Journal of Applied Logic 10 (2012) 179–186

Contents lists available at SciVerse ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

An optimal construction of Hanf sentences

Benedikt Bollig [a], Dietrich Kuske [b,*]

[a] Laboratoire Spécification et Vérification, École Normale Supérieure de Cachan & Centre National de la Recherche Scientifique, France
[b] Institut für Theoretische Informatik, TU Ilmenau, Germany

ARTICLE INFO

Article history:
Received 6 June 2011
Accepted 9 January 2012
Available online 12 January 2012

Keywords:
Locality of first-order logic
Formulas in normal form

ABSTRACT

We give a new construction of formulas in Hanf normal form that are equivalent to first-order formulas over structures of bounded degree. This is the first algorithm whose running time is shown to be elementary. The triply exponential upper bound is complemented by a matching lower bound.

1. Introduction

Various syntactical normal forms for semantical properties of structures are known. For example, every first-order definable property that is preserved under extensions of structures is definable by an existential first-order sentence (Łoś–Tarski [26,22]). Gaifman's normal form is another example that formalizes the observation that first-order logic can only express local properties [10]. A third example in this line is Hanf's theorem, giving another formalization of locality of first-order logic (at least for structures of bounded degree) [14,7]. Finally, we should also mention the normal form by Schwentick and Barthelmann [24] that rejoins the two formalizations of locality by Gaifman and by Hanf.

Gaifman's and Hanf's theorems have found applications in finite model theory and in particular in parametrized complexity. Namely, they lead to efficient parametrized algorithms deciding whether a formula holds in a (finite) structure [25,19,8,9,16,3,21,17,18] and even to more general algorithms that list all the satisfying assignments [5,15]. Hanf's theorem was also used in the transformation of logical formulas into different automata models [27,13,24,2,11,1,12].

In [4], it was shown that passing from arbitrary formulas to those in Łoś–Tarski or Gaifman normal form leads to a non-elementary blowup. The same paper also proves that for structures of bounded degree, the blowup for Gaifman's normal form is between 2- and 4-fold exponential, and that for Łoś–Tarski normal forms (for a restricted class of structures) is between 2- and 5-fold exponential.

This paper shows that Hanf's normal form can be computed in three-fold exponential time and that this is optimal since there is a necessary blowup of three exponentials when passing from general first-order formulas to their Hanf normal form. We remark (as already observed by Seese [25]) that the first construction of Hanf normal forms [6] is not effective since satisfiability of first-order formulas in graphs of bounded degree is undecidable, also when we restrict to finite structures [28]. Only Seese [25] gave a small additional argument showing that Hanf normal forms can indeed be computed. But his algorithm is not primitive recursive. This was improved later to a primitive-recursive algorithm by Durand and Grandjean [5] and (independently) by Lindell [21]. Their papers do not give an upper bound for the construction of Hanf normal forms, but on the face of it, the algorithm seems not to be elementary.[1] Their algorithm is a quantifier-elimination procedure that

* Corresponding author.
E-mail address: Dietrich.Kuske@tu-ilmenau.de (D. Kuske).

[1] In the meantime, A. Durand has informed us of ongoing work aiming at an elementary upper bound for their algorithm.

1570-8683/$ – see front matter
doi:10.1016/j.jal.2012.01.002

only works if the signature consists of finitely many injective functions (following Seese, one can bi-interpret every structure of bounded degree in such a structure, so this is no real restriction of the algorithm). Differently, our algorithm follows the original proof of Hanf's theorem very closely by examining spheres of bounded diameter, but avoiding the detour via Ehrenfeucht–Fraïssé-games.

2. Definitions and background

Throughout this paper, let L be a finite relational signature and let L_m denote the extension of L by the constants $c_1, c_2, \dots, c_m$. Let $\mathcal{A}$ be an L_m-structure. We write $a \in \mathcal{A}$ when we mean that a is an element of the universe of $\mathcal{A}$. Furthermore, $\bar{a}$ denotes a tuple $(a_1, \dots, a_n)$ of length n of elements of some structure $\mathcal{A}$ and $\bar{x}$ is the list of variables $(x_1, \dots, x_n)$. In both cases, n will be determined by the context. Finally, we define a distance (from $\mathbb{N} \cup \{\infty\}$) on the universe of $\mathcal{A}$ setting $\mathrm{dist}^{\mathcal{A}}(a, b) = 0$ iff $a = b$ and $\mathrm{dist}^{\mathcal{A}}(a, c) = d + 1$ if there exists $b \in \mathcal{A}$ with $\mathrm{dist}^{\mathcal{A}}(a, b) \leqslant d$, there is some tuple in some of the relations of $\mathcal{A}$ that contains both, b and c, and there is no such $b \in \mathcal{A}$ with $\mathrm{dist}^{\mathcal{A}}(a, b) < d$, and $\mathrm{dist}^{\mathcal{A}}(a, b) = \infty$ if $\mathrm{dist}^{\mathcal{A}}(a, b) \neq d$ for all $d \in \mathbb{N}$. Next, the *degree* of $a \in \mathcal{A}$ is the number of elements $b \in \mathcal{A}$ with $\mathrm{dist}^{\mathcal{A}}(a, b) = 1$, the degree of $\mathcal{A}$ is the supremum of the degrees of $a \in \mathcal{A}$.

Let $\mathcal{A}$ be an L-structure, $\bar{a} = (a_1, \dots, a_n) \in \mathcal{A}$, and $d > 0$. Then $B_d^{\mathcal{A}}(\bar{a})$ is the set of elements $b \in \mathcal{A}$ with $\mathrm{dist}^{\mathcal{A}}(a_i, b) < d$ for some $1 \leqslant i \leqslant n$.[2] The *d-sphere around* $\bar{a}$ is the L_n-structure

$$S_d^{\mathcal{A}}(\bar{a}) = \left(\mathcal{A} \restriction B_d^{\mathcal{A}}(\bar{a}), \bar{a}\right).$$

A *d-sphere* (*with n centers*) is an L_n-structure $(\mathcal{A}, \bar{a})$ with $S_d^{\mathcal{A}}(\bar{a}) = \mathcal{A}$. The L_n-structure $(\mathcal{A}, \bar{a})$ is a *sphere* if there exists $d > 0$ such that $(\mathcal{A}, \bar{a})$ is a d-sphere; the least such d is denoted $d(\tau)$ and is the *radius of* $(\mathcal{A}, \bar{a})$. The d-sphere τ is *realized by* $\bar{a}$ *in* $\mathcal{A}$ if

$$\tau \cong S_d^{\mathcal{A}}(\bar{a}).$$

If two L-structures $\mathcal{A}$ and $\mathcal{B}$ satisfy exactly the same first-order sentences, then we write $\mathcal{A} \equiv \mathcal{B}$. If they satisfy the same sentences of quantifier rank $\leqslant r$, then $\mathcal{A} \equiv_r \mathcal{B}$. Provided the degrees of $\mathcal{A}$ and $\mathcal{B}$ are finite, both these concepts can be characterized using the number of realizations of spheres.

Theorem 2.1 *(Hanf [14]). For any L-structures $\mathcal{A}$ and $\mathcal{B}$, we have $\mathcal{A} \equiv \mathcal{B}$ whenever any sphere in $\mathcal{A}$ or $\mathcal{B}$ is finite and any sphere is realized in $\mathcal{A}$ and in $\mathcal{B}$ the same number of times or $\geqslant \aleph_0$ times.*

This result was sharpened by Fagin, Stockmeyer & Vardi (see also Ebbinghaus & Flum [6]) to characterize the relation $\equiv_r$:

Theorem 2.2 *(Fagin et al. [7]). For all $r, f \in \mathbb{N}$ there exist $d, m \in \mathbb{N}$ (where d depends on r, only) such that for any L-structures $\mathcal{A}$ and $\mathcal{B}$ of degree $\leqslant f$, we have $\mathcal{A} \equiv_r \mathcal{B}$ whenever any d-sphere with one center is realized in $\mathcal{A}$ and in $\mathcal{B}$ the same number of times or $\geqslant m$ times.*

Proof of both theorems. The proof proceeds by showing that the respective counting property implies that duplicator has a winning strategy in the Ehrenfeucht–Fraïssé-game [6,20]. This then implies the respective equivalence of $\mathcal{A}$ and $\mathcal{B}$. $\square$

This theorem has (at least) three different applications: The first application (and its original motivation in [7]) is a technique to prove that certain properties $\mathfrak{P}$ are not expressible in first-order logic: One provides two lists of structures $\mathcal{A}_r$ and $\mathcal{B}_r$ where $\mathcal{A}_r$ has the desired property and $\mathcal{B}_r$ does not. Furthermore, for any r, $\mathcal{A}_r$ and $\mathcal{B}_r$ satisfy the counting condition from Theorem 2.2 with d and m determined by r and the degree f of $\mathcal{A}_r$ and $\mathcal{B}_r$. This implies $\mathcal{A}_r \equiv_r \mathcal{B}_r$ and therefore the property $\mathfrak{P}$ cannot be expressed by a first-order sentence of quantifier rank r. Since this holds for all r, the property is not first-order expressible. The simplest such property is connectivity of a graph where $\mathcal{A}_r$ can be chosen a circle of size $\max(m, 2d)$ and $\mathcal{B}_r$ a disjoint union of two copies of $\mathcal{A}_r$ (m and d are the constants from Theorem 2.2 for $f = 2$).

The second application is an efficient evaluation of first-order properties on finite structures of bounded degree [25,9,5]: The idea is to count the number of realizations of spheres up to the threshold m and, depending on the vector obtained that way, decide whether the formula holds or not (we will come back to this aspect later in this section).

The third application is a normal form for first-order sentences [6]. For a finite d-sphere τ with n centers, let $\mathrm{sph}_\tau(\bar{x})$ denote a formula such that $(\mathcal{A}, \bar{a}) \models \mathrm{sph}_\tau$ iff $S_d^{\mathcal{A}}(\bar{a}) \cong \tau$. A *Hanf sentence* asserts that there are at least m realizations of the finite sphere τ with one center. Formally, it has the form

$$\exists x_1, x_2, \dots, x_m : \bigwedge_{1 \leqslant i < j \leqslant m} x_i \neq x_j \wedge \forall x : \left(\left(\bigvee_{1 \leqslant i \leqslant m} x = x_i\right) \to \mathrm{sph}_\tau(x)\right)$$

[2] In the literature, one usually defines $B_d^{\mathcal{A}}(\bar{a})$ as the closed ball. Here, we prefer to consider the open ball which slightly simplifies some later calculations.

which we abbreviate as

$$\exists^{\geqslant m} x : \mathrm{sph}_\tau(x).$$

A sentence is in *Hanf normal form* if it is a Boolean combination of Hanf sentences.

Let φ and ψ be two formulas with free variables in $x_1, \ldots, x_n$. To simplify notation, we will say that φ and ψ are f*-equivalent* if, for all structures $\mathcal{A}$ of degree $\leqslant f$, we have

$$\mathcal{A} \models \forall x_1 \forall x_2 \cdots \forall x_n : (\varphi \leftrightarrow \psi).$$

Corollary 2.3 *(Ebbinghaus & Flum [6]). For every sentence φ and all $f \in \mathbb{N}$, there exists an f-equivalent sentence ψ in Hanf normal form.*

Proof. Let r be the quantifier rank of φ and let d and m denote the numbers from Theorem 2.2. Then there are only finitely many d-spheres of degree $\leqslant f$ with one center; let $(\tau_1, \ldots, \tau_n)$ be the list of these spheres. Now we associate with every structure $\mathcal{A}$ of degree $\leqslant f$ a tuple $t^{\mathcal{A}} \in \{0, 1, \ldots, m\}^n$ as follows: For $1 \leqslant i \leqslant n$, let $t_i^{\mathcal{A}}$ denote the minimum of m and the number of $a \in \mathcal{A}$ with $S_d^{\mathcal{A}}(a) \cong \tau_i$. Note that there are only finitely many tuples $t^{\mathcal{A}}$. Now ψ is a disjunction. It has one disjunct for every $t \in \{0, 1, \ldots, m\}^n$ for which there exists a structure $\mathcal{A}$ of degree $\leqslant f$ with $\mathcal{A} \models \varphi$ and $t = t^{\mathcal{A}}$. This disjunct is the conjunction of the following formulas for $1 \leqslant i \leqslant n$:

$$\begin{cases} \exists^{=t_i} x : \mathrm{sph}_\tau(x) & \text{if } t_i < m, \\ \exists^{\geqslant m} x : \mathrm{sph}_\tau(x) & \text{if } t_i = m. \end{cases} \qquad \square$$

Note that φ is satisfiable if and only if the disjunction ψ is not empty. Hence an effective construction of ψ would allow us to decide satisfiability of first-order formulas in structures of degree $\leqslant f$ which is not possible [28].

We now turn to finite structures. Clearly, the disjunction ψ as in the above corollary is also equivalent to φ for all finite structures of degree $\leqslant f$. But in this context, we can also define another disjunction ψ_{fin} by taking only those $t \in \{0, 1, \ldots, m\}^n$ for which there exists a *finite* structure $\mathcal{A}$ of degree $\leqslant f$ with $\mathcal{A} \models \varphi$ and $t = t^{\mathcal{A}}$ (cf., e.g., [20, p. 101]). As above, an effective construction of ψ_{fin} would allow us to decide satisfiability of first-order formulas in *finite* structures of degree $\leqslant f$ which, again, is not possible [28].

Despite the fact that the proof of Corollary 2.3 is not constructive, Seese showed that some sentence ψ as required in Corollary 2.3 can be computed.

Theorem 2.4 *(Seese [25, p. 523]). From a sentence φ and $f \in \mathbb{N}$, one can compute an f-equivalent sentence in Hanf normal form.*

Proof. Let β express that a structure has degree $\leqslant f$. Then search for a tautology of the form $\beta \to (\varphi \leftrightarrow \psi)$ where ψ is a sentence in Hanf normal form. Since the set of tautologies is recursively enumerable, we can do this search effectively. And since we know from Theorem 2.2 that an f-equivalent sentence in Hanf normal form exists, this search will eventually terminate successfully. $\square$

Note that Seese's procedure to compute ψ is not primitive recursive. A primitive recursive construction of a Hanf normal form was described by Durand and Grandjean [5] and independently by Lindell [21]. They present a quantifier elimination scheme and do not rest their reasoning on Ehrenfeucht–Fraïssé-games. But so far, no elementary upper bound for the running time of their algorithm is known. The main result of this paper is an elementary procedure for the computation of a Hanf normal form. This is achieved by a new (direct) proof of Corollary 2.3 that does not use games.

The effective constructions of Hanf normal forms led Seese [25], Durand and Grandjean [5] and Lindell [21] to efficient algorithms for the evaluation of first-order queries on structures of bounded degree. Seese showed that sentences in Hanf normal form can be evaluated in time linear in the structure and the Hanf normal form. Consequently, the set of pairs $(\mathcal{A}, \varphi)$ with $\mathcal{A}$ a structure of degree $\leqslant f$ and φ a sentence with $\mathcal{A} \models \varphi$ can be decided in time

$$g_1\big(|\varphi|, f\big) + g_2\big(|\varphi|, f\big) \cdot |\mathcal{A}| \tag{1}$$

where $g_1(|\varphi|, f)$ is the time needed to compute the Hanf normal form and $g_2(|\varphi|, f)$ is its size[3] (it can be shown that the function g_2 is elementary since the radiuses appearing in the Hanf normal form can be bound). Since Seese's construction is not primitive recursive, the function g_1 is not primitive recursive. The constructions by Durand and Grandjean and by Lindell show that g_1 can be replaced by a primitive recursive function g_1'. Since they get another Hanf normal form, also the function g_2 changes to g_2', say (but as for Seese's Hanf normal form, also this function is elementary).

In addition, Durand and Grandjean and Lindell show that the set of tuples $\bar{a}$ from $\mathcal{A}$ with $\mathcal{A} \models \varphi(\bar{a})$ can be computed in time

[3] It should be noted that Frick and Grohe proved this problem to be solvable with g_1 the identity and g_2 triply exponential in $|\varphi|$ and f [9].

$$g'_1\big(|\varphi|, f\big) + g'_2\big(|\varphi|, f\big) \cdot \big(|\mathcal{A}| + \big|\{\bar{a} \mid \mathcal{A} \models \varphi(\bar{a})\}\big|\big) \tag{2}$$

where φ is a first-order formula and f is the degree of the structure $\mathcal{A}$. This was recently improved by Kazana and Segoufin who compute this set in time

$$2^{2^{2^{O(|\varphi|)}}} \cdot \big(|\mathcal{A}| + \big|\{\bar{a} \mid \mathcal{A} \models \varphi(\bar{a})\}\big|\big).$$

Here, the triply exponential factor originates from the work by Frick and Grohe [9] and the summand g'_1 is avoided since they do not precompute a Hanf normal form. Our result in this paper will show that the Hanf normal form can be computed in triply exponential time. Consequently, the functions from (1) and from (2) can be replaced by triply exponential functions. As a result, the model checking algorithm by Seese and the enumeration algorithm by Durand and Grandjean and by Lindell perform as well as the algorithms by Frick and Grohe and by Kazana and Segoufin, resp.

3. Construction of a Hanf normal form

A *Hanf formula with free variables from* $x_1, \ldots, x_n$ is a formula of the form

$$\exists^{\geqslant m} y : \mathrm{sph}_\tau(\bar{x}, y)$$

where τ is a sphere with $n+1$ centers. A formula is in *Hanf normal form* if it is a Boolean combination of Hanf formulas.

Theorem 3.1. *From a formula Φ with free variables among $\bar{x}$ and $f \geqslant 1$, one can construct an f-equivalent formula Ψ in Hanf normal form. This construction can be carried out in time*

$$2^{f^{2^{O(|\Phi|)}}}.$$

The construction of Ψ from Φ will be done by structural induction on Φ. The central part in this induction is described by the following lemma (the proof of Theorem 3.1 can be found at the end of this section).

Lemma 3.2. *From a formula φ in Hanf normal form with free variables among $\bar{x}, x_{n+1}$ and $f \geqslant 1$, one can construct a formula ψ in Hanf normal form with free variables in $\bar{x}$ such that $\exists x_{n+1} : \varphi$ and ψ are f-equivalent. This construction can be carried out in time $|\varphi| \cdot 2^{n^{O(1)} \cdot f^{O(d)}}$ where d is the maximal radius of a sphere appearing in φ. Furthermore, the largest radius appearing in ψ is $3d$.*

Proof. Set $e = 3d$. The formula ψ will be a disjunction with one disjunct for every e-sphere τ' with $n+1$ centers. This disjunct will have the form

$$\psi_{\tau'} = \varphi_{\tau'} \wedge \exists^{\geqslant 1} x_{n+1} : \mathrm{sph}_{\tau'}.$$

We next describe how $\varphi_{\tau'}$ is obtained from φ. For this, let $\alpha = \exists^{\geqslant m} x_{n+2} : \mathrm{sph}_\tau$ be some Hanf formula appearing in φ. This formula will be replaced by the Hanf formula α' that we construct next. In this construction, we distinguish two cases, namely whether the $d(\tau)$-sphere around $c_{n+1}c_{n+2}$ in τ is connected or not.

(a) $S^\tau_{d(\tau)}(c_{n+1}c_{n+2})$ is connected.
Let p denote the number of elements $c \in B^{\tau'}_{2d(\tau)}(c_{n+1})$ with

$$S^{\tau'}_d(\bar{c}c_{n+1}c) \cong \tau$$

and set

$$\alpha' = \begin{cases} \text{true} & \text{if } p \geqslant m \\ \text{false} & \text{otherwise.} \end{cases}$$

(b) $S^\tau_{d(\tau)}(c_{n+1}c_{n+2})$ is not connected.
Let

$$\sigma = S^\tau_{d(\tau)}(\bar{c}c_{n+2})$$

and write p for the number of $c \in B^{\tau'}_{2d(\tau)}(c_{n+1})$ with

$$S^{\tau'}_{d(\tau)}(\bar{c}c) \cong \sigma.$$

In this case, set

$$\alpha' = \exists^{\geqslant m+p} x_{n+2} : \mathrm{sph}_\sigma(\bar{x}, x_{n+2}).$$

This finishes the construction of $\varphi_{\tau'}$ and therefore of the disjunction ψ. Clearly, ψ is in Hanf normal form.
Now let $a_{n+1} \in \mathcal{A}$ with $S_d^{\mathcal{A}}(\bar{a}a_{n+1}) \cong \tau'$. We will show

$$(\mathcal{A}, \bar{a}a_{n+1}) \models \alpha \iff (\mathcal{A}, \bar{a}) \models \alpha'$$

again distinguishing the two cases above.

(a) First let $S^{\tau}_{d(\tau)}(c_{n+1}c_{n+2})$ be connected. Then, for $a_{n+2} \in \mathcal{A}$ with $\tau \cong S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2})$, we have

$$\operatorname{dist}^{\mathcal{A}}(a_{n+1}, a_{n+2}) = \operatorname{dist}^{\tau}(c_{n+1}, c_{n+2}) \leqslant 2d(\tau) - 1 < 2d(\tau)$$

and therefore $a_{n+2} \in B^{\mathcal{A}}_{2d(\tau)}(a_{n+1})$. Hence

$$\begin{aligned}
&\left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2}) \cong \tau\right\}\right| \\
&\quad = \left|\left\{a_{n+2} \in B^{\mathcal{A}}_{2d(\tau)}(a_{n+1}) \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2}) \cong \tau\right\}\right| \\
&\quad = \left|\left\{c \in B^{\tau'}_{2d(\tau)}(c_{n+1}) \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{c}c_{n+1}c) \cong \tau\right\}\right| = p
\end{aligned}$$

where the last equality follows from $S_e^{\mathcal{A}}(\bar{a}a_{n+1}) \cong \tau'$ and $e \geqslant 3d(\tau)$. Hence we showed

$$\begin{aligned}
(\mathcal{A}, \bar{a}a_{n+1}) \models \alpha \quad &\iff \quad (\mathcal{A}, \bar{a}a_{n+1}) \models \exists^{\geqslant m} x_{n+2} : \operatorname{sph}_\tau(\bar{x}, x_{n+1}) \\
&\iff \quad p \geqslant m \\
&\iff \quad (\mathcal{A}, \bar{a}) \models \alpha'.
\end{aligned}$$

(b) Next consider the case that $S^{\tau}_{d(\tau)}(c_{n+1}c_{n+2})$ is not connected. Then, for $a_{n+2} \in \mathcal{A}$, we have $S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2}) \cong \tau$ if and only if

$$\operatorname{dist}^{\mathcal{A}}(a_{n+1}, a_{n+2}) \geqslant 2d(\tau) \quad \text{and} \quad S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma.$$

But this implies

$$\begin{aligned}
&\left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2}) \cong \tau\right\}\right| \\
&\quad = \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; \operatorname{dist}^{\mathcal{A}}(a_{n+1}, a_{n+2}) \geqslant 2d(\tau),\, S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| \\
&\quad = \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| \\
&\qquad - \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; \operatorname{dist}^{\mathcal{A}}(a_{n+1}, a_{n+2}) < 2d(\tau),\, S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| \\
&\quad = \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| \\
&\qquad - \left|\left\{c \in \tau' \;\middle|\; \operatorname{dist}^{\tau'}(c_{n+1}, c) < 2d(\tau),\, S^{\tau'}_{d(\tau)}(\bar{c}c) \cong \sigma\right\}\right| \\
&\quad = \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| - p.
\end{aligned}$$

Hence

$$\begin{aligned}
(\mathcal{A}, \bar{a}a_{n+1}) \models \alpha \quad &\iff \quad (\mathcal{A}, \bar{a}a_{n+1}) \models \exists^{\geqslant m} x_{n+2} : \operatorname{sph}_\tau \\
&\iff \quad \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+1}a_{n+2}) \cong \tau\right\}\right| \geqslant m \\
&\iff \quad \left|\left\{a_{n+2} \in \mathcal{A} \;\middle|\; S^{\mathcal{A}}_{d(\tau)}(\bar{a}a_{n+2}) \cong \sigma\right\}\right| \geqslant m + p \\
&\iff \quad (\mathcal{A}, \bar{a}) \models \alpha'.
\end{aligned}$$

We next evaluate the size of the formula ψ. Since ψ is a disjunction of formulas $\psi_{\tau'}$, we first fix some e-sphere τ' with $n+1$ centers (with $e = 3d$). Then τ' has $\leqslant f^{3d-1} \cdot (n+1)$ elements. Hence the formula $\operatorname{sph}_{\tau'}$ has size $\leqslant (f^{3d-1} \cdot (n+1))^{O(1)}$ (the constant $O(1)$ depends on the signature L). Now we deal with the formula $\varphi_{\tau'}$. It results from φ by the replacement of subformulas of the form $\alpha = \exists^{\geqslant m} x_{n_2} : \operatorname{sph}_\tau$. In the first case, $|\alpha'| \leqslant |\alpha|$. In the second case, note that σ is a subsphere of τ, so $|\operatorname{sph}_\sigma| \leqslant |\operatorname{sph}_\tau| < |\alpha|$. Furthermore, $p \leqslant f^{2d(\tau)-1} \leqslant f^{2d-1}$. Recall that the formula

$$\alpha' = \exists^{\geqslant m+p} x_{n_2} : \operatorname{sph}_\sigma(\bar{x}, x_{n+2})$$

is shorthand for

$$\exists y_1, y_2, \ldots, y_{m+p} : \bigwedge_{1 \leqslant i < j \leqslant m+p} y_i \neq y_j \wedge \forall y \Biggl(\Biggl(\bigvee_{1 \leqslant i \leqslant m+p} y = y_i\Biggr) \rightarrow \operatorname{sph}_\sigma(\bar{x}, y)\Biggr).$$

The size of this formula is bounded by

$$O(p^2) + |\mathrm{sph}_\sigma| \leqslant O(f^{4d-2}) + |\alpha|.$$

Since $\varphi_{\tau'}$ is obtained from φ by at most $|\varphi|$ replacements, we obtain

$$|\varphi_{\tau'}| = |\varphi| \cdot O(f^{4d-2}) + |\varphi| \leqslant |\varphi| \cdot f^{O(d)}$$

and therefore

$$\left|\varphi_{\tau'} \wedge \exists^{\geqslant 1} x_{n+1} : \mathrm{sph}_{\tau'}\right| = f^{O(d)} \cdot \left(|\varphi| + n^{O(1)}\right).$$

The number of disjuncts of ψ equals the number of $3d$-spheres with $n+2$ centers. Since any such sphere has at most $f^{3d-1}(n+1)$ elements, the number of these spheres is bounded by

$$2^{(n \cdot f^{3d-1})^{O(1)}} = 2^{n^{O(1)} \cdot f^{O(d)}}$$

which finally results in

$$|\psi| \leqslant 2^{n^{O(1)} \cdot f^{O(d)}} \cdot f^{O(d)} \cdot \left(|\varphi| + n^{O(1)}\right) \leqslant 2^{n^{O(1)} \cdot f^{O(d)}}.$$

We finally come to the evaluation of the time needed to compute ψ. The crucial point in our estimation is the time needed to compute the numbers p in (a) and (b); we only discuss (a).

There are $\leqslant f^{2d(\tau)+1} - 1$ candidates c in $B^{\tau'}_{2d(\tau)}(c_{n+1})$. For any of them, we have to compute the set $B^{\tau'}_d(c)$ (which can be done in time $f^{2d+1} - 1$). Then, isomorphism of τ and $S^{\tau'}_d(\bar{c}c_{n+1}c)$ has to be decided. But these are two structures of degree $\leqslant f$ and of size $(n+2) \cdot (f^{d+1} - 1) \leqslant |\varphi| \cdot (f^{d+1} - 1)$. Hence, by [23], this isomorphism test can be performed in time polynomial in the size of the structures (the degree of the polynomial depends on f).[4] Hence, the number p can indeed be computed within the given time bound. $\square$

We now come to the proof of the central result of this paper:

Proof of Theorem 3.1. The proof is carried out by induction on the construction of the formula Φ. So first, let φ be a quantifier-free subformula of Ψ whose free variables are among $x_1, \ldots, x_n$. Let T be the set of all 1-spheres τ of degree $\leqslant f$ with $n+1$ centers such that the constants $c_1, \ldots, c_n$ of τ satisfy φ. Then set

$$\psi = \bigvee_{\tau \in T} \exists^{\geqslant 1} x_{n+1} : \mathrm{sph}_\tau.$$

Note that any 1-sphere with $n+1$ centers has precisely $n+1$ elements. Furthermore, $n \leqslant |\Phi|$ since φ is a subformula of Φ. Hence the formula sph_τ has size $n^{O(1)} \leqslant |\Phi|^{O(1)}$ and there are $2^{|\Phi|^{O(1)}}$ disjuncts in the formula ψ (where the constants $O(1)$ depend on the signature L), i.e., $|\psi| = 2^{|\Phi|^{O(1)}}$.

We now come to the induction step. The computation of Hanf normal forms of $\neg\varphi$ and of $\varphi \vee \varphi'$ are straightforward from Hanf normal forms of φ and φ'. The only critical point in the induction are subformulas of the form $\exists x_{n+1} : \beta$. By the induction hypothesis, β can be transformed into an f-equivalent Hanf normal form φ and then Lemma 3.2 is invoked yielding an f-equivalent Hanf normal form for $\exists x_{n+1} : \beta$. We have to invoke Lemma 3.2 at most $|\Phi|$ times where the number n is always bounded by $|\Phi|$. Each invocation increases the radius of the spheres considered by a factor of three, so the maximal radius will be $3^{|\Phi|} = 2^{O(|\Phi|)}$. Hence, each invocation of Lemma 3.2 increases the formula by a factor of $2^{f^{2^{O(|\Phi|)}}}$. Putting this to the power of $|\Phi|$ does not change the expression. $\square$

4. Optimality

In this section, we give a matching lower bound for the size of an f-equivalent formula in Hanf normal form. Namely, we prove

Theorem 4.1. *There is a family of sentences $(\chi_n)_{n \in \mathbb{N}}$ such that $|\chi_n| \in O(n)$ and every 3-equivalent formula ψ_n in Hanf normal form has $\geqslant 2^{2^{2^n+1}-1}$ subformulas, so $|\psi_n| \geqslant 2^{2^{2^n+1}-1}$.*

[4] For this result to apply, one has to code the L-structure into a graph. This standard technique is explained, e.g., in [25, Proof of Theorem 3.2].

The formulas χ_n will speak about labeled trees. More formally, our signature L consists of two binary relations S_0 and S_1 and one unary relation U. A structure $\mathcal{A} = (A, S_0^{\mathcal{A}}, S_1^{\mathcal{A}}, U^{\mathcal{A}})$ over this signature is a *tree* if there is a finite, nonempty and prefix-closed set $X \subseteq \{0,1\}^*$ such that

$$\mathcal{A} \cong \big(X, \big\{(u, u0) \mid u0 \in X\big\}, \big\{(u, u1) \mid u1 \in X\big\}, H\big)$$

for some $H \subseteq X$. Note that every tree has degree at most 3. The tree is *complete* if every inner node has two children and any two maximal paths have the same length, this length is called the *height* of the tree (i.e., $X = \{0,1\}^{\leqslant h}$ where h is the height). A *forest* is a disjoint union of trees. As in [9, Lemma 23], one can construct formulas χ_n of size $O(n)$ such that for every forest $\mathcal{A}$, we have

$$\mathcal{A} \models \chi_n \quad \text{if and only if any two complete trees of height } 2^n \text{ in } \mathcal{A} \text{ are non-isomorphic.} \tag{3}$$

Lemma 4.2. *Let ψ be a formula in Hanf normal form that is* 3*-equivalent to χ_n. Then there are $2^{2^{2^n+1}-1}$ non-isomorphic spheres σ such that the formula* sph_σ *appears in ψ.*

Proof. Suppose, towards a contradiction, that ψ contains $< 2^{2^{2^n+1}-1}$ subformulas of the form sph_σ.

Let M be the maximal number m such that $\exists^{\geqslant m} x : \mathrm{sph}_\sigma$ appears in ψ (for any sphere σ). We can assume that ψ does not contain any formula sph_σ where σ is a 1-sphere. The complete tree of height 2^n has $2^{2^n+1} - 1$ nodes. Hence there are $2^{2^{2^n+1}-1}$ ways to color such a tree. By our assumption on ψ, there is one such tree $\mathcal{B}$ (with root r) such that the formula $\mathrm{sph}_{(\mathcal{B},r)}$ does not appear in ψ.

Next, we need a bit of terminology. If $\mathcal{A}$ is a tree, a a node in τ, and $d \in \mathbb{N}$, then also $\tau \restriction B_d^{\mathcal{A}}(a)$ is a tree that we denote $N_d^{\mathcal{A}}(a)$. Recall that the sphere $S_d^{\mathcal{A}}(a) = (N_d^{\mathcal{A}}(a), a)$ about a of radius d has an additional constant.

Now we define a structure $\mathcal{A}_0$. It consists of $M+1$ copies of any of the structures $N_d^{\mathcal{B}}(b)$ where

(1) $1 < d \leqslant 2^n$ and b is not the root of $\mathcal{B}$ or
(2) $d < 2^n$.

Finally, let $\mathcal{A}_2 = \mathcal{A}_0 \uplus \mathcal{B} \uplus \mathcal{B}$ be the disjoint union of $\mathcal{A}_0$ and two copies of the tree $\mathcal{B}$. Then, by (3), we have $\mathcal{A}_2 \not\models \psi$. Since $\mathcal{A}_0$ does not contain any complete tree of height 2^n, we get $\mathcal{A}_0 \models \chi_n$ and therefore $\mathcal{A}_0 \models \psi$. Note that any sphere realized in $\mathcal{A}_0$ or $\mathcal{A}_2$ is also realized in $\mathcal{B}$. So let $b \in \mathcal{B}$, and $d \in \mathbb{N}$. We distinguish several cases:

(1) $1 < d \leqslant 2^n$ and b is not the root of $\mathcal{B}$. Then the sphere $(N_d^{\mathcal{B}}(b), b)$ is realized in $\mathcal{A}_0$ more than M times, hence the same holds for $\mathcal{A}_2$.
(2) $d < 2^n$. Then $(N_d^{\mathcal{B}}(b), b)$ is realized in $\mathcal{A}_0$ more than M times, hence the same holds for $\mathcal{A}_2$.
(3) b is the root of the tree $\mathcal{B}$ and $d = 2^n$. Then $N_d^{\mathcal{B}}(b) = \mathcal{B}$. Hence $S_d^{\mathcal{B}}(b)$ is not realized in $\mathcal{A}_0$ and it is realized twice in $\mathcal{A}_2$. But validity of ψ does not depend on this number since ψ does not mention the formula $\mathrm{sph}_{(\mathcal{B},b)}$.

Hence, we obtain $\mathcal{A}_2 \models \psi$, contrary to our assumption that χ_n and ψ are 3-equivalent. □

The theorem now follows immediately from this lemma.

Acknowledgements

We would like to thank Luc Segoufin and Arnaud Durand for comments on an earlier version and for hints to the literature that improved this paper.

References

[1] B. Bollig, D. Kuske, Muller message-passing automata and logics, Information and Computation 206 (2008) 1084–1094.
[2] B. Bollig, M. Leucker, Message-passing automata are expressively equivalent to EMSO logic, Theoretical Computer Science 358 (2–3) (2006) 150–172.
[3] A. Dawar, M. Grohe, S. Kreutzer, N. Schweikardt, Approximation schemes for first-order definable optimisation problems, in: LICS'06, IEEE Computer Society Press, 2007, pp. 411–420.
[4] A. Dawar, M. Grohe, S. Kreutzer, N. Schweikardt, Model theory makes formulas large, in: ICALP'07, in: Lecture Notes in Computer Science, vol. 4596, Springer, 2007, pp. 913–924.
[5] A. Durand, J. Grandjean, First-order queries on structures of bounded degree are computable with constant delay, ACM Transactions on Computational Logic 8 (4) (2007).
[6] H.-D. Ebbinghaus, J. Flum, Finite Model Theory, Springer, 1995.
[7] R. Fagin, L. Stockmeyer, M. Vardi, On monadic NP vs. monadic co-NP, Information and Computation 120 (1) (1995) 78–92.
[8] M. Frick, M. Grohe, Deciding first-order properties of locally tree-decomposable structures, Journal of the ACM 48 (2001) 1184–1206.
[9] M. Frick, M. Grohe, The complexity of first-order and monadic second-order logic revisited, Annals of Pure and Applied Logic 130 (1–3) (2004) 3–31.
[10] H. Gaifman, On local and nonlocal properties, in: J. Stern (Ed.), Logic Colloquium'81, North-Holland, 1982, pp. 105–135.

[11] P. Gastin, D. Kuske, Uniform satisfiability in PSPACE for local temporal logics over Mazurkiewicz traces, Fundamenta Informaticae 80 (2007) 169–197.
[12] P. Gastin, D. Kuske, Uniform satisfiability problem for local temporal logics over Mazurkiewicz traces, Information and Computation 208 (2010) 797–816.
[13] D. Giammarresi, A. Restivo, S. Seibert, W. Thomas, Monadic second-order logic over rectangular pictures and recognizability by tiling systems, Information and Computation 125 (1) (1996) 32–45.
[14] W. Hanf, Model-theoretic methods in the study of elementary logic, in: J. Addison, L. Henkin, A. Tarski (Eds.), The Theory of Models, North Holland, 1965, pp. 132–145.
[15] W. Kazana, L. Segoufin, First-order query evaluation on structures of bounded degree, Logical Methods in Computer Science 7 (2(20)) (2011) 1–8.
[16] D. Kuske, M. Lohrey, Decidable first-order theories of one-step rewriting in trace monoids, Theory of Computing Systems 38 (2005) 39–81.
[17] D. Kuske, M. Lohrey, First-order and counting theories of ω-automatic structures, Journal of Symbolic Logic 73 (2008) 129–150.
[18] D. Kuske, M. Lohrey, Automatic structures of bounded degree revisited, Journal of Symbolic Logic 76 (4) (2011) 1352–1380.
[19] L. Libkin, Logics with counting and local properties, ACM Transactions on Computational Logics 1 (2000) 33–59.
[20] L. Libkin, Elements of Finite Model Theory, Springer, 2004.
[21] S. Lindell, A normal form for first-order logic over doubly-linked data structures, International Journal of Foundations of Computer Science 19 (1) (2008) 205–217.
[22] J. Łoś, On the extending of models (i), Fundamenta Mathematicae 42 (1955) 38–54.
[23] E. Luks, Isomorphism of graphs of bounded valence can be tested in polynomial time, Journal of Computer and System Sciences 25 (1) (1982) 42–65.
[24] T. Schwentick, K. Barthelmann, Local normal forms for first-order logic with applications to games and automata, Discrete Mathematics and Computer Science 3 (1999) 109–124.
[25] D. Seese, Linear time computable problems and first-order descriptions, Mathematical Structures in Computer Science 6 (1996) 506–526.
[26] A. Tarski, Contributions to the theory of models. I, Konink. Nederl. Adak. Wetensch. Proc. Ser. A 57 (1954) 572–581.
[27] W. Thomas, Languages, automata, and logic, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Springer-Verlag, 1997, pp. 389–455.
[28] R. Willard, Hereditary undecidability of some theories of finite structures, Journal of Symbolic Logic 59 (4) (1994) 1254–1262.