Florent R. Madelaine and Iain A. Stewart. Some problems not definable using structure homomorphisms. *Ars Comb.*, 67:153–160, 2003.

43 Antoine Mottet. Promise and infinite-domain constraint satisfaction. In Aniello Murano and Alexandra Silva, editors, *32nd EACSL Annual Conference on Computer Science Logic, CSL 2024, February 19-23, 2024, Naples, Italy*, volume 288 of *LIPIcs*, pages 41:1–41:19. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.CSL.2024.41.

44 Antoine Mottet and Michael Pinsker. Smooth approximations and csps over finitely bounded homogeneous structures. In Christel Baier and Dana Fisman, editors, *LICS '22: 37th Annual ACM/IEEE Symposium on Logic in Computer Science, Haifa, Israel, August 2–5, 2022*, pages 36:1–36:13. ACM, 2022. available at arXiv:2011.03978. doi:10.1145/3531130.3533353.

45 Antoine Mottet, Michael Pinsker, and Tomáš Nagy. An order out of nowhere: a new algorithm for infinite-domain CSPs. In *51st EATCS International Colloquium on Automata, Languages and Programming (ICALP 2024)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. to appear, available at arXiv:2301.12977. doi:10.48550/arXiv.2301.12977.

46 Shan-Hwei Nienhuys-Cheng and Ronald de Wolf. *Foundations of Inductive Logic Programming*, volume 1228 of *Lecture Notes in Computer Science*. Springer, 1997. doi:10.1007/3-540-62927-0.

47 Jakub Rydval. Homogeneity and homogenizability: Hard problems for the logic snp. *CoRR*, abs/2108.00452, 2021. doi:10.48550/arXiv.2108.00452.

48 François Schwarzentruber. The complexity of tiling problems. *CoRR*, abs/1907.00102, 2019. doi:10.48550/arXiv.1907.00102.

49 Jeffrey O. Shallit. *A second course in formal languages and automata theory*. Cambridge University Press, 2008. URL: http://www.cambridge.org/gb/knowledge/isbn/item1173872/?site_locale=en_GB.

50 Peter van Emde Boas. The convenience of tilings. In *Complexity, Logic, and Recursion Theory*, pages 331–363. CRC Press, 2019. doi:10.1201/9780429187490.

51 Dmitriy Zhuk. A proof of CSP dichotomy conjecture. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 331–342. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.38.

52 Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM*, 67(5):30:1–30:78, 2020. doi:10.1145/3402029.