An optimal construction of Hanf sentences *

Benedikt Bollig\textsuperscript{1}  
Dietrich Kuske\textsuperscript{2}

\textsuperscript{1}Laboratoire Spécification et Vérification
\textsuperscript{2}Technische Universität Ilmenau

First-order logic can only express local properties. Every first-order sentence is equivalent to ...

- [Gaifman '82] a boolean combination of formulas of the form
  “there are at least $k$ elements $x$ of pairwise distance $>2d$
  whose $d$-sphere satisfies $\varphi(x)$”
  where $\varphi(x)$ is $d$-local around $x$.

- [Schwentick & Barthelmann '99] a formula of the form
  $\exists x_1 \ldots \exists x_k \forall x \varphi(x_1, \ldots, x_k, x)$
  where $\varphi(x_1, \ldots, x_k, x)$ is $d$-local around $x$.

- [Hanf '65] a boolean combination of formulas of the form
  “there are at least $k$ elements $x$ whose $d$-sphere has
  isomorphism type $\tau$”
  (over structures of bounded degree).
Hanf normal form

Applications of Hanf normal forms

The construction

Optimality
Hanf normal form

Applications of Hanf normal forms

The construction

Optimality
Terminology

- \( L \) is a finite relational signature
  \( L_n \) is its extension with \( n \) constants

- a \textit{d-sphere} (with \( n \) centers) is an \( L_n \)-structure where every element has distance \( \leq d \) to some constant

- distance in Gaifman graph of \( L_n \)-structure \( \mathcal{A} \):
  \( \{a, b\} \in \text{Edges} \) if (\( \ldots, a, \ldots, b, \ldots \)) \( \in R^\mathcal{A} \) for some \( R \in L \)

- the \textit{degree} of \( \mathcal{A} \) is the degree of the Gaifman graph of \( \mathcal{A} \)
Theorem [Hanf ’65]

∀ \mathcal{A}, \mathcal{B} \text{ locally finite } L\text{-structures}

\mathcal{A} \equiv \mathcal{B} \text{ whenever }

any one-centered sphere is realized in \mathcal{A} and in \mathcal{B} the same number of times or infinitely often. (*)

\hline
\mathcal{A} \equiv \mathcal{B}: \text{ FO sentences cannot distinguish } \mathcal{A} \text{ and } \mathcal{B}.
\mathcal{A} \equiv_r \mathcal{B}: \text{ FO sentences of quantifier rank } r \text{ cannot distinguish } \mathcal{A} \text{ and } \mathcal{B}. 
\hline
Theorem [Fagin, Stockmeyer, Vardi ’95]
\[ \forall r, f \in \mathbb{N} \exists d, m \in \mathbb{N} \forall A, B \ L\text{-structures of degree } \leq f: \]
\[ A \equiv_r B \text{ whenever} \]
any one-centered \textit{d-sphere} is realized in \( A \) and in \( B \)
the same number of times or \( \geq m \) times. (*)

\textbf{Proof:} (*) implies winning strategy for duplicator

Theorem implies existence of Hanf normal form:
a boolean combination of formulas “\textit{d-sphere} \( \tau \) is realized at least \( k \) times”

\[ A \equiv B: \text{ FO sentences cannot distinguish } A \text{ and } B. \]
\[ A \equiv_r B: \text{ FO sentences of quantifier rank } r \text{ cannot distinguish } A \text{ and } B. \]
Terminology

- $\tau_1, \ldots, \tau_n$ all one-centered $d$-spheres of degree $\leq f$
- $t_i^A = \text{minimum of } m \text{ and number of realisations of } \tau_i \text{ in } A$
- $t^A = (t_1^A, t_2^A, \ldots, t_n^A) \in \{0, 1, \ldots, m\}^n$
- $T_\varphi = \{t^A \mid A \text{ of degree } \leq f \text{ with } A \models \varphi\}$

Corollary

For every sentence $\varphi$ of quantifier rank $\leq r$ and every $L$-structure $A$ of degree $\leq f$,

$$A \models \varphi \text{ if and only if } t^A \in T_\varphi.$$ 

Note: The set $T_\varphi$ from proof is empty iff $\varphi$ is contradictory. Hence above $T_\varphi$ cannot be computed from $\varphi$, also if we restrict to finite structures [Willard ’94].
Definition

• Let $S^A_d(\bar{x})$ denote the $d$-sphere around $\bar{x}$ in $A$.

• For a $d$-sphere $\tau$ with $n$ centers, the formula $\text{sph}_d(\bar{x}) \cong \tau$ expresses “the $d$-sphere around $\bar{x}$ is isomorphic to $\tau$”.

• A formula $\psi(\bar{x})$ is in Hanf normal form if it is a Boolean combination of formulas $\exists \geq k x : \text{sph}_d(\bar{x}, x) \cong \tau$.

• Two formulas are $f$-equivalent if they are equivalent on all structures of degree $\leq f$.

Corollary

For every formula $\varphi$ and every $f \in \mathbb{N}$, there exists an $f$-equivalent formula in Hanf normal form.
Corollary [Seese ’96]

From a formula $\varphi$ and $f \in \mathbb{N}$, one can compute an $f$-equivalent formula in Hanf normal form.

Proof: Let $\beta$ express “the structure has degree $\leq f$”. Enumerate all tautologies until you find one of the form

$$\beta \rightarrow (\varphi \leftrightarrow \psi)$$

where $\psi$ is in Hanf normal form.

Output $\psi$. 
Remark

- this construction is not primitive recursive
- [Durand & Grandjean '07] and [Lindell '08]: primitive recursive constructions by elimination of quantifiers and change of signature (superfluous analysis: non-elementary)
- [Clochard '12]: construction by Durand & Grandjean is 4-fold exponential
Hanf normal form

Applications of Hanf normal forms

The construction

Optimality
Automata theory

- logical characterization of automata that can “compute spheres”:
  - graph acceptors [Thomas '90]
  - finite automata, tree automata, ...
  - communicating finite-state machines [B. & Leucker '04]
- automata are expressively equivalent to EMSO logic
  \[ \exists X_1 \ldots \exists X_n : \varphi(X_1, \ldots, X_n) \] with \( \varphi \) a first-order formula

Model checking

- \( \mathcal{A} \models \varphi \) is decidable in linear time (for fixed \( \varphi, f \)) [Seese '96]
- algorithm computes \( t^A \) in linear time and checks \( t^A \in T_\varphi \)
Hanf normal form

Applications of Hanf normal forms

The construction

Optimality
Let

\[(\neg)\exists x_k \ldots (\neg)\exists x_1 : \varphi(\overline{x}, x_1, \ldots, x_k)\]

be a formula in prenex normal form. We want to construct, “fast”, an $f$-equivalent formula $\psi(\overline{x})$ in Hanf normal form. We proceed by induction.

**Base step:** Let $T$ be the set of all 0-spheres $\tau$ of degree $\leq f$ with $|\overline{x}| + k + 1$ centers $\overline{c}, c_1, \ldots, c_k, d$ such that

\[\forall n \quad \tau \models \varphi(\overline{c}, c_1, \ldots, c_k) .\]

Then, $\varphi$ and

\[\bigvee_{\tau \in T} \exists \geq 1 y : \text{sph}_0(\overline{x}, x_1, \ldots, x_k, y) \equiv \tau\]

are $f$-equivalent.

**Induction step:** negation is trivial
**Inductive step:**

\[ \Phi = \exists x \varphi(\overline{x}, x) \quad \text{HNF} \]

\[ \Rightarrow \quad \psi(\overline{x}) = \bigvee_{\tau} (\varphi' \land \exists \geq 1 x : \text{sph}_e(\overline{x}, x) \cong \tau) \quad \text{HNF} \]

\[ \exists^m y \text{sph}_{d'}(\overline{x}, x, y) \cong \sigma \quad f\text{-equivalent to } \Phi \]

\[ \sigma \text{ has } n + 2 \text{ centers; } d' \leq d \]

\[ \tau \text{ ranging over all } e = (3d + 1)\text{-spheres of degree } \leq f \text{ with } n + 1 \text{ centers} \]

We obtain \( \varphi' \) from \( \varphi \) by replacing

\[ \alpha = \exists^m y \text{sph}_{d'}(\overline{x}, x, y) \cong \sigma \text{ by } \alpha' : \]

**Case 1: dist}_\sigma(x, y) \leq 2d' + 1 \]

\[ \alpha' = \begin{cases} \text{true} & \text{if } p \geq m \\ \text{false} & \text{else} \end{cases} \]

where \( p = |\{y \in S_{2d'+1}(x) \mid S_{d'}(\overline{x}, x, y) \cong \sigma\}| \)
Inductive step:

\[ \Phi = \exists x \varphi(\overline{x}, x) \quad \Rightarrow \quad \psi(\overline{x}) = \bigvee_\tau (\varphi' \land \exists \geq 1 x : \text{sph}_e(\overline{x}, x) \cong \tau) \]

- HNF
- HNF

\[ \exists \geq m y \text{sph}_{d'}(\overline{x}, x, y) \cong \sigma \]
- \( \sigma \) has \( n + 2 \) centers; \( d' \leq d \)
- \( \tau \) ranging over all \( e = (3d + 1) \)-spheres of degree \( \leq f \) with \( n + 1 \) centers

We obtain \( \varphi' \) from \( \varphi \) by replacing \( \alpha = \exists \geq m y \text{sph}_{d'}(\overline{x}, x, y) \cong \sigma \) by \( \alpha' \):

Case 2: \( \text{dist}_\sigma(x, y) = \infty \)

\[ \alpha' = \begin{cases} 
\text{false} & \text{if } \not\subseteq S^\sigma_{d'}(\overline{x}, x) \\
\exists \geq m + p y \text{sph}_{d'}(\overline{x}, y) \cong S^\sigma_{d'}(\overline{x}, y) & \text{else}
\end{cases} \]

where \( p = |\{ y \in S^\tau_{2d' + 1}(x) | \cong S^\sigma_{d'}(\overline{x}, y) \}| \)
Size and computation time:

$$\psi(\overline{x}) = \bigvee_{\tau \leq 2^{n^{O(1)} \cdot f^{O(d)}}} \left( \varphi' \leq |\varphi| \cdot f^{O(d)} \wedge \exists \geq 1 \ x \ \text{sph}_{3d+1}(\overline{x}, x) \approx \tau \right)$$

$$|\psi(\overline{x})| \leq |\varphi| \cdot 2^{n^{O(1)} \cdot f^{O(d)}}$$

Radius $d$ is exponential in the size of the original formula.
Theorem
From a formula $\varphi$ and $f > 1$, one can construct an $f$-equivalent formula $\psi$ in Hanf normal form of triply exponential size. This construction can be carried out in time

$$2^{f^{2^O(|\varphi|)}}.$$
Hanf normal form

Applications of Hanf normal forms

The construction

Optimality
Theorem
There is a family of sentences \((\varphi_n)_{n \in \mathbb{N}}\) with \(|\varphi_n| \in O(n)\) such that every 3-equivalent formula \(\psi_n\) in HNF has \(\geq 2^{2^{2^n+1} - 1}\) subformulas of the form \(\exists \geq m x : \text{spH}_d(x) \approx \sigma\), so \(|\psi_n| \geq 2^{2^{2^n+1} - 1}\).

Proof:
Uses forests \(\mathcal{A}\) consisting of binary ordered trees with a unary predicate.

[Frick & Grohe ’04]: there are sentences \(\varphi_n\) of size \(O(n)\) such that \(\mathcal{A} \models \varphi_n\) if and only if

no two complete binary trees of height \(2^n\) in \(\mathcal{A}\) are isomorphic.
Let $\psi_n$ be in HNF, 3-equivalent to $\varphi_n$, and $|\psi_n| < 2^{2^{2^n}+1} - 1$.

There is one complete binary tree $B$ with root $r$ and of height $2^n$ such that $(B, r)$ does not occur in $\psi_n$.

Let $M \in \mathbb{N}$ be maximal such that $\exists^{\geq M}$ appears in $\psi_n$.

$$A_0 = \left( \bigcup_{1 \leq d \leq 2^n} S_d^B(u) \uplus \ldots \uplus S_d^B(u) \right) \setminus \{B\}$$

Then, $A_0$ does not contain any complete binary tree of height $2^n$, so $A_0 \models \psi_n$.

Let $A_2 = A_0 \uplus B \uplus B$. Then, $A_2$ realises the same spheres mentioned in $\psi_n$ as $A_0$, the same number of times (up to $M$). Thus, $A_2 \models \psi_n$. But $A_2 \not\models \varphi_n$, a contradiction. \qed
Summary

Theorem
From a formula $\varphi$ and $f > 1$, one can construct an $f$-equivalent formula $\psi$ in Hanf normal form of triply exponential size in triply exponential time – and this is optimal.