Hrushovski's Encoding and ω-Categorical CSP Monsters

Pierre Gillibert
Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Austria

Julius Jonušas
Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Austria

Michael Kompatscher
Department of Algebra, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Antoine Mottet
Department of Algebra, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
http://www.karlin.mff.cuni.cz/~mottet/

Michael Pinsker
Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Austria
Department of Algebra, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Abstract

We produce a class of ω-categorical structures with finite signature by applying a model-theoretic construction – a refinement of an encoding due to Hrushosvki – to ω-categorical structures in a possibly infinite signature. We show that the encoded structures retain desirable algebraic properties of the original structures, but that the constraint satisfaction problems (CSPs) associated with these structures can be badly behaved in terms of computational complexity. This method allows us to systematically generate ω-categorical templates whose CSPs are complete for a variety of complexity classes of arbitrarily high complexity, and ω-categorical templates that show that membership in any given complexity class cannot be expressed by a set of identities on the polymorphisms. It moreover enables us to prove that recent results about the relevance of topology on polymorphism clones of ω-categorical structures also apply for CSP templates, i.e., structures in a finite language. Finally, we obtain a concrete algebraic criterion which could constitute a description of the delineation between tractability and NP-hardness in the dichotomy conjecture for first-order reducts of finitely bounded homogeneous structures.

2012 ACM Subject Classification Mathematics of computing → Combinatoric problems

Keywords and phrases Constraint satisfaction problem, complexity, polymorphism, pointwise convergence topology, height 1 identity, ω-categoricity, orbit growth

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.131

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

Related Version A long version of this abstract is available on arXiv [22], https://arxiv.org/abs/2002.07054.

Funding *Pierre Gillibert*: received funding from the Austrian Science Fund (FWF) through projects No P27600 and P32337.
Julius Jonušas: received funding from the Austrian Science Fund (FWF) through Lise Meitner grant No M 2555.
Michael Kompatscher: supported by the grants PRIMUS/SCI/12 and UNCE/SCI/022 of Charles University Research Centre programs, as well as grant No 18-20123S of the Czech Science Foundation.
Antoine Mottet: received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 771005, CoCoSym).
Michael Pinsker: received funding from the Austrian Science Fund (FWF) through projects No P27600 and P32337 and from the Czech Science Foundation (grant No 18-20123S).

1 Introduction

1.1 Constraint Satisfaction Problems

The *Constraint Satisfaction Problem*, or CSP for short, over a relational structure $\mathbb{A}$ is the computational problem of deciding whether a given finite relational structure $\mathbb{B}$ in the signature of $\mathbb{A}$ can be homomorphically mapped into $\mathbb{A}$. The structure $\mathbb{A}$ is known as the *template* or *constraint language* of the CSP, and the CSP of the particular structure $\mathbb{A}$ is denoted by $\mathrm{CSP}(\mathbb{A})$. A host of interesting computational problems can be modelled using CSPs by choosing an appropriate template. For example, if $\mathbb{A}$ is the structure with domain $\{0,1\}$ and all binary relations on the set $\{0,1\}$, then $\mathrm{CSP}(\mathbb{A})$ is precisely the 2-SAT problem, and if $\mathbb{A}$ is the complete graph on three vertices, then $\mathrm{CSP}(\mathbb{A})$ is the 3-colouring problem. Note that the template $\mathbb{A}$ which defines the problem can also be infinite – only the input structure $\mathbb{B}$ is required to be finite in order to obtain a computational problem. Many well-known computational problems can be modelled, and can in fact only be modelled, using an infinite template. One example is the CSP of the order of the rational numbers $(\mathbb{Q};<)$, which is equivalent to the problem of deciding whether a given finite directed graph is acyclic. The size of the signature of the template $\mathbb{A}$, or in other words the number of its relations, is however generally required to be finite: otherwise, the encoding of its relational symbols might influence the computational complexity of $\mathrm{CSP}(\mathbb{A})$, so that this complexity is not well-defined as per the structure $\mathbb{A}$ itself. To emphasize the importance of this requirement, we shall henceforth only call relational structures in a finite signature *CSP templates*.

The general aim in the study of CSPs is to understand the structural reasons for the hardness or the tractability of such problems. This has been successfully achieved for CSPs of structures over a finite domain. As it turns out, every finite template either has, in a certain precise sense, as little symmetry as the 3-colouring problem above, in which case its CSP is NP-complete; or it has more symmetry and its CSP is polynomial-time solvable, just like the 2-SAT problem. This dichotomy result was conjectured by Feder and Vardi [19, 20], and proved, almost 25 years later, independently by Bulatov [18] and Zhuk [28].

1.2 A dichotomy conjecture and local identities

The algebraic approach behind these proofs does not require the template to be finite, but also works under the assumption of ω-categoricity. And although every computational decision problem is polynomial-time Turing-equivalent to the CSP of some infinite template [9], for a large and natural class of ω-categorical templates, which considerably expands the class of finite templates, a similar conjecture as for finite-domain CSPs has been formulated.

▶ **Conjecture 1** (see [3, 5, 17]). *Let $\mathbb{A}$ be a CSP template which is a first-order reduct of a countable finitely bounded homogeneous structure. Then one of the following holds.*

- *$\mathbb{A}$ satisfies some non-trivial set of h1 identities locally, i.e., on every finite subset of its domain, and $\mathrm{CSP}(\mathbb{A})$ is in P.*
- *There exists a finite subset of its domain on which $\mathbb{A}$ satisfies no non-trivial set of h1 identities, and $\mathrm{CSP}(\mathbb{A})$ is NP-complete.*

The conjectured P/NP-complete dichotomy has been demonstrated for numerous subclasses: for example for all CSPs in the class MMSNP [11], as well as for the CSPs of the first-order reducts of $(\mathbb{Q};<)$ [10], of any countable homogeneous graph [12] (including the random graph [15]), and of the random poset [24].

It is thus the *local h1 identities* which are believed to be the "right" measure of symmetry of a template $\mathbb{A}$ – according to the conjecture, they determine tractability or hardness of its CSP. The distinction between local and global h1 identities is, of course, void in the case of a finite template $\mathbb{A}$, and the above-stated dichotomy is true by the theorem of Bulatov and Zhuk and results from [5]. One of the main challenges towards proving Conjecture 1 is to determine whether this distinction could be void as well for structures within its range. The satisfaction of non-trivial h1 identities in $\mathbb{A}$ is characterised by the non-existence of particular maps called *minion homomorphisms* from the *polymorphism clone* $\mathrm{Pol}(\mathbb{A})$ of $\mathbb{A}$ to the *projection clone* $\mathscr{P}$. The mentioned distinction between local and global h1 identities is then mirrored by the distinction between those maps that are *uniformly continuous* and those that are not. Hence, the question whether non-trivial local h1 identities imply non-trivial global h1 identities in a relational structure $\mathbb{A}$ raises the following question.

▶ **Question 2.** *Does the existence of a minion homomorphism from* $\mathrm{Pol}(\mathbb{A})$ *to* $\mathscr{P}$ *imply the existence of a uniformly continuous minion homomorphism from* $\mathrm{Pol}(\mathbb{A})$ *to* $\mathscr{P}$*?*

2 Results

Among all finite structures, it is known that CSP templates (i.e., those structures with finite signature) have considerably better algebraic properties than other structures [2, 1]. We refine a model-theoretic trick due to Hrushovski [23] to encode ω-categorical structures with an infinite signature into ω-categorical CSP templates whilst preserving certain properties of the original, showing that a similar phenomenon does not seem to appear within the class of ω-categorical structures. Using this method, we produce ω-categorical CSP templates with various "untame" properties of both algebraic and complexity-theoretic nature.

2.1 Local versus global identities

Recently, in [13, 14], an example of an ω-categorical structure answering Question 2 in the negative was given; however, this structure had an infinite language and therefore this result had *a priori* no consequence for the study of CSPs. Using our encoding, we provide a negative answer within the realm of CSP templates.

▶ **Theorem 3.** *There is an ω-categorical CSP template $\mathbb{U}$ with slow orbit growth such that there exists a minion homomorphism from* $\mathrm{Pol}(\mathbb{U})$ *to* $\mathscr{P}$*, but no uniformly continuous one.*

We also encode a counterexample from [17] for *clone homomorphisms*, which are mappings preserving arbitrary (not only h1) identities, into a finite language. Clone homomorphisms appear in the original (and equivalent [3, 4]) formulation of Conjecture 1 from [17] (also see [6, 7]).

▶ **Theorem 4.** *There exists an ω-categorical CSP template $\mathbb{U}$ with a clone homomorphism from* $\mathrm{Pol}(\mathbb{U})$ *to* $\mathscr{P}$ *that is not uniformly continuous.*

2.2 Dissected weak near-unanimity identities

Our proof of the fact that the template $\mathbb{U}$ from Theorem 3 satisfies non-trivial h1 identities locally is constructive: we exhibit a concrete set of such identities which we call *dissected weak near-unanimity*. Moreover, we obtain quite general conditions on the symmetry of a structure which force our identities to be satisfied locally. It follows that the original infinite-language structure from [13, 14] satisfies them; this contrasts the indirect proof in [13, 14] which does not provide any concrete set of h1 identities.

▶ **Theorem 5.** *Let $\mathbb{U}$ be a homogeneous structure. Let F be a finite subset of its domain, and let $k \geq 1$. Suppose that:*

(i) *Only relations of arity smaller than k hold for tuples of elements in F.*

(ii) *There is an embedding from $\mathbb{U}^2$ into $\mathbb{U}$.*

Then $\mathbb{U}$ satisfies (n, k) dissected weak near-unanimity identities on F for all $n > k$.

Dissected weak near-unanimity identities can be viewed as a generalization of *weak near-unanimity* identities. It follows from [26] and [5] that if $\mathbb{U}$ is a finite relational structure satisfying non-trivial h1 identities, then $\mathbb{U}$ satisfies weak near-unanimity identities. The satisfaction of dissected weak near-unanimity identities has been proven for a large number of structures within the range of Conjecture 1 in [4, 3]; since they now reappeared in the rather different context of Theorem 3, the following question emerges naturally.

▶ **Question 6.** *Let $\mathbb{U}$ be an ω-categorical structure with slow orbit growth which satisfies non-trivial h1 identities locally. Does $\mathbb{U}$ satisfy dissected weak near-unanimity identities locally?*

2.3 ω-categorical CSP monsters

The complexity of CSP($\mathbb{A}$) is, for every ω-categorical CSP template $\mathbb{A}$, determined by Pol($\mathbb{A}$) viewed as a topological clone: if there exists a topological clone isomorphism Pol($\mathbb{A}$) $\rightarrow$ Pol($\mathbb{B}$) and $\mathbb{A}$ and $\mathbb{B}$ are ω-categorical, then CSP($\mathbb{A}$) and CSP($\mathbb{B}$) are equivalent under log-space reductions [16]. In other words, the local (not necessarily h1) identities satisfied in Pol($\mathbb{A}$) encode the complexity of CSP($\mathbb{A}$). Conjecture 1 even postulates that for every template $\mathbb{A}$ within its scope, membership of CSP($\mathbb{A}$) in P only depends on the local h1 identities of $\mathbb{A}$. The latter is equivalent to the statement that polynomial-time tractability is characterised by the global satisfaction of the single identity $\alpha s(x, y, x, z, y, z) = \beta s(y, x, z, x, z, y)$ [4, 7].

Using our encoding, we prove that global identities do not characterise membership in P – or, in fact, in any other non-trivial class of languages containing AC^0 – for the class of homogeneous CSP templates.

▶ **Theorem 7.** *Let $\mathcal{C}$ be any class of languages that contains AC^0 and that does not intersect every Turing degree. Then there is no countable set Θ of identities such that for all homogeneous CSP templates membership in $\mathcal{C}$ is equivalent to the satisfaction of Θ.*

The proof of Theorem 7 relies on encoding arbitrary languages as CSPs of homogeneous templates. These templates are obtained by applying our encoding to structures which have only empty relations, but a complicated infinite signature. On the way, we obtain a new proof of a result by Bodirsky and Grohe [9].

▶ **Theorem 8.** *Let $\mathcal{C}$ be a complexity class such that there exist* co*NP*$^{\mathcal{C}}$*-complete problems. Then there exists a homogeneous CSP template that satisfies non-trivial h1 identities and whose CSP is* co*NP*$^{\mathcal{C}}$*-complete. Moreover, if P $\neq$* co*NP, then there exists a CSP template with these algebraic properties whose CSP has* co*NP-intermediate complexity.*

In particular, Theorem 8 gives complete problems for classes such as Π_n^{P} for every $n \geq 1$, Pspace, ExpTime, or even every fast-growing time complexity class $\mathbf{F}_\alpha$ where $\alpha \geq 2$ is an ordinal (such as the classes Tower or Ackermann, see [27]).

3 Outline

The most important definitions, in particular those of notions which appear in the introduction and the results, are given in Section 4. A complete exposition of all notions, as well as most proofs are left to the appendix due to space restrictions; we also refer to the long version of this extended abstract which is available on arXiv [22]. Our variant of Hrushovski's encoding and its most important properties are presented in Section 5. Dissected weak near-unanimity identities, as well as the proof of Theorem 3, are the contents of Section 6. In Section 7, we sketch the proofs of Theorem 7 and of the first statement of Theorem 8. Due to space restrictions, we cannot address the proofs of Theorems 4 and of the second statement of Theorem 8 at all – they can be found in the long version.

4 Preliminaries

If $\mathbb{A}$ is a relational structure in a finite signature, called a *CSP template*, then $\mathrm{CSP}(\mathbb{A})$ is the set of all finite structures $\mathbb{B}$ in the same signature with the property that there exists a homomorphism from $\mathbb{B}$ into $\mathbb{A}$. This set can be viewed as a computational problem where we are given a finite structure $\mathbb{B}$ in that signature, and we have to decide whether $\mathbb{B} \in \mathrm{CSP}(\mathbb{A})$.

We will tacitly assume that all relational structures, as well as their signatures, are at most countably infinite.

4.1 Polymorphisms, identities, and clone and minion homomorphisms

A *polymorphism* of a relational structure $\mathbb{A}$ is a homomorphism from some finite power $\mathbb{A}^n$ of the structure into $\mathbb{A}$. The set of all polymorphisms of $\mathbb{A}$ is called the *polymorphism clone* of $\mathbb{A}$ and is denoted by $\mathrm{Pol}(\mathbb{A})$.

An *identity* is a formal expression $s(x_1, \ldots, x_n) = t(y_1, \ldots, y_m)$ where s and t are abstract terms of function symbols, and $x_1, \ldots, x_n, y_1, \ldots, y_m$ are the variables that appear in these terms. The identity is of *height 1*, and called *h1 identity*, if the terms s and t contain precisely one function symbol, i.e., no nesting of function symbols is allowed, and no term may be just a variable.

A set Θ of identities is *satisfied* in $\mathbb{A}$ if the function symbols of Θ can be assigned functions in $\mathrm{Pol}(\mathbb{A})$ in such a way that all identities of Θ become true for all possible values of their variables in $\mathbb{A}$. If F is a finite subset of the domain of $\mathbb{A}$, then Θ *is satisfied locally on* F if the above situation holds where only values within F are considered for the variables.

A set of identities is called *trivial* if it is satisfied in the *projection clone* $\mathscr{P}$ consisting of the projection operations on the set $\{0, 1\}$. Otherwise, the set is called *non-trivial*. We say that $\mathbb{A}$ satisfies non-trivial identities *locally* if on every finite subset of its domain it locally satisfies some non-trivial set of identities. We shall use similar terminology for h1 identities.

A map $\xi\colon \mathrm{Pol}(\mathbb{A}) \to \mathrm{Pol}(\mathbb{B})$ is called a *clone homomorphism* if it preserves arities, maps the i-th n-ary projection in $\mathrm{Pol}(\mathbb{A})$ to the i-th n-ary projection in $\mathrm{Pol}(\mathbb{B})$ for all $1 \leq i \leq n$, and satisfies $\xi(f \circ (g_1, \ldots, g_n)) = \xi(f) \circ (\xi(g_1), \ldots, \xi(g_n))$ for all $n, m \geq 1$, all n-ary $f \in \mathrm{Pol}(\mathbb{A})$, and all m-ary $g_1, \ldots, g_n \in \mathrm{Pol}(\mathbb{A})$. This is the case if and only if the map ξ *preserves identities*, i.e., whenever some functions in $\mathrm{Pol}(\mathbb{A})$ witness the satisfaction of some identity in $\mathrm{Pol}(\mathbb{A})$, then their images under ξ witness the satisfaction of the same identity in $\mathrm{Pol}(\mathbb{B})$.

A map $\xi\colon \mathrm{Pol}(\mathbb{A}) \to \mathrm{Pol}(\mathbb{B})$ is called a *minion homomorphism* if it preserves arities and composition with projections; the latter meaning that for all $n, m \geq 1$, all n-ary $f \in \mathrm{Pol}(\mathbb{A})$, and all m-ary projections $p_1, \ldots, p_n \in \mathrm{Pol}(\mathbb{A})$, we have $\xi(f \circ (p_1, \ldots, p_n)) = \xi(f) \circ (p'_1, \ldots, p'_n)$, where p'_i is the m-ary projection in $\mathrm{Pol}(\mathbb{B})$ onto the same variable as p_i, for all $1 \leq i \leq n$. This is the case if and only if the map ξ preserves h1 identities in the sense above.

The existence of clone and minion homomorphisms $\mathrm{Pol}(\mathbb{A}) \to \mathscr{P}$ is connected to the satisfaction of non-trivial identities in a relational structure $\mathbb{A}$. Namely, there exists a clone homomorphism $\mathrm{Pol}(\mathbb{A}) \to \mathscr{P}$ if and only if every set of identities satisfied in $\mathbb{A}$ is trivial; and there exists a minion homomorphism $\mathrm{Pol}(\mathbb{A}) \to \mathscr{P}$ if and only if every set of h1 identities satisfied in $\mathbb{A}$ is trivial.

Similarly, the *local satisfaction* of identities and h1 identities can be characterised via *uniformly continuous* clone and minion homomorphisms, respectively [16, 21, 5]. However, the reader will not need any knowledge of topology and can only keep in mind that the topology reflects the local/global distinction.

4.2 Homogeneity, boundedness, reducts, ω-categoricity, orbit growth

Let $\mathcal{C}$ be a class of finite structures in a common relational signature which is closed under isomorphisms. We define the following properties the class $\mathcal{C}$ might have.

Hereditary property (HP): if $\mathbb{A} \in \mathcal{C}$ and if $\mathbb{B}$ is a substructure of $\mathbb{A}$, then $\mathbb{B} \in \mathcal{C}$;

Amalgamation property (AP): if $\mathbb{A}, \mathbb{B}, \mathbb{C} \in \mathcal{C}$ and if $f_1 \colon \mathbb{A} \to \mathbb{B}$ and $f_2 \colon \mathbb{A} \to \mathbb{C}$ are embeddings, then there exist $\mathbb{D} \in \mathcal{C}$ and embeddings $g_1 \colon \mathbb{B} \to \mathbb{D}$ and $g_2 \colon \mathbb{C} \to \mathbb{D}$ such that $g_1 \circ f_1 = g_2 \circ f_2$;

Strong amalgamation property (SAP): $\mathcal{C}$ satisfies AP and in addition g_1 and g_2 can be chosen to have disjoint ranges, except for the common values enforced by above equation.

A relational structure $\mathbb{C}$ is *homogeneous* if every isomorphism between finite induced substructures extends to an automorphism of the entire structure $\mathbb{C}$. In that case, $\mathbb{C}$ is uniquely determined, up to isomorphism, by its *age*, i.e., the class of its finite induced substructures up to isomorphism. This is a consequence of the following theorem.

▶ **Theorem 9** (Fraïssé's Theorem, see [23]). *Let σ be a relational signature and let $\mathcal{C}$ be a class of finite σ-structures which is closed under isomorphisms and satisfies HP and AP. Then there exists a σ-structure $\mathbb{A}$ such that $\mathbb{A}$ is countable, homogeneous, and the age of $\mathbb{A}$ equals $\mathcal{C}$. Furthermore $\mathbb{A}$ is unique up to isomorphism.*

The structure $\mathbb{A}$ above is called the *Fraïssé limit* of $\mathcal{C}$, and the class $\mathcal{C}$ a *Fraïssé class*.

A class $\mathcal{C}$ of finite structures in the same finite signature is *finitely bounded* if it is given by a finite set $\mathcal{F}$ of forbidden finite substructures, i.e., $\mathcal{C}$ consists precisely of those finite structures in its signature which do not embed any member of $\mathcal{F}$. A class $\mathcal{C}$ of finite structures in the same signature is *homomorphically bounded* by a (possibly infinite) set $\mathcal{F}$ of finite structures if it is defined by forbidding the structures in $\mathcal{F}$ homomorphically, i.e., $\mathcal{C}$ consists precisely of those finite structures in its signature which do not contain a homomorphic image of any member of $\mathcal{F}$ as a substructure. A structure $\mathbb{A}$ is finitely bounded (homomorphically bounded) if its age is.

A *first-order reduct* of a relational structure $\mathbb{C}$ is a relational structure $\mathbb{A}$ on the same domain which is first-order definable without parameters in $\mathbb{C}$. Every first-order reduct $\mathbb{A}$ of a finitely bounded homogeneous structure is *ω-categorical*, i.e., the automorphism group $\mathrm{Aut}(\mathbb{A})$ has finitely many orbits in its componentwise action on A^n, for all finite $n \geq 1$. In fact, if $\mathbb{A}$ is such a first-order reduct, then the number of orbits in the action of $\mathrm{Aut}(\mathbb{A})$ on A^n grows exponentially in n; in general, we say that ω-categorical structures where this number grows less than double exponentially in n have *slow orbit growth*.

For a relational structure $\mathbb{A}$ in signature $\sigma = (R_i)_{i \in I}$, and $J \subseteq I$, we call the structure $(A; (R_i^{\mathbb{A}})_{i \in J})$ in signature $\rho := (R_i)_{i \in J}$ the *ρ-reduct* of $\mathbb{A}$; conversely $\mathbb{A}$ is called an *expansion* of any of its reducts, and a *first-order expansion* of a reduct if all of its relations have a first-order definition in the reduct. We say that a structure is *homogenizable* if it has a homogeneous first-order expansion. All ω-categorical structures are homogenizable. A homogenizable structure $\mathbb{A}$ has *no algebraicity* if the age of any, or equivalently some, homogeneous first-order expansion of $\mathbb{A}$ has SAP.

A formula is *primitive positive*, in short *pp*, if it contains only existential quantifiers, conjunctions, equalities, and relational symbols. If $\mathbb{A}$ is a relational structure, then a relation is *pp-definable* in $\mathbb{A}$ if it can be defined by a pp-formula in $\mathbb{A}$. A structure $\mathbb{A}$ *pp-interprets* $\mathbb{B}$ if a structure isomorphic to $\mathbb{B}$ can be constructed from $\mathbb{A}$ by pp-defining a subset S of some finite power of its domain, then pp-defining an equivalence relation $\sim$ on S, and then pp-defining relations on the equivalence classes of $\sim$.

5 The encoding

We present the encoding of an arbitrary homogenizable structure with no algebraicity into a CSP template, which will be the basis of our results. The construction is originally due to Hrushovski [23, Section 7.4]; it was designed to capture properties of the first-order theory and consequently the automorphism group of the original structure. We refine his construction in order to also compare the polymorphism clones of the original structure and its encoded counterpart, and to control the complexity of the CSPs of the produced templates.

5.1 Encoding and Decoding

Let Σ be a finite alphabet, and let $\Sigma^{\geq 2}$ denote the set of all finite words over Σ of length at least two. We are going to encode structures with a signature of the form $\rho = (R_w)_{w\in W}$, where $W \subseteq \Sigma^{\geq 2}$ and where the arity of each symbol R_w equals the length $|w|$ of the word w. For the rest of this section we fix Σ and ρ. Our goal is to encode any homogenizable ρ-structure $\mathbb{A}$ with no algebraicity into a structure $\mathbf{E}\,\mathbb{A}$ (where $\mathbf{E}$ stands for *E. Hrushovski*) in a finite signature θ which is disjoint from ρ and only depends on Σ.

Note that by renaming its signature, and possibly artificially inflating the arity of its relations (by adding dummy variables), any arbitrary structure with countably many relations can be given a signature of the above form without changing, for example, its polymorphism clone. However, the encoding will depend on these modifications, and their effect on the algebraic and combinatorial properties of the encoding is beyond the scope of this article. The original encoding [23, Section 7.4] roughly corresponds to the case where $|\Sigma| = 1$, and our generalization allows us to avoid such modifications for the structures we wish to encode, making in particular our complexity-theoretic results possible.

▶ **Definition 10.** *Let θ denote the signature $\{P, \iota, \tau, S\} \cup \{H_s \mid s \in \Sigma\}$, where P, ι, τ are unary relation symbols, H_s is a binary relation symbol for each $s \in \Sigma$, and S is a 4-ary relation symbol. For every signature σ disjoint from θ, define σ^+ to be the union $\sigma \cup \theta$.*

The encoding of a ρ-structure $\mathbb{A}$ will roughly be obtained as follows: first, one takes a homogeneous first-order expansion $\mathbb{B}$ in some signature σ; from its age K, one defines a class K^+ of finite structures in signature σ^+; and the encoding is the θ-reduct of the Fraïssé limit of K^+. In order to define the class K^+, we need the following definitions.

▶ **Definition 11.** *Let σ be a signature disjoint from θ, let $\mathbb{A}$ be a σ^+-structure, and let $w \in \Sigma^{\geq 2}$. A tuple $(a_1, \ldots, a_{|w|}, c_1, \ldots, c_{|w|})$ of elements of $\mathbb{A}$ is a* valid w-code *in $\mathbb{A}$ if the following hold:*

(a) $a_1, \ldots, a_{|w|} \in P^{\mathbb{A}}$;

(b) $H^{\mathbb{A}}_{w_i}(c_i, c_j)$ *for all* $1 \leq i, j \leq |w|$ *such that* $j \equiv i + 1 \pmod{|w|}$;

(c) $\iota^{\mathbb{A}}(c_1)$ *and* $\tau^{\mathbb{A}}(c_{|w|})$;

(d) $S^{\mathbb{A}}(a_i, a_j, c_i, c_j)$ *for all* $1 \leq i, j \leq |w|$ *with* $i \neq j$.

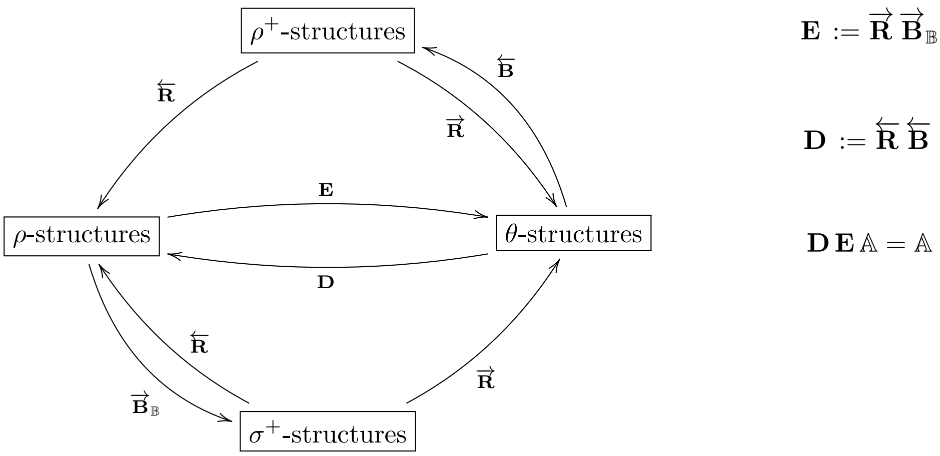

Figure 1 Sources and destinations of the operators related to the encoding and decoding.

▶ **Definition 12.** *Let σ be a signature disjoint from θ, and let $\mathbb{A}$ be a σ^+-structure. Then $\mathbb{A}$ is called* separated *if*

(i) *$H_s^{\mathbb{A}}$ only relates pairs within $A \setminus P^{\mathbb{A}}$ for all $s \in \Sigma$;*
(ii) *$\iota^{\mathbb{A}}, \tau^{\mathbb{A}}$ are contained in $A \setminus P^{\mathbb{A}}$;*
(iii) *If $(a, b, c, d) \in S^{\mathbb{A}}$, then $c, d \in A \setminus P^{\mathbb{A}}$ and $c \neq d$.*

It follows from (iii) above that in a separated structure a valid w-code can only exist if $|w| \geq 2$; this is the reason for the exclusion of unary relation symbols from ρ.

▶ **Definition 13.** *Let $\mathbb{A}$ be a ρ-structure and let $\mathbb{B}$ be a homogeneous first-order expansion of $\mathbb{A}$ with signature σ and age K. Define K^+ to be the class of all finite σ^+-structures $\mathbb{C}$ with the following properties:*

(1) *The σ-reduct of the restriction of $\mathbb{C}$ to $P^{\mathbb{C}}$ is an element of K;*
(2) *$\mathbb{C}$ is separated and for every $R \in \sigma$ the relation $R^{\mathbb{C}}$ only relates tuples which lie entirely within $P^{\mathbb{C}}$;*
(3) *If $R_w \in \rho$ and $(a_1, \dots, a_{|w|}, c_1, \dots, c_{|w|})$ is a valid w-code in $\mathbb{C}$, then $(a_1, \dots, a_{|w|}) \in R_w^{\mathbb{C}}$.*

▶ **Lemma 14.** *Let $\mathbb{A}$ be a ρ-structure and let $\mathbb{B}$ be a homogeneous first-order expansion of $\mathbb{A}$ with age K. If K has the HP and the SAP, then K^+ has the HP and the SAP as well.*

By Lemma 14, if $\mathbb{A}$ has no algebraicity[1], and $\mathbb{B}$ is a homogeneous first-order expansion of $\mathbb{A}$ with age K, then K^+ has a Fraïssé limit, allowing us to define our encoding as follows.

▶ **Definition 15.** *Let $\mathbb{A}$ be a ρ-structure with no algebraicity and let $\mathbb{B}$ be a homogeneous first-order expansion of $\mathbb{A}$ with age K. We define $\overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A}$, the* encoding blow up *of $\mathbb{A}$, to be the Fraïssé limit of K^+. Moreover, we define $\overrightarrow{\mathbf{R}}\,\mathbb{C}$ to be the θ-reduct of any structure $\mathbb{C}$ with signature containing θ. The* Hrushovski-encoding *$\mathbf{E}\,\mathbb{A}$ is defined by $\mathbf{E}\,\mathbb{A} := \overrightarrow{\mathbf{R}}\,\overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A}$.*

▶ Remark 16. All ω-categorical structures have a homogeneous first-order expansion. This expansion is not unique, but the encoding $\mathbf{E}\,\mathbb{A}$ of $\mathbb{A}$ does not depend on it.

[1] Contrary to a claim in [23], AP of K is not a sufficient assumption for AP of K^+ in Lemma 14.

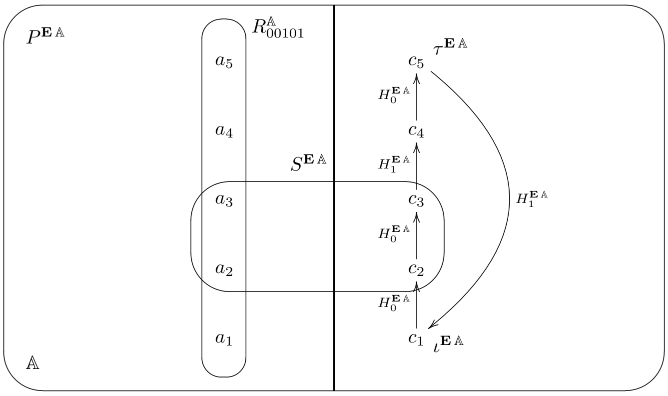

Figure 2 The encoding $\mathbf{E}\,\mathbb{A}$ of a structure $\mathbb{A}$.

It might be of help to the reader if we note that the operators used in the encoding of a structure, i.e., $\overrightarrow{\mathbf{B}}_{\mathbb{B}}$ and $\overrightarrow{\mathbf{R}}$, bear arrows from left to right; the operators used in the decoding of a structure, defined below, bear arrows in the opposite direction. Table 1 contains an informal summary of all operators, and Figure 1 describes on which classes of structures they operate.

Table 1 The meaning of the operators.

Operator	Name	Description
$\overrightarrow{\mathbf{B}}_{\mathbb{B}}$	encoding blow up	The first step in an encoding, extends the domain and defines relations for the signature θ via a homogeneous expansion $\mathbb{B}$ of the input.
$\overrightarrow{\mathbf{R}}$	θ-reduct	Returns the θ-reduct of a structure.
$\mathbf{E}$	encoding	Combines $\overrightarrow{\mathbf{B}}_{\mathbb{B}}$ and $\overrightarrow{\mathbf{R}}$ to obtain a θ-structure from a ρ-structure.
$\overleftarrow{\mathbf{B}}$	decoding blow up	The first step in decoding a θ-structure, it converts valid codes into corresponding relations in ρ.
$\overleftarrow{\mathbf{R}}$	relativised reduct	Restricts a structure to the set named by P and forgets relations not in ρ.
$\mathbf{D}$	decoding	Combines $\overleftarrow{\mathbf{R}}$ and $\overleftarrow{\mathbf{B}}$ to obtain the ρ-structure $\mathbb{A}$ from the encoded θ-structure $\mathbf{E}\,\mathbb{A}$.
$\mathbf{C}$	canonical code	Defines in a canonical way a finite θ-structure from a finite ρ-structure in which every relation which holds in the input is witnessed by a valid code.

Like the encoding of a structure, the decoding of a structure is a composition of two steps; first a *decoding blow up*, and then a *relativised reduct.*

▶ **Definition 17.** *Let $\mathbb{C}$ be a θ-structure. Then the* decoding blow up $\overleftarrow{\mathbf{B}}\,\mathbb{C}$ *of $\mathbb{C}$ is the expansion of $\mathbb{C}$ in signature ρ^+, where for any symbol $R_w \in \rho$ the relation $R_w^{\overleftarrow{\mathbf{B}}\,\mathbb{C}}$ is defined to consist of those tuples $(a_1, \ldots, a_{|w|})$ for which there exist $c_1, \ldots, c_{|w|}$ such that the tuple $(a_1, \ldots, a_{|w|}, c_1, \ldots, c_{|w|})$ is a valid w-code in $\mathbb{C}$.*

For a structure $\mathbb{D}$ in a signature containing ρ^+, the relativised reduct $\overleftarrow{\mathbf{R}}\,\mathbb{D}$ *of $\mathbb{D}$ is defined to be the ρ-reduct of $\mathbb{D}$ restricted to $P^{\mathbb{D}}$.*

Finally, we set $\mathbf{D}\,\mathbb{C} := \overleftarrow{\mathbf{R}}\,\overleftarrow{\mathbf{B}}\,\mathbb{C}$, the decoding of $\mathbb{C}$*, for any θ-structure $\mathbb{C}$.*

The following proposition states that the operator $\mathbf{D}$ indeed decodes $\mathbf{E}\,\mathbb{A}$. It also allows us to identify $\mathbb{A}$ and $\mathbf{D}\,\mathbf{E}\,\mathbb{A}$, an assumption we shall thenceforth make.

▶ **Proposition 18.** *Let $\mathbb{A}$ be a homogenizable ρ-structure with no algebraicity. Then $\mathbb{A}$ and $\mathbf{D}\,\mathbf{E}\,\mathbb{A}$ are isomorphic. Moreover, for any θ-structure $\mathbb{D}$, the structure $\mathbf{D}\,\mathbb{D}$ has a pp-interpretation in $\mathbb{D}$.*

5.2 The relationship between $\mathbb{A}$ and $\mathbf{E}\,\mathbb{A}$

We derive the following main properties of the encoding $\mathbf{E}\,\mathbb{A}$ of a ρ-structure $\mathbb{A}$:

- $\mathbf{E}\,\mathbb{A}$ is ω-categorical if and only if $\mathbb{A}$ is, and has slow orbit growth if and only if $\mathbb{A}$ does (Proposition 19);
- There exists a uniformly continuous clone homomorphism ξ from $\mathrm{Pol}(\mathbf{E}\,\mathbb{A})$ into $\mathrm{Pol}(\mathbb{A})$; conversely, if $\mathbb{A}$ is ω-categorical, then injective polymorphisms of $\mathbb{A}$ essentially extend to polymorphisms of $\mathbf{E}\,\mathbb{A}$ (Proposition 20).

We start by investigating the relationship of the orbits of $\mathrm{Aut}(\mathbb{A})$ with those of $\mathrm{Aut}(\mathbf{E}\,\mathbb{A})$, showing that ω-categoricity and slow orbit growth are preserved by the encoding.

▶ **Proposition 19.** *Let $\mathbb{A}$ be a homogenizable ρ-structure with no algebraicity.*

1. *$\mathbb{A}$ is ω-categorical if and only if $\mathbf{E}\,\mathbb{A}$ is.*
2. *Let $\mathbb{A}$ be ω-categorical. For $n \geq 1$, write $f(n)$ and $g(n)$ for the number of orbits of n-tuples under the action of $\mathrm{Aut}(\mathbb{A})$ and $\mathrm{Aut}(\mathbf{E}\,\mathbb{A})$, respectively. Then $f(n) \leq g(n) \leq 2^{6|\Sigma|n^4} f(n)$ for all $n \geq 1$. In particular, $\mathbb{A}$ has slow orbit growth if and only if $\mathbf{E}\,\mathbb{A}$ does.*

We now turn to the polymorphism clones of $\mathbb{A}$ and $\mathbf{E}\,\mathbb{A}$. An immediate consequence of Proposition 18 is that polymorphisms of $\mathbf{E}\,\mathbb{A}$ can be restricted to polymorphisms of $\mathbb{A}$. Conversely, one can prove that assuming ω-categoricity of $\mathbb{A}$, for every injective $f \in \mathrm{Pol}(\mathbb{A})$ there exists an embedding u of $\mathbb{A}$ such that uf can be extended to a polymorphism of $\mathbf{E}\,\mathbb{A}$.

▶ **Proposition 20.** *Let $\mathbb{A}$ be a structure with no algebraicity, and let $\mathbb{B}$ be a homogeneous first-order expansion of $\mathbb{A}$. Then the following hold:*

(1) *For every $f \in \mathrm{Pol}(\mathbf{E}\,\mathbb{A})$, the restriction $f|_{P^{\mathbf{E}\,\mathbb{A}}}$ of f to $P^{\mathbf{E}\,\mathbb{A}}$ is a polymorphism of $\mathbb{A}$. The map $f \mapsto f|_{P^{\mathbf{E}\,\mathbb{A}}}$ is a uniformly continuous clone homomorphism $\mathrm{Pol}(\mathbf{E}\,\mathbb{A}) \to \mathrm{Pol}(\mathbb{A})$.*

(2) *If $\mathbb{A}$ is ω-categorical, then for every injective $f \in \mathrm{Pol}(\mathbb{A})$ there exists an embedding $u\colon \overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A} \to \overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A}$ such that uf extends to a polymorphism of $\mathbf{E}\,\mathbb{A}$.*

(3) *If $\mathbb{A}$ is ω-categorical, then for all $k \geq 1$, $\mathbb{B}^k$ embeds into $\mathbb{B}$ if and only if $(\overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A})^k$ embeds into $\overrightarrow{\mathbf{B}}_{\mathbb{B}}\,\mathbb{A}$.*

Propositions 19 and 20 are the fundamental results upon which the following sections rely. They allow us to relate $\mathrm{Pol}(\mathbf{E}\,\mathbb{A})$ and $\mathrm{Pol}(\mathbb{A})$ and to transfer the exotic behaviour of the latter (for a well-chosen $\mathbb{A}$ with infinite signature) into the former.

5.3 Homomorphisms and the encoding

We now examine the relationship between the finite structures that homomorphically map into a structure $\mathbb{A}$ with those that homomorphically map into its encoding $\mathbf{E}\,\mathbb{A}$ (which is precisely $\mathrm{CSP}(\mathbf{E}\,\mathbb{A})$). This will be particularly relevant in Section 7 where we investigate the complexity of CSPs of structures encoded with our encoding.

▶ **Proposition 21.** *There exists a log-space computable function $\mathbb{C} \mapsto \mathbf{C}\,\mathbb{C}$ from the set of finite ρ-structures to the set of finite θ-structures satisfying the following properties:*

- *For every finite ρ-structure $\mathbb{C}$, we have $\mathbf{D}\,\mathbf{C}\,\mathbb{C} = \mathbb{C}$.*
- *If $\mathbb{C}$ is a finite ρ-structure, and $\mathbb{D}$ is a θ-structure, then there exists a homomorphism from $\mathbb{C}$ to $\mathbf{D}\,\mathbb{D}$ if and only if there exists a homomorphism from $\mathbf{C}\,\mathbb{C}$ to $\mathbb{D}$.*

The properties from Proposition 21 are enough to give a concrete description of CSP($\mathbf{E}\,\mathbb{A}$) when $\mathbb{A}$ is homomorphically bounded.

▶ **Proposition 22.** *Let $\mathbb{A}$ be a homogenizable ρ-structure with no algebraicity which is homomorphically bounded by a set $\mathcal{G}$ of finite ρ-structures. Let $\mathbb{X}$ be a θ-structure. Then the following are equivalent.*

(1) *There exists an embedding of $\mathbb{X}$ into $\mathbf{E}\,\mathbb{A}$;*
(2) *There exists a homomorphism from $\mathbb{X}$ to $\mathbf{E}\,\mathbb{A}$;*
(3) *$\mathbb{X}$ is separated and for all $\mathbb{G} \in \mathcal{G}$ there exists no homomorphism from $\mathbf{C}\,\mathbb{G}$ to $\mathbb{X}$.*

Note that being separated can be characterised by not containing the homomorphic image of any element of a finite set $\mathcal{S}$ of finite θ-structures. As an immediate consequence of Proposition 22 we therefore obtain the following corollary.

▶ **Corollary 23.** *Let $\mathbb{A}$ be a homogenizable ρ-structure with no algebraicity which is homomorphically bounded by a set $\mathcal{G}$ of finite ρ-structures. Then $\mathbf{E}\,\mathbb{A}$ is homomorphically bounded by $\{\mathbf{C}\,\mathbb{G} \mid \mathbb{G} \in \mathcal{G}\} \cup \mathcal{S}$.*

6 Height 1 identities: local without global

Recall that in [13] a negative answer to Question 2 of the introduction was established by an infinite-language example which is ω-categorical and has slow orbit growth. We are now going to prove that the encoding of that structure, or in fact, of a simplification $\mathbb{S}$ thereof, also provides an example. Since $\mathbf{E}\,\mathbb{S}$ is a CSP template, and since both ω-categoricity and slow orbit growth are preserved by the encoding, $\mathbf{E}\,\mathbb{S}$ is a witness for the truth of Theorem 3. While the non-satisfaction of non-trivial global h1 identities lifts from $\mathbb{S}$ to $\mathbf{E}\,\mathbb{S}$ by virtue of Proposition 20 (1), we do not know in general when this is the case for the local satisfaction of non-trivial h1 identities. Our proof thus relies on specific structural properties of $\mathbb{S}$; we show that both $\mathbb{S}$ and $\mathbf{E}\,\mathbb{S}$ locally satisfy *dissected weak near-unanimity identities*.

6.1 Dissected weak near-unanimity identities

▶ **Definition 24.** *Let $n > k > 1$, let $g_1, \dots, g_n$ be binary function symbols, and for every injective function $\psi \colon \{1, \dots, k\} \to \{1, \dots, n\}$ let f_ψ be a k-ary function symbol. Then the set of (n, k)* dissected weak near-unanimity identities *consists of the identities*

$$f_\psi(x, \dots, x, \overset{i}{\downarrow}{y}, x, \dots, x) = g_{\psi(i)}(x, y)$$

for all injective functions $\psi \colon \{1, \dots, k\} \to \{1, \dots, n\}$ and $i \in \{1, \dots, k\}$.

Note that any polymorphism clone which satisfies identities of the form

$$f(y, x, \dots, x) = \dots = f(x, \dots, x, y),$$

called *k-ary weak near-unanimity identities* when f is k-ary for some $k \geq 3$, must also satisfy the (n, k) dissected weak near-unanimity identities for all $n > k$. This can be seen by setting $f_\psi = f$ for every ψ. Moreover, there exist polymorphism clones which satisfy dissected weak near-unanimity identities, but do not satisfy any weak near-unanimity identities: one example is the polymorphism clone consisting of all injective functions (up to dummy variables) on a countable set, see [4]. Hence, we can regard dissected weak near-unanimity identities as a strict weakening of the weak near-unanimity identities.

Further note that, for all parameters $m \geq n > k > 1$, the (n, k) dissected weak near-unanimity identities form a subset of the (m, k) dissected weak near-unanimity identities. Thus for every fixed $k > 1$ the family of (n, k) dissected weak near-unanimity identities form an infinite chain of h1 identities of increasing strength. In the special case $k = 2$, the satisfaction of any of the $(n, 2)$ dissected weak near-unanimity identities is equivalent to the existence of a binary commutative polymorphism (as they imply $g_1(x, y) = g_2(y, x) = g_3(x, y) = g_1(y, x)$).

▶ **Lemma 25.** *For all $n > k > 1$ the (n, k) dissected weak near-unanimity identities are non-trivial.*

Proof. Assume to the contrary that there exist projections $g_1, \ldots, g_n \in \mathscr{P}$ and $f_\psi \in \mathscr{P}$ for every injection $\psi\colon \{1, \ldots, k\} \to \{1, \ldots, n\}$ that satisfy the (n, k) dissected weak near-unanimity identities. First, suppose that there are two distinct $1 \leq i, j \leq k$ such that g_i, g_j are both the projection onto the second coordinate. Then let ψ be an injective function with $\psi(1) = i, \psi(2) = j$. It follows from the identities that $f_\psi(y, x, \ldots, x) = f_\psi(x, y, \ldots, x) = y$ holds for all values of the variables, which contradicts f_ψ being a projection. Therefore at most one operation g_i equals the projection to its second coordinate. Since $n > k$, there is an injective function $\psi\colon \{1, \ldots, k\} \to \{1, \ldots, n\}$ such that $g_{\psi(i)}$ is the first projection for all $i \in \{1, \ldots, k\}$. Then f_ψ satisfies the weak near-unanimity identities, which again contradicts f_ψ being a projection. ◀

▶ **Lemma 26.** *Let $\mathbb{U}$ be a relational structure and let $n \geq 2$. Then there exists an embedding from $\mathbb{U}^2$ into $\mathbb{U}$ if and only if there exists an embedding from $\mathbb{U}^n$ into $\mathbb{U}$.*

Proof. If there is an embedding $f\colon \mathbb{U}^n \to \mathbb{U}$ for some $n \geq 2$, then $g\colon \mathbb{U}^2 \to \mathbb{U}$, defined by $g(x, y) := f(x, y, \ldots, y)$, is also an embedding. On the other hand, if for some $n \geq 2$ there exist embeddings $g\colon \mathbb{U}^2 \to \mathbb{U}$ and $h\colon \mathbb{U}^n \to \mathbb{U}$, then the composition $f(x_1, \ldots, x_{n+1}) := g(h(x_1, \ldots, x_n), x_{n+1})$ is an embedding from $\mathbb{U}^{n+1}$ into $\mathbb{U}$. Hence by induction the existence of an embedding from $\mathbb{U}^2$ into $\mathbb{U}$ implies the existence of an embedding from $\mathbb{U}^n$ into $\mathbb{U}$ for all $n \geq 2$. ◀

Proof of Theorem 5. For all $l \geq 2$, define $X_l \subseteq F^l$ by

$$X_l := \bigcup_{a,b \in F} \{(a, \ldots, a, b), (a, \ldots, a, b, a), \ldots, (b, a, \ldots, a)\},$$

and let $\mathbb{X}_l$ be the substructure which X_l induces in $\mathbb{U}^l$.

The first step of our proof is to show that if $n \geq k$, then there exists an embedding $h\colon \mathbb{X}_k \to \mathbb{X}_n$ such that $\mathbf{x}$ is an initial segment of $h(\mathbf{x})$ for all $\mathbf{x} \in X_k$. Let us first assume that $k \geq 3$. For every tuple $\mathbf{x} \in \mathbb{X}_k$ we denote the unique element of F which occurs more than once among its entries by $s(\mathbf{x})$. Define $h\colon X_k \to X_n$ to be the map that extends the tuple $\mathbf{x}$ by $n - k$ many entries with value $s(\mathbf{x})$. In order to prove that h is an embedding let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in X_k$ be such that $R^{\mathbb{U}^k}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ holds for some m-ary relation symbol R in the signature of $\mathbb{U}$. By assumption (i) we have $m < k$. Thus there exists $1 \leq j \leq k$ such that the projection of each $\mathbf{x}_i$ to its j-th coordinate equals $s(\mathbf{x}_i)$. Therefore $(s(\mathbf{x}_1), \ldots, s(\mathbf{x}_m)) \in R^{\mathbb{U}}$, and hence h is a homomorphism. Also its inverse – the projection of n-tuples to the first k-coordinates – is a homomorphism, and thus h is an embedding. Now assume the remaining case where $k = 2$. Define a map $h\colon X_2 \to X_n$ by $(x_1, x_2) \mapsto (x_1, x_2, \ldots, x_2)$. To check that that h is an embedding, by assumption (i), we only need to check that h is an embedding with respect to unary relations, which however follows from its definition.

Observe that h was defined in such a way that, for each index $1 \leq i \leq k$, the i-th projection of $h(\mathbf{x})$ is equal to x_i. By permuting the coordinates of its image in a suitable manner, we can obtain embeddings $h_\psi\colon \mathbb{X}_k \to \mathbb{X}_n$ for every injection $\psi\colon \{1, \ldots, k\} \to \{1, \ldots, n\}$ such that the $\psi(i)$-th projection of $h_\psi(\mathbf{x})$ is equal to x_i for all $1 \leq i \leq k$.

In order to construct the operations f_ψ on F, let $f\colon \mathbb{U}^k \to \mathbb{U}$ and $g\colon \mathbb{U}^n \to \mathbb{U}$ be embeddings, which exist by (ii) and Lemma 26. For every injection $\psi : \{1, \ldots, k\} \to \{1, \ldots, n\}$ define the map $u_\psi\colon f(\mathbb{X}_k) \to g(\mathbb{X}_n)$ by

$$u_\psi(f(a, \ldots, a, \underset{i^{\text{th}}}{b}, a, \ldots, a)) := g(a, \ldots, a, \underset{\psi(i)^{\text{th}}}{b}, a, \ldots, a). \tag{1}$$

Then u_ψ is equal to $g \circ h_\psi \circ f^{-1}$. Since h_ψ is an embedding, $u_\psi\colon f(\mathbb{X}_k) \to u_\psi(f(\mathbb{X}_k))$ is an isomorphism between finite substructures of $\mathbb{U}$. By the homogeneity of $\mathbb{U}$, it can be extended to an automorphism v_ψ of $\mathbb{U}$. Set $f_\psi := v_\psi \circ f$ and, for all $1 \leq i \leq n$, define $g_i(x, y) := g(x, \ldots, x, y, x, \ldots, x)$, where the only y appears at the i-th coordinate of g. It then follows from (1) that these polymorphisms satisfy the (n, k) dissected weak near-unanimity identities on F, concluding the proof. ◀

6.2 Revisiting the infinite-language counterexample

We now investigate the simplification $\mathbb{S}$ of the infinite-language structure from [13]; $\mathbb{S}$ satisfies the same local and global h1 identities as the structure in [13]. Namely, we consider the superposition as in [13, Construction 6.4], but directly of the CSS structures in the proof of [13, Lemma 6.3] rather than of their model-complete cores; we are able to do this due to our constructive, rather than indirect, proof of the local satisfaction of h1 identities. The structure $\mathbb{S}$ has the following properties.

▶ **Proposition 27** (Consequence of the results from [13]). *There exist ω-categorical structures $\mathbb{S}$ and $\mathbb{H}$ with slow orbit growth and without algebraicity, as well as a strictly increasing function $\alpha\colon \mathbb{N} \to \mathbb{N}$ such that the following hold:*

(1) *$\mathbb{H}$ is a homogeneous expansion of $\mathbb{S}$ by pp-definable relations.*

(2) *Every relation of $\mathbb{H}$ has arity $k \cdot \alpha(n)$ for some $k, n \geq 1$, and for every $n \geq 1$ there exist only finitely many relations of arity of the form $k \cdot \alpha(n)$. Moreover, if $(a_1, \ldots, a_{k\cdot\alpha(n)}) \in R$ for some relation R of $\mathbb{H}$, then $\{a_1, \ldots, a_{k\cdot\alpha(n)}\}$ has size at least $\alpha(n)$.*

(3) *$\mathbb{H}$ is homomorphically bounded.*

(4) *There exists a minion homomorphism from $\mathrm{Pol}(\mathbb{S})$ to $\mathscr{P}$.*

Proof of Theorem 3. Let $\mathbb{S}, \mathbb{H}$ be as in Proposition 27. We can assume that $\mathbb{S}$ has signature ρ as in Section 5 since this change does not affect the properties claimed in Proposition 27; see the remark at the beginning of Section 5.1. Since $\mathbb{S}$ has no algebraicity, it has an ω-categorical finite-language encoding $\mathbb{U} = \mathbf{E}\,\mathbb{S}$. Moreover, since $\mathbb{S}$ has slow orbit growth, so does $\mathbb{U}$ by Proposition 19. There is a minion homomorphism $\mathrm{Pol}(\mathbb{U}) \to \mathrm{Pol}(\mathbb{S})$ by Proposition 20 (1) and a minion homomorphism $\mathrm{Pol}(\mathbb{S}) \to \mathscr{P}$ by Proposition 27, thus we obtain a minion homomorphism $\mathrm{Pol}(\mathbb{U}) \to \mathscr{P}$ by composition.

It remains to prove that there is no uniformly continuous minion homomorphism from $\mathrm{Pol}(\mathbb{U})$ to $\mathscr{P}$. We do so by showing that for every finite subset F of the domain, there exists $k > 1$ such that $\mathrm{Pol}(\mathbb{U})$ satisfies the (n, k) dissected weak near-unanimity identities on F for all

$n > k$. Note that $\mathbb{U} = \mathbf{E}\,\mathbb{S}$ is a reduct of the blowup $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$, and hence $\mathrm{Pol}(\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}) \subseteq \mathrm{Pol}(\mathbf{E}\,\mathbb{S})$. It is therefore sufficient to prove that there exists some $k > 1$ for which $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$ satisfies the (n, k) dissected weak near-unanimity identities on F for all $n > k$.

In order to prove this statement, we verify that conditions (i) and (ii) of Theorem 5 hold for $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$, F, and a suitable $k > 1$. Since $\mathbb{H}$ is homomorphically bounded, it is well-known that $\mathbb{H}^2$ embeds into $\mathbb{H}$. By Proposition 20 (3), there exists an embedding of $(\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S})^2$ into $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$, and thus condition (ii) holds.

It remains to check (i) which states that there exists an upper bound on the arity of tuples in F that satisfy some relation from $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$. Denote the signature of $\mathbb{H}$ by σ. Suppose that $R^{\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}}$ contains a tuple entirely within F for some $R \in \sigma^+$, the language of $\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}$. Since $\sigma^+ = \sigma \cup \theta$, and all relations in θ have arity at most 4, we may assume that $R \in \sigma$. Then any tuple in $R^{\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}}$ must lie entirely within $P^{\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}}$, and essentially the proof of Proposition 18 shows that this implies that the tuple is an element of $R^{\mathbb{H}}$.

By Proposition 27 (2), R has arity $k \cdot \alpha(n)$ for some $n, k \geq 1$ and at least $\alpha(n)$ many of the values of any tuple in $R^{\mathbb{H}}$ are distinct. Therefore, $\alpha(n)$ must be smaller than $|F|$. Since α is a strictly increasing function, it follows that only finitely many relations of $R^{\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}}$ have tuples that lie entirely in F. Let $k > 1$ be a strict upper bound on the arity of those relations. For this choice of k we have that (i) of Theorem 5 holds, and thus $R^{\overrightarrow{\mathbf{B}}_{\mathbb{H}}\,\mathbb{S}}$ satisfies the (n, k) dissected weak near-unanimity identities on F for all $n > k$. ◀

▶ Remark 28. It follows that the original structure $\mathbb{S}$ satisfies dissected weak near-unanimity identities locally as well, since by Proposition 20 (1), there is a uniformly continuous minion homomorphism from $\mathrm{Pol}(\mathbf{E}\,\mathbb{S})$ to $\mathrm{Pol}(\mathbb{S})$. This result is new and no other explicit description of non-trivial local h1 identities of $\mathbb{S}$ was given in [13].

7 Identities and CSPs with Homogeneous Templates

7.1 Encoding arbitrary languages as CSPs

Let Σ be a finite alphabet and $W \subseteq \Sigma^{\geq 2}$. Let ρ_W be the signature consisting of one $|w|$-ary relation symbol R_w for every word $w \in W$. The *trivial structure* $\mathbb{T}_W$ is the countable ρ_W-structure whose relations are all empty. For every word $w \in W$, the *w-edge structure* $\mathbb{F}_w$ is the ρ_W-structure on the set $F_w = \{1, \ldots, |w|\}$ whose only non-empty relation is $R_w^{\mathbb{F}_w} = \{(1, \ldots, |w|)\}$.

The trivial structure $\mathbb{T}_W$ is homomorphically bounded by the set of all edge-structures $\mathbb{F}_w$ with $w \in W$. Moreover, $\mathbb{T}_W$ has no algebraicity. It is not hard to see that $\mathbf{E}\,\mathbb{T}_W$ is homogeneous. Applying Theorem 5 and a compactness argument, one can see that $\mathbf{E}\,\mathbb{T}_W$ satisfies non-trivial h1 identities.

Since $\mathbb{T}_W$ is homomorphically bounded, Corollary 23 can be used to give an explicit description of $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$, which we use to prove the results of this section. We employ the notion of coNP-many-one reduction first defined in [8]: a language K coNP-many-one reduces to L if there is a non-deterministic polynomial-time Turing machine M such that for all words w, we have $w \in K$ if and only if each computational path of M, on input w, produces a word in L. Note that if K has a coNP-many-one reduction to L, then in particular K is in coNP^L.

▶ **Proposition 29.** *Let $L \subseteq \Sigma^{\geq 2}$ and $W := \Sigma^{\geq 2} \setminus L$ both be nonempty. Then L has a log-space many-one reduction to $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$, and $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ coNP-many-one reduces to L.*

Proof. The function $w \mapsto \mathbf{C}\,\mathbb{F}_w$, for every $w \in \Sigma^{\geq 2}$, is computable in logarithmic space with respect to $|w|$ by Proposition 21. Also note that there is a homomorphism $\mathbf{C}\,\mathbb{F}_u \to \mathbf{C}\,\mathbb{F}_w$ if and only if $w = u$. Moreover, it follows from Proposition 22 that there is a homomorphism $\mathbf{C}\,\mathbb{F}_w \to \mathbf{E}\,\mathbb{T}_W$ if and only if $w \in L$. Thus L has a log-space many-one reduction to $\mathrm{CSP}(\mathbb{T}_W)$.

For the other reduction, let $\mathbb{X}$ be a finite θ-structure, an instance of $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$. If there is no homomorphism $\mathbb{X} \to \mathbf{E}\,\mathbb{T}_W$, by Proposition 22, either $\mathbb{X}$ is not separated (which can be checked in polynomial time), or there is a word $w \in W$ not longer than the size of the domain of $\mathbb{X}$ and a homomorphism $f\colon \mathbf{C}\,\mathbb{F}_w \to \mathbb{X}$. The reduction does the following: if $\mathbb{X}$ is not separated, we map it to a fixed element of W. Otherwise, we guess a word w not longer than the size of the domain of $\mathbb{X}$ and a function $f\colon \mathbf{C}\,\mathbb{F}_w \to \mathbb{X}$. If this function is not a homomorphism, we map $\mathbb{X}$ to a fixed word of L. If f is a homomorphism, we map $\mathbb{X}$ to w. Thus, if $\mathbb{X} \in \mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ then all runs of the reduction output a word of L. Moreover, if $\mathbb{X} \notin \mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$, then at least one run outputs word in W. ◀

We can now prove the first statement of Theorem 8.

Proof of the first statement of Theorem 8. Let $L \subseteq \Sigma^{\geq 2}$ be a $\mathrm{coNP}^{\mathcal{C}}$-complete language, and let W be its complement. Then L reduces to $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ by Proposition 29, so $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ is $\mathrm{coNP}^{\mathcal{C}}$-hard. Moreover, it follows from Proposition 29 and the fact that $\mathrm{coNP}^{\mathcal{C}}$ is closed under coNP-many-one reductions that $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ belongs to $\mathrm{coNP}^{\mathcal{C}}$. ◀

The second statement of Theorem 8 follows from the following proposition whose proof is inspired by Ladner's proof on the existence of NP-intermediate problems [25].

▶ **Proposition 30.** *Let $L \subseteq \{0,1\}^{\geq 2}$ be a language in coNP $\setminus$ P. Then there is a unary language $I \subseteq \{0\}^{\geq 2}$ such that $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_I)$ is also in coNP $\setminus$ P, but L is not polynomial-time reducible to $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_I)$.*

Finally, we are ready to prove Theorem 7. In the following, let $\mathcal{L}$ be the extension of existential second-order logic allowing countably many second-order quantifiers, followed by a countable conjunction of first-order formulas. It can be seen that the upward direction of Łoś's theorem and the downward Löwenheim-Skolem theorem hold for this logic.

Proof of Theorem 7. We prove the following: there is no countable set Θ of θ-formulas in $\mathcal{L}$ such that the equivalence $\mathbb{A} \models \Theta \Leftrightarrow \mathrm{CSP}(\mathbb{A}) \in \mathcal{C}$ holds for all homogeneous θ-structures $\mathbb{A}$. This proves the theorem, as the satisfaction of a countable set of identities by polymorphisms can be expressed in $\mathcal{L}$.

Assume that such a Θ exists. Let L be a language over Σ whose Turing-degree is not intersected by $\mathcal{C}$, and let $W = \Sigma^{\geq 2} \setminus L$. For every $n \in \mathbb{N}$, let $W \cap \Sigma^{\leq n}$ be the set of words of length at most n in W. Corollary 23 implies that $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}})$ is finitely bounded, hence $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}})$ is in AC^0. Therefore, since $\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}}$ is homogeneous, our assumption implies that $\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}} \models \Theta$. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$, and let $\mathbb{X}$ be the ultraproduct $\left(\prod_{n\in\mathbb{N}} \mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}}\right)/\mathcal{U}$. Then $\mathbb{X} \models \Theta$ by Łoś's theorem and $\mathbb{X}$ is homogeneous, as all the factors in the ultraproduct are homogeneous. By the Löwenheim-Skolem theorem, $\mathbb{X}$ has a countable elementary substructure $\mathbb{Y}$ that also satisfies Θ. Note that $\mathbb{Y}$ is homogeneous and has the same age as $\mathbb{X}$, as it is an elementary substructure of $\mathbb{X}$.

Finally, we claim that $\mathbb{X}$ and $\mathbf{E}\,\mathbb{T}_W$ have the same age. Every finite substructure of $\mathbf{E}\,\mathbb{T}_W$ embeds into $\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}}$ for all $n \in \mathbb{N}$, by Corollary 23, and therefore into their ultraproduct, which is $\mathbb{X}$. Conversely, assume that a finite structure $\mathbb{C}$ embeds into $\mathbb{X}$. This precisely means that $I := \{n \in \mathbb{N} \mid \mathbb{C} \text{ embeds into } \mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}}\}$ is in $\mathcal{U}$. Moreover, since $\mathcal{U}$ is not principal, I is infinite. Let $w \in W$. Since I is infinite there is an $n \geq |w|$ such that $\mathbb{C}$ embeds into

$\mathbf{E}\,\mathbb{T}_{W\cap\Sigma^{\leq n}}$. Since $w \in W \cap \Sigma^{\leq n}$, Corollary 23 gives that $\mathbf{C}\,\mathbb{F}_w$ does not homomorphically map to $\mathbb{C}$, and that $\mathbb{C}$ is separated. Since this holds for all $w \in W$, it follows that $\mathbb{C}$ embeds into $\mathbf{E}\,\mathbb{T}_W$.

By Theorem 9, the two structures $\mathbb{Y}$ and $\mathbf{E}\,\mathbb{T}_W$ are isomorphic. By Proposition 29, L and $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ have the same Turing-degree, therefore $\mathrm{CSP}(\mathbf{E}\,\mathbb{T}_W)$ is not in $\mathcal{C}$, a contradiction. ◀

References

1 Libor Barto. Finitely related algebras in congruence distributive varieties have near unanimity terms. *Canadian Journal of Mathematics*, 65(1):3–21, 2013.

2 Libor Barto. Finitely related algebras in congruence modular varieties have few subpowers. *Journal of the European Mathematical Society*, 20(6):1439–1471, 2018.

3 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 *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12, 2017.

4 Libor Barto, Michael Kompatscher, Miroslav Olšák, Trung Van Pham, and Michael Pinsker. Equations in oligomorphic clones and the constraint satisfaction problem for ω-categorical structures. *Journal of Mathematical Logic*, 19(2):#1950010, 2019.

5 Libor Barto, Jakub Opršal, and Michael Pinsker. The wonderland of reflections. *Israel Journal of Mathematics*, 223(1):363–398, 2018.

6 Libor Barto and Michael Pinsker. The algebraic dichotomy conjecture for infinite domain constraint satisfaction problems. In *Proceedings of the 31th Annual IEEE Symposium on Logic in Computer Science – LICS'16*, pages 615–622, 2016.

7 Libor Barto and Michael Pinsker. Topology is irrelevant. *SIAM Journal on Computing*, 49(2):365–393, 2020.

8 Richard Beigel, Richard Chang, and Mitsunori Ogiwara. A relationship between difference hierarchies and relativized polynomial hierarchies. *Mathematical Systems Theory*, 26(3):293–310, 1993.

9 Manuel Bodirsky and Martin Grohe. Non-dichotomies in constraint satisfaction complexity. In *Proceedings of the International Colloquium on Automata, Languages and Programming (ICALP)*, Lecture Notes in Computer Science, pages 184–196. Springer Verlag, 2008.

10 Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *Journal of the ACM*, 57(2):1–41, 2009. A conference version appeared in the Proceedings of the Symposium on Theory of Computing (STOC) 2008.

11 Manuel Bodirsky, Florent Madelaine, and Antoine Mottet. A universal-algebraic proof of the complexity dichotomy for monotone monadic snp. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '18, pages 105–114, 2018.

12 Manuel Bodirsky, Barnaby Martin, Michael Pinsker, and András Pongrácz. Constraint satisfaction problems for reducts of homogeneous graphs. *SIAM Journal on Computing*, 48(4):1224–1264, 2019. A conference version appeared in the Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, pages 119:1–119:14.

13 Manuel Bodirsky, Antoine Mottet, Miroslav Olšák, Jakub Opršal, Michael Pinsker, and Ross Willard. Topology is relevant (in the infinite-domain dichotomy conjecture for constraint satisfaction problems). In *Proceedings of the Symposium on Logic in Computer Science – LICS'19*, 2019.

14 Manuel Bodirsky, Antoine Mottet, Miroslav Olšák, Jakub Opršal, Michael Pinsker, and Ross Willard. ω-categorical structures avoiding height 1 identities. *Transactions of the American Mathematical Society*, 2020. Accepted for publication.

15 Manuel Bodirsky and Michael Pinsker. Schaefer's theorem for graphs. *Journal of the ACM*, 62(3):52 pages (article number 19), 2015. A conference version appeared in the Proceedings of STOC 2011, pages 655–664.

16 Manuel Bodirsky and Michael Pinsker. Topological Birkhoff. *Transactions of the American Mathematical Society*, 367:2527–2549, 2015.

17 Manuel Bodirsky, Michael Pinsker, and András Pongrácz. Projective clone homomorphisms. *Journal of Symbolic Logic*, 2019. To appear. Preprint `arXiv:1409.4601`. `doi:10.1017/jsl.2019.23`.

18 Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017*, pages 319–330, 2017.

19 Tomás Feder and Moshe Y. Vardi. Monotone monadic SNP and constraint satisfaction. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 612–622, 1993.

20 Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. *SIAM Journal on Computing*, 28:57–104, 1999.

21 Mai Gehrke and Michael Pinsker. Uniform Birkhoff. *Journal of Pure and Applied Algebra*, 222(5):1242–1250, 2018.

22 Pierre Gillibert, Yannis Jonušas, Michael Kompatscher, Antoine Mottet, and Michael Pinsker. When symmetries are not enough: a hierarchy of hard constraint satisfaction problems. Preprint, 2020. `arXiv:2002.07054`.

23 Wilfrid Hodges. *Model theory*. Cambridge University Press, 1993.

24 Michael Kompatscher and Trung Van Pham. A complexity dichotomy for poset constraint satisfaction. *Journal of Applied Logics*, 5(8):1663–1695, 2018.

25 Richard E. Ladner. On the structure of polynomial time reducibility. *Journal of the ACM*, 22(1):155–171, 1975.

26 Miklós Maróti and Ralph McKenzie. Existence theorems for weakly symmetric operations. *Algebra Universalis*, 59(3-4):463–489, 2008.

27 Sylvain Schmitz. Complexity hierarchies beyond elementary. *ACM Transactions on Computation Theory (TOCT)*, 8(1):3:1–3:36, 2016. `doi:10.1145/2858784`.

28 Dmitriy N. Zhuk. A proof of CSP dichotomy conjecture. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017*, pages 331–342, 2017.