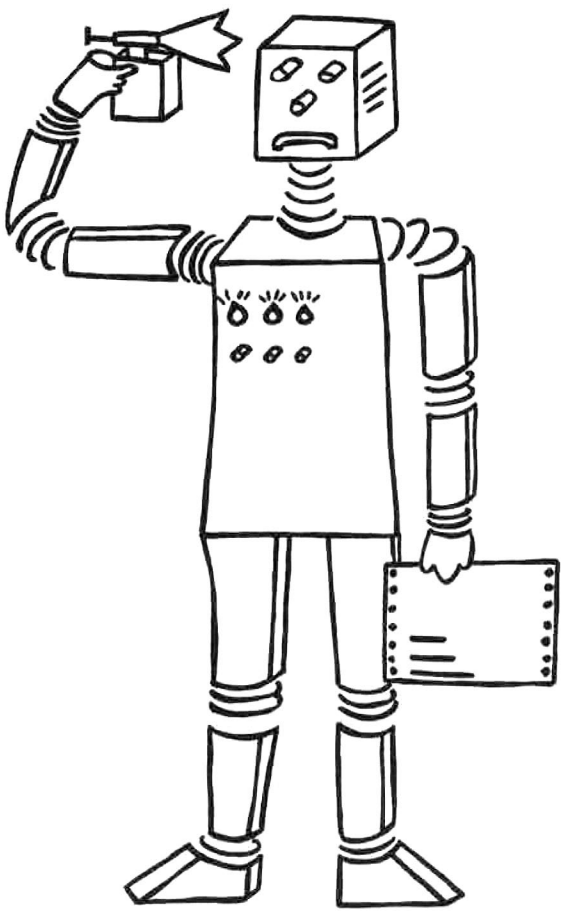
Fachbereich Informatik
Universität Kaiserslautern
Postfach 3049
D-6750 Kaiserslautern 1, W. Germany

Artificial
Intelligence
Laboratories

SEKI-REPORT



The Undecidability of the D_A-Unification Problem

J. Siekmann P. Szabo

December1986 SEKI-REPORT SR-86-19

THE UNDECIDABILITY OF THE D_A-UNIFICATION PROBLEM

J. Siekmann P. Szabó
Universität Kaiserslautern
Fachbereich Informatik
Postfach 3047
6750 Kaiserslautern
WEST GERMANY

ABSTRACT:

We show that the D_A-Unification problem is undecidable. That is, given two binary function symbols $\oplus$ and $\otimes$, variables and constants, it is undecidable if two terms built from these symbols can be unified provided the following D_A-axioms hold:

$$(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$$
$$x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$$
$$x \oplus (y \oplus z) = (x \oplus y) \oplus z$$

Two terms are D_A-unifiable (i.e. an equation is solvable in D_A) if there exist terms to be substituted for their variables such that the resulting terms are equal in the equational theory D_A.

This is the smallest currently known axiomatic subset of Hilbert's Tenth Problem for which an undecidability result has been obtained.

1. UNIFICATION IN EQUATIONAL THEORIES

Not least because of its applications in Computer Science and Artificial Intelligence ***unification theory*** is currently under intense development [Ki 87] and provides some of the theoretical background for a new generation of computing machinery.

The field is concerned with problems of the following kind: Let f and g be function symbols, a and b constants, and let x and y be variables and consider the two ***first order terms*** s and t built from these symbols:

$$s = f(x\ g(a\ b))$$
$$t = f(g(y\ b)\ x)$$

The problem is whether or not there exist terms which can be substituted for the variables x and y such that the two terms thus obtained from s and t become equal: in the example g(a b) and a are two such terms. We shall write

$$\delta = \{x \leftarrow g(a\ b),\ y \leftarrow a\}$$

for such a unifying substitution: δ is a ***unifier*** of s and t since $\delta\, s = \delta\, t$.

In addition to the *decision problem* above there is also the problem of finding a ***unification algorithm*** which enumerates the unifiers for a given pair s and t. Such algorithms are at the very heart of present day computing, in fact they are the central processing unit of the "Fifth Generation Computers" and for efficiency reasons often implemented in silicon.

Consider a variation of the problem above, which arises when we assume that f is commutative:

(C) $\qquad f(x\ y) = f(y\ x)$

Now δ is still a unifying substitution and moreover $\sigma = \{y \leftarrow a\}$ is also a unifier for s and t since

$$\sigma\, s = f(x\ g(a\ b)) =_C f(g(a\ b)\ x) = \sigma\, t$$

But σ is ***more general*** than δ, since δ is an instance of σ obtained as the ***composition*** $\lambda \circ \sigma$ with $\lambda = \{x \leftarrow g(a\ b)\}$; hence a unification algorithm only needs to compute σ.
In some cases there is a single and essentially unique least upper bound on the generality lattice of unifiers, called the ***most general unifier.***

Under commutativity however, there are pairs of terms which have more than one most general unifier, but they always have at most *finitely many*. This is in contrast for example to the above situation of free terms, where every pair has at most *one* most general unifying substitution.
The problem becomes entirely different when we assume that the function denoted by f is associative:

(A) $f(x\ f(y\ z)) = f(f(x\ y)\ z)$

In that case δ is still a unifying substitution, but

$$\tau = \{x \leftarrow f(g(a\ b)\ g(a\ b)),\ y \leftarrow a\}$$

is also a unifier:

$$\tau s = f(f(g(a\ b)\ g(\ a\ b))\ g(a\ b)) =_A f(g(a\ b)\ f(g(a\ b)\ g(a\ b))) = \tau t\ .$$

But

$$\tau' = \{x \leftarrow f(g(a\ b)\ f(g(a\ b)\ g(a\ b))),\ y \leftarrow a\}$$

is again a unifying substitution and by iteration of this process it is not difficult to see that there are *infinitely* many unifiers, all of which are most general.

Finally, if we assume that both axioms (A) and (C) hold for f then the situation changes yet again and for any pair of terms there are at most *finitely* many most general unifiers under (A) and (C).

These and other examples of equational theories motivated the two central notions of unification theory: the ***set of most general unifiers*** μU of two terms and the ***unification hierarchy*** defined on the cardinality of this set:

(i) a theory T is ***unitary***	if μU always exists and has at most one element for every pair of terms;
(ii) a theory T is ***finitary***	if μU always exists and is finite for every pair of terms;
(iii) a theory T is ***infinitary***	if μU always exists and μU is infinite for at least one pair of terms;
(iv) a theory T is ***nullary***	otherwise.

2. THE D_A-UNIFICATION PROBLEM

Given two binary function symbols $\oplus$ and $\otimes$, a single constant **1** and a denumerable set of Variables $\mathbb{V}$ let $\mathbb{T}1$ be the set of terms built from these symbols (for convenience in infix form). When terms in T1 are interpreted by the following D_A-axioms

$$D_A: \quad \begin{aligned} &((x \oplus y) \otimes z) = ((x \otimes z) \oplus (y \otimes z)) \\ &(x \otimes (y \oplus z)) = ((x \otimes y) \oplus (x \otimes z)) \\ &(x \oplus (y \oplus z)) = ((x \oplus y) \oplus z) \end{aligned}$$

we have the equationally defined theory D_A over the ***signature*** $(\oplus,\otimes,1,\mathbb{V})$.
A ***D_A -unification problem*** is a pair of terms for which a D_A-unifier is to be found,written as:

$$<s = t>_{DA} \qquad \text{for } s, t \in \mathbb{T}1$$

and a D_A-unifier is a substitution σ such that D_A equationally implies $\sigma s = \sigma t$,which we shall abbreviate to $\sigma s =_{DA} \sigma t$ in the sequel.We are interested in the existence of a decision procedure that will answer yes if the problem above has a D_A-unifier and no otherwise.

Hilbert's Tenth Problem [Hi19] asks for the existence of a decision procedure, which will tell for a given polynomial Diophantine equation with integer coefficients, whether or not it has a solution in positive integers. A polynomial equation in n unknowns of degree m may be represented as

$$(2.1) \qquad \sum a_i x_1^{i1} \cdot x_2^{i2} \cdot x_3^{i3} \cdot \ldots \cdot x_n^{in} = \sum a_i x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in} \quad \text{for integers } a_i \geq 0 \text{ and } ik \leq m.$$

Y. Matiyasevic showed that there is no decision procedure [Ma 70]; M. Davis gives a lucid account in [Da 73].
Let H10 be the set of axioms containing all necessary axioms of number theory that interpret $\otimes$ as integer multiplication and $\oplus$ as integer addition (and note: D_A is a subset of H10). Then:

$$(2.2) \qquad <s = t>_{H10}$$

phrases Hilbert's problem as a unification problem, where s and t are terms in some suitable extension of T1 such that (2.1) is expressable.
However, we like to represent equations such as (2.1) using only terms in T1 and for technical convenience they should also be in a certain normal form, called "compatible expanded terms", to be formally defined later on.

For that reason let:

step 1: transform every subexpression in (2.1) of the form

$$a_i \cdot x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in}$$

into a_i product terms

$$x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in} + x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in} + \ldots + x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in}$$

step 2: transform every product term

$$x_1^{i1} \cdot x_2^{i2} \cdot \ldots \cdot x_n^{in}$$

into

$$\underbrace{1 \cdot 1 \cdot \ldots \cdot 1}_{M-\Sigma i_k} \cdot \underbrace{x_1 \cdot x_1 \cdot \ldots \cdot x_1}_{i_1} \cdot \underbrace{x_2 \cdot x_2 \cdot \ldots \cdot x_2}_{i_2} \cdot \ldots \cdot \underbrace{x_n \cdot x_n \cdot \ldots \cdot x_n}_{i_n}$$

where M is the maximal sum of the i_k in all product terms.

step 3: replace $\cdot$ by $\otimes$ and $+$ by $\oplus$ and insert appropriate brackets.

Hence every diophantine equation (2.1) can be represented as an H10-unification problem

(2.3) $\langle s = t \rangle_{H10}$ with compatible expanded terms s,t in $\mathbb{T}1$.

Example: Let $x^2y + 2z = 3y^2$ be an instance of (2.1), then M = 3.

step1: $x^2y + z + z = y^2 + y^2 + y^2$

step2: $x \cdot x \cdot y + 1 \cdot 1 \cdot z + 1 \cdot 1 \cdot z = 1 \cdot y \cdot y + 1 \cdot y \cdot y + 1 \cdot y \cdot y$

step3: $(x \otimes (x \otimes y)) \oplus (1 \otimes (1 \otimes z)) \oplus (1 \otimes (1 \otimes z)) = (1 \otimes (y \otimes y)) \oplus (1 \otimes (y \otimes y)) \oplus (1 \otimes (y \otimes y))$

Obviously these syntactical manipulations do not affect solvability under the interpretation of the axioms in H10, i.e. (2.1.) is solvable iff (2.3.) is.
We shall show the undecidability of the D_A-unification problem by coding it into Hilbert's Tenth Problem; i.e. the trick is that in a certain sense the problem is coded recursively into itself.

3. DEFINITIONS AND NOTATION

To give a more precise meaning to the previous notions we shall repeat some of the terminology of unification theory and term rewriting systems.
Unification theory rests upon the basic notions of universal algebra (see e.g.[Gr 79], [BS 81])with the familiar concept of an algebra $\mathbf{A} = (A, F)$ where A is the ***carrier*** and F is a family of ***operators*** given with their arities, usually called the ***signature*** of $\mathbf{A}$.

Assuming that there is at least one constant (operator of arity 0) in F and a denumerable set of variables $\mathbb{V}$, we define T, the set of first order terms, over F and $\mathbb{V}$, as the least set with (i) $\mathbb{V} \subseteq T$, and if arity(f) = 0 for $f \in F$ then $f \in T$ and (ii) if $t_1,...,t_n \in T$ and arity(f) = n then $f(t_1...t_n) \in T$. Let $\mathbb{V}(s)$ be the set of variables occurring in a term s.
Let $\mathbf{T}$ denote the algebra with carrier T and its operators are the term constructors corresponding to each operator of F. $\mathbf{T}$ is called the absolutely free (term) algebra, i.e. it just gives an algebraic structure to T and we shall drop the distinction between T and $\mathbf{T}$ in the sequel.
A ***substitution*** $\sigma: T \to T$ is an endomorphism on T, which is identical almost everywhere on $\mathbb{V}$ and hence can be represented as a finite set of pairs $\sigma = \{x_1 \leftarrow t_1,...,x_n \leftarrow t_n\}$. The application of a substitution σ to a term t is written as σt. The composition of substitutions is defined as the usual composition of mappings: $(\sigma \circ \tau)t = \sigma(\tau t)$ for $t \in T$.
An ***equation*** $s = t$ is a pair of terms. For a set of equations E, the ***equational theory T presented by*** E (in short: the equational theory T) is defined as the finest congruence $=_T$ on T containing all pairs $\sigma s = \sigma t$ for $s = t$ in E and all substitutions σ (i.e. it is the substitution-invariant congruence relation generated by E).
Two terms s,t are ***T-equal*** if $s =_T t$, i.e.E equationally implies $s = t$ .We extend T-equality in T to the set of substitutions by:

$$\sigma =_T \tau \qquad \text{iff} \quad \forall x \in \mathbb{V} \;\; \sigma x =_T \tau x.$$

A substitution σ is ***more general*** than τ (or τ ***is a T-instance of*** σ):

$$\tau \geq_T \sigma \quad \text{iff } \exists\, \lambda \text{ such that } \tau =_T \lambda\sigma .$$

Given two terms s,t and an equational theory T we say a ***unification problem*** $\langle s = t \rangle_T$ is T-unifiable iff there exists a substitution σ such that $\sigma s =_T \sigma t$ and σ is called a ***T-unifier*** of s and t. For the set of all T-unifiers of s and t we write $U_T(s,t)$.
For a given unification problem $\langle s = t \rangle_T$, it is unnecessary to compute the whole set of unifiers $U_T(s,t)$, which is always recursively enumerable for a decidable theory T, but rather a smaller set useful in representing U_T. Therefore $cU_T(s,t)$, the ***complete set of unifiers of s and t*** , is

defined as:

(i) $cU_T \subseteq U_T$ (correctness)

(ii) $\forall\, \delta \in U_T \;\; \exists\, \sigma \in cU_T\colon \;\; \delta \geq_T \sigma$ (completeness)

The ***set of most general unifiers*** $\mu U_T(s,t)$ is defined as a set with (i), (ii) and

(iii) $\forall \sigma, \tau \in \mu U_T\colon \;\; \sigma \geq_T \tau$ implies $\sigma = \tau$ (minimality).

The set μU_T does not always exist [FH 86], [Sc 86], [Ba 86] ; if it does then it is unique up to equivalence, see [FH 86].

Although unification theory is not necessarily restricted to equationally defined theories, most results have been obtained within this frame. The reason is that equational theories can often be directed and used for a computational treatment of $=_T$.

Suppose the equational theory is presented as $T = \{l_1 = r_1, l_2 = r_2, \ldots, l_n = r_n\}$. A term s is said to be ***demodulated*** to t , if there is a subterm s' in s and a pair $l_i = r_i$ in T such that $s' = \mu l_i$ (or $s' = \mu r_i$) for some substitution μ and term t is obtained from s by replacement of s' by μr_i (by μl_i) [WR 67].

The idea of demodulation has been taken further by D.Knuth [KB 70], who observed that it is often possible to find an equivalent and canonical set of equations, called ***a term rewriting system*** (TRS), which is directed from left to right $R = \{l_1 \Rightarrow r_1, l_2 \Rightarrow r_2, \ldots, l_n \Rightarrow r_n\}$ with $V(r_i) \subseteq V(l_i)$. Two terms are in the rewrite relation $s \rightarrow_R t$ iff s can be demodulated to t using one rule in R only from left to right.

R is equivalent to T if for all s,t : $s =_T t$ iff $s \overset{*}{\rightarrow}_R t$,where $\overset{*}{\rightarrow}_R$ is the reflexive and transitive closure of $\rightarrow_R$. If there are no infinite sequences $s_1 \rightarrow_R s_2 \rightarrow_R s_3 \rightarrow_R \ldots$ the relation $\rightarrow_R$ is said to be finitely terminating or ***Noetherian.*** The relation $\rightarrow_R$ is called ***confluent*** if for every r, s, t with $r \rightarrow_R s$ and $r \rightarrow_R t$ there exists a term u such that $s \overset{*}{\rightarrow}_R u$ and $t \overset{*}{\rightarrow}_R u$. A confluent and Noetherian relation (a TRS) is called ***canonical.*** Canonical TRS are an important basis for a computational treatment of equational logic, since they define a unique normal form for every term .

A ***critical pair*** is a pair of terms $\langle s,t \rangle$ computed by superposition in the KB-algorithm (see [KB 70]).This algorithm is used to compute a canonical TRS for a given equational theory T.

It is often useful to partition a given equational theory T into two sets of equations R and E, such that $T = R \cup E$ and only R has a canonical TRS.This is justified by the following theorem due to G.Huet [Hu 80],which we shall use in the next section:

3.Theorem 1: Let $T = R \cup e$ be an equational theory such that

(i) for all $(l_i \Rightarrow r_i) \in R$: $\mathbb{V}(r_i) \subseteq \mathbb{V}(l_i)$ and l_i is linear.

(ii) for all $l_i = r_i \in E$: $\mathbb{V}(l_i) = \mathbb{V}(r_i)$

(iii) the composition of $\rightarrow_R$ and $=_E$, $\rightarrow_R \circ =_E$, is Noetherian

(iv) for all critical pairs $\langle p,q \rangle$ we have $\|p\| =_E \|q\|$

Then for all terms s,t :

$$s =_T t \text{ iff } \|s\| =_E \|t\|$$

A term is linear if every variable occurs only once and $\|p\|$ denotes an irreducible term: $p \overset{*}{\rightarrow}_R \|p\|$.

The fields of term rewriting systems, unification theory and their applications are surveyed in [HO 80], [Bu 85] and in [Si 86].

4. EXPANSION AND ADAPTATION

This section contains some technical preliminaries useful for the proof of the main result as presented in the next section.

A term in $\mathbb{T}$ that contains no $\oplus$-symbol but at least one $\otimes$-symbol is called a ***product term***.. Because of the associativity of $\oplus$ we often omit the brackets around $(s \oplus t)$, but note that we do not have an associativity axiom for $\otimes$.[(i)] A term t is called an ***expanded term*** if it is of the form $t = t_1 \oplus t_2 \oplus \ldots \oplus t_n$ where each t_i is a product term.

4. Proposition 2: For every term $t \in \mathbb{T}1$ there exists an expanded term t' with $t =_{DA} t'$.

We can compute some of the expanded terms of a given term using the distributivity axioms only from left to right:

$$((x \oplus y) \otimes z) \Rightarrow ((x \otimes z) \oplus (y \otimes z))$$

(D)

$$(x \otimes (y \oplus z)) \Rightarrow ((x \otimes y) \oplus (x \otimes z))$$

Let $\rightarrow_D$ be the corresponding rewrite relation and define :

$$EXP(t) = \{ t' \mid t \overset{*}{\rightarrow}_D t' \text{ and } t' \text{ is irreducible under D} \}$$

A term is irreducible if no rewrite rule can be applied to it. Obviously we have :

4.Proposition 3: For all $t' \in EXP(t)$: $t' =_{DA} t$

4.Proposition 4: $EXP(t) \subseteq \{ t' \mid t' =_{DA} t \text{ and } t' \text{ is an expanded term} \}$

The set of expanded terms EXP(t) may contain more than one element, however they all have the same number of $\oplus$-symbols. This observation,which we shall prove now in a more general setting, is the crucial technical prerequisite for the encoding in section 5.

(i) The main result of this paper would still hold if we had; it would be technically even a little simpler to present. However, as we shall discuss later on, the rule of the game is to find a minimal set of axioms.

Let $D_{AC} = D_A \cup \{x \oplus y = y \oplus x\}$ then we have

4.Theorem 5: For $s,t \in \mathbb{T}$ let $s' \in EXP(s)$, $t' \in EXP(t)$:

$$s =_{DAC} t \quad \text{iff} \quad s' =_{AC} t'$$

Proof: The proof is based on 3.Theorem 1 by letting R of 3.Theorem 1 be the two rewrites D and setting E to AC = $\{(x \oplus y) \oplus z = x \oplus (y \oplus z),\ x \oplus y = y \oplus x\}$. The proof consists just of showing that the hypothesis of 3.Theorem 1 is fulfilled.

(i) Obviously $\mathbb{V}(r) \subseteq \mathbb{V}(l)$ for every $(l \Rightarrow r) \in D$ and l is linear (i.e. D is leftlinear).

(ii) Obviously $\mathbb{V}(l) = \mathbb{V}(r)$ for every $(l = r) \in E$.

(iii) Termination of $\rightarrow_R \circ =_E$ can be shown by the following polynomial mapping:

$|t| = 2$ if t is a variable or a constant

$|s \oplus t| = |s| + |t| + 1$

$|s \otimes t| = |s| \cdot |t|$

Then we have:

$$|(r \oplus s) \otimes t| = |r| \cdot |t| + |s| \cdot |t| + |t|$$

$$|(r \otimes t) \oplus (s \otimes t)| = |r| \cdot |t| + |s| \cdot |t| + 1$$

Since $|t| \geq 2$ we have the required orientation for $((x \oplus y) \otimes z) \Rightarrow ((x \otimes z) \oplus (y \otimes z))$ and similarily for $(x \otimes (y \oplus z)) \Rightarrow ((x \otimes y) \oplus (x \otimes z))$ in D.

Since $s =_{AC} t$ implies $|s| = |t|$ we conclude that $\rightarrow_R \circ =_E$ is Noetherian.

(iv) We have to consider the critical pairs between the following rules and equations:

D: (1) $((x \oplus y) \otimes z) \rightarrow ((x \otimes z) \oplus (y \otimes z))$

(2) $(x \otimes (y \oplus z)) \rightarrow ((x \otimes y) \oplus (x \otimes z))$

AC: (3) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$

(4) $x \oplus y = y \oplus x$

Superposition of (1) with (2) and (2) with (1):

$$\begin{array}{ccc} & (x \oplus y) \otimes (u \oplus v) & \\ & \downarrow (2) \qquad\qquad \downarrow (1) & \\ ((x \oplus y) \otimes u) \oplus ((x \oplus y) \otimes v) & & (x \otimes (u \oplus v)) \oplus (y \otimes (u \oplus v)) \\ \downarrow (1) & & \downarrow (2) \\ ((x \otimes u) \oplus (y \otimes u)) \oplus ((x \otimes v) \oplus (y \otimes v)) & =_{AC} & ((x \otimes u) \oplus (x \otimes v)) \oplus ((y \otimes u) \oplus (y \otimes v)) \end{array}$$

Superposition of (1) with LHS of (3) :

$$\begin{array}{ccc} & (x \oplus (y \oplus z)) \otimes u & \\ & \downarrow(1) \qquad\qquad \downarrow(3) & \\ (x \otimes u) \oplus ((y \oplus z) \otimes u) & & ((x \oplus y) \oplus z) \otimes u \\ \downarrow(1) & & \downarrow(1) \\ (x \otimes u) \oplus ((y \otimes u) \oplus (z \otimes u)) & =_{AC} & ((x \otimes u) \oplus (y \otimes u)) \oplus (z \otimes u) \end{array}$$

Superposition of (1) with RHS of (3) and of (2) with (3) are the same.
Superposition of (1) with LHS of (4):

$$\begin{array}{ccc} (x \oplus y) \otimes z & & \\ | \qquad \downarrow(4) & & \\ \downarrow(1) \quad (y \oplus x) \otimes z & & \\ \qquad\qquad \downarrow(1) & & \\ (x \otimes z) \oplus (y \otimes z) & =_{AC} & (y \otimes z) \oplus (x \otimes z) \end{array}$$

Superposition of (1) with RHS of (4) and of (2) with (3) and (4) are the same.
Hence all hypothesis of 3.Theorem 1. are fulfilled and we conclude:

$$s =_{DAC} t \quad \text{iff} \quad s' =_{AC} t'$$

■

Since $s =_{DA} t$ implies $s =_{DAC} t$ we have:

4.Corrollary 6: For $s' \in EXP(s)$ and $t' \in EXP(t)$:

$$\text{If } s =_{DA} t \text{ then } s' =_{AC} t'$$

For a term t let $|t|_{\oplus} : T1 \to \mathbb{N}$ be the number of occurrences of the $\oplus$–symbol in t and let $|t|$ be the number of (all) symbols in t. Now since two terms that are equal under AC must have the same number of $\oplus$–symbols, we finally conclude:

4. Theorem 7: For all $s' \in EXP(s)$, $t' \in EXP(t)$:

$$\text{If } s =_{DA} t \text{ then } |s'|_{\oplus} = |t'|_{\oplus}$$

For technical reasons we define the set $EXT1 \subset T1$ as the set of expanded terms with right associative product terms of equal length, i.e. $t \in EXT1$ if $t = t_1 \oplus t_2 \oplus ... \oplus t_n$ such that the t_i are product terms with $|t_i| = |t_k|$, $1 \le i,k \le n$ and each t_i is associated to the right.
Two expanded terms $s, t \in EXT1$ are called ***compatible*** if $s \oplus t \in EXT1$, i.e. all their respective product terms are of equal length. A ***term is a sum*** if it does not contain a product term, a ***substitution σ is a sum*** if for all $x \in \mathbb{V}$: $\sigma(x)$ is a sum.
If σ is a sum and t an expanded term in EXT1 then the product terms of $EXP(\sigma(t))$ are in EXT1 and they are all of the same length.

For example let $s = ((x \otimes y) \oplus (z \otimes 1))$ and $\sigma = \{x \leftarrow (1 \oplus v)\}$
then $\sigma(s) = ((1 \oplus v) \otimes y) \oplus (z \otimes 1) = (1 \otimes y) \oplus (v \otimes y) \oplus (z \otimes 1)$.

As an abbreviation we use the notation

$$\bigoplus_{1 \le i \le n} t_i = t_1 \oplus t_2 \oplus ... \oplus t_n \qquad \text{for expanded term}$$

4. Proposition 8: Let $s = \bigoplus_{1 \le i \le n} s_i \in EXT1$ and let σ be a sum.
Then for all $s' \in EXP(\sigma(s))$ with $s' = \bigoplus_{1 \le i \le m} s_i'$:

$$|s_i'| = |s_j'| = |s_k| \quad \text{with } 1 \le i,j \le m,\ 1 \le k \le n \quad \text{and} \quad s' \in EXT1.$$

Proof: The proof is by double induction on n and on q, where q is the number of variables moved by σ. ■

A substitution α is a $\oplus$–***adapter*** for terms s and t if there exists $s' =_{DA} \alpha(s)$ and $t' =_{DA} \alpha(t)$ with $|s'|_\oplus = |t'|_\oplus$. In other words, a $\oplus$–adapter is a substitution that introduces enough symbols to make two terms equal with respect to their number of product terms (w.r.t .their number of $\oplus$–symbols).This is a weaker notion than that of a unifier; if two terms are unifiable then there exists an adapter, but not necessarily vice versa:

4.Proposition 9: If two terms are D_A-unifiable then there exists an adapter.

Proof: Use 4.Theorem 7.

■

5. The D_A-Unification Problem is Undecidable

The reasoning in this section is as follows: we show that in the restricted class of terms EXT1, a D_A-unification problem is unifiable iff there exists an adapter and there exists an adapter iff there exists a solution to Hilbert's Tenth Problem.

5. Lemma 9: For compatible terms $s,t \in$ EXT1:

There exists an adapter for s and t iff $\langle s=t\rangle_{DA}$ is unifiable.

Proof: If $\langle s=t\rangle_{DA}$ is unifiable with σ then $\sigma(s) =_{DA} \sigma(t)$ and hence by 4. Proposition 9. there is an adapter.

If there exists an adapter α for $\langle s=t\rangle_{DA}$ it can be represented using 4.Prposition 2 as:

$$\alpha = \left\{ x_1 \to \bigoplus_{1\le j\le k1} r_{1j},\ x_2 \to \bigoplus_{1\le j\le k2} r_{2j},\ \ldots,\ x_n \to \bigoplus_{1\le j\le kn} r_{nj} \right\}$$

such that $Var(s,t) = \{x_1,\ldots,x_n\}$ and the terms r_{ij} are product terms, i.e.the terms to be substituted are expanded. But then

$$\sigma = \left\{ x_1 \mapsto \bigoplus_{1\le j\le k1} \mathbf{1},\ x_2 \mapsto \bigoplus_{1\le j\le k2} \mathbf{1},\ \ldots,\ x_n \mapsto \bigoplus_{1\le j\le kn} \mathbf{1} \right\}$$

where $\mathbf{1}$ is the single constant of the D_A-signature is also an adapter for s and t, since it substitutes the same number of $\oplus$-symbols.

But now it is not difficult to see that σ is a unifier: Since s and t are compatible terms, let

$$s=\bigoplus_{1\le i\le m} s_i \quad \text{and} \quad t=\bigoplus_{1\le i\le n} t_i \quad \text{with} \quad |s_i| = |t_i| \quad \text{for all } i.$$

Let $s' \in EXP(\sigma(s))$ and $t' \in EXP(\sigma(t))$. Since σ is an adapter, we have $s'=\bigoplus_{1\le i\le k} s_i'$ and $t'=\bigoplus_{1\le i\le k} t_i'$ for some fixed $k \ge m,n$.

By 4. Proposition 8 we have, since σ is a sum: $|s_i'| = |t_i'|$, $s_i', t_i' \in$ EXT1 and each s_i', t_i' consists just of $\mathbf{1}$'s with brackets associated to the right.

Hence $\sigma(s) =_{DA} s'$ for $s' \in EXP(\sigma(s))$ by 4.Lemma1

$= t'$ since $s_i' = t_i'$ and each s_i', t_i' consists just of $\mathbf{1}$'s in right associative form

$=_{DA} \sigma(t)$ for $t' \in EXP(t)$ by 4.Lemma 1

■

5. *Proposition 10:* Given a product term t with $\mathbb{V}(t) = \{x_1,...,x_n\}$ and a substitution

$$\sigma = \left\{ x_1 \to \bigoplus_{1\le j\le k1} u_{1j},\ x_2 \to \bigoplus_{1\le j\le k2} u_{2j},\ \dots,\ x_n \to \bigoplus_{1\le j\le kn} u_{nj} \right\}$$

Then for all $t' \in EXP(\sigma(t))$:

$$|t'|_{\oplus} = \prod_{1\le i\le n} (k_i)^{m_i} - 1 \qquad \text{with } m_i = |t|_{x_i},\ 1\le i\le n.$$

Example: Let $t = (a \otimes (x_1 \otimes (x_2 \otimes x_3)))$ be the product term and $\sigma = \{x_1 \to \bigoplus_{1\le i\le 3} u_{1j}\}$.
Then $n = 1$, $k_1 = 3$ and by default $k_2 = 1$
$\sigma t = (a \otimes ((u_{11} \oplus u_{12} \oplus u_{13}) \otimes ((u_{11} \oplus u_{12} \oplus u_{13}) \otimes x_2)))$.
For $t' \in EXP(\sigma t)$ we obtain
$t' = (a\otimes(u_{11}\otimes(u_{11}\otimes x_2)))\oplus(a\otimes(u_{11}\otimes(u_{12}\otimes x_2)))\oplus....\oplus(a\otimes(u_{13}\otimes(u_{13}\otimes x_2)))$
and $|t'|_{\oplus} = 3^2 \cdot 1^1 - 1 = 8$.

■

For the following we note that the signature of the D_A-problems is a subset of the signature that is necessary to represent Hilbert's H10-problem. But if we restrict the terms in H10 to those in EXT1, we can interpret these terms in D_A as well as in H10, which provides the background for the following lemma:

5. *Lemma 11:* For compatible terms $s,t \in EXT1$: There exists an adapter for $\langle s=t\rangle_{DA}$ iff there exists a diophantine solution for $\langle s=t\rangle_{H10}$.

Proof: Let $s = \bigoplus_{1\le i\le m} s_i$ and $t = \bigoplus_{1\le i\le n} t_i$ where s_i and t_i are product terms with $|s_i| = |t_i|$
Let $\mathbb{V}(s,t) = \{x_1,x_2,...,x_p\}$ be all the variables occurring in s and t.
If there exists an adapter for s and t then there exists an adapter of the form

(i) $$\alpha = \left\{ x_1 \to \bigoplus_{1\le j\le k1} u_{1j},\ x_2 \to \bigoplus_{1\le j\le k2} u_{2j},\ \dots,\ x_p \to \bigoplus_{1\le j\le kp} u_{pj} \right\}$$

with u_{ij} new variables not in $\mathbb{V}(s,t)$.
Now since α is an adapter we have for all $s' \in EXP(\alpha(s))$ and $t' \in EXP(\alpha(t))$:
$|s'|_{\oplus} = |t'|_{\oplus}$.
Setting $u_{ij}=1$ and interpreting $\oplus$ as integer addition and $\otimes$ as integer multiplication in H10, we see that $\{x_1 \to k_1, x_2 \to k_2,....,x_n \to k_n\}$, for $k_i \in \mathbb{N}$ as in (i) above, is in fact a diophantine solution for $\langle s=t\rangle_{H10}$: The number of $\oplus$-symbols is the integer value in H10, since all product terms are of the form $1 \otimes 1 \otimes ... \otimes 1 =_{H10} 1$ in $\mathbb{N}$.

Conversely if $\sigma = \{x_1 \to k_1, \ldots, x_p \to k_p\}$, $k_i \in \mathbb{N}$, is a diophantine solution for $\langle s=t \rangle_{H10}$ then

(ii) $\alpha = \left\{ x_1 \to \bigoplus_{1 \le j \le k1} u_{1j},\ x_2 \to \bigoplus_{1 \le j \le k2} u_{2j}, \ \ldots\ , \ x_p \to \bigoplus_{1 \le j \le kp} u_{pj} \right\}$

is an adapter for s and t, since by 5. Proposition 10, we have

for all $s' \in EXP(\alpha(s))$: $|s'|_{\oplus} = |s|_{\oplus} + \sum_{1 \le j \le m} \left(\prod_{1 \le i \le p} (k_i)^{m_i} - 1 \right)$ with $m_i = |s_j|_{x_i}$

and

for all $t' \in EXP(\alpha(t))$: $|t'|_{\oplus} = |t|_{\oplus} + \sum_{1 \le j \le n} \left(\prod_{1 \le i \le p} (k_i)^{l_i} - 1 \right)$ with $l_i = |t_j|_{x_i}$

With $|s|_{\oplus} = m-1$ and $|t|_{\oplus} = n-1$ we obtain:

$$|s'|_{\oplus} = (m-1) + \sum_{1 \le j \le m} \left(\prod_{1 \le i \le p} (k_i)^{m_i} - 1 \right) \quad \text{for } m_i = |s_j|_{x_i}$$

$$= \text{integer-value-of}(\sigma(s)) + 1 \quad \text{in H10}$$

$$= \text{integer-value-of}(\sigma(t)) + 1 \quad \text{in H10 since } \sigma \text{ is a Diophantine solution}$$

$$= (n-1) + \sum_{1 \le j \le n} \left(\prod_{1 \le i \le p} (k_i)^{l_i} - 1 \right) \quad \text{for } l_i = |t_j|_{x_i}$$

$$= |t'|_{\oplus}$$

■

We are now ready to formulate the main theorem :

5.Theorem 12: The D_A-unification problem for compatible terms in EXT1 is decidable iff Hilbert's Tenth Problem is decidable.

Proof: Given any Diophantine equation of the form (2.1) transform it as in section 2 into $\langle s=t \rangle_{H10}$ with compatible terms $s,t \in$ EXT1.

(1) Suppose the D_A-unification problem for compatible terms is decidable.
Let $\langle s=t \rangle_{H10}$ be any Hilbert Problem with terms in EXT1 and consider $\langle s=t \rangle_{D_A}$: our supposed decision procedure will say yes or no to its solvability.

(1.1) No: i.e. $\langle s=t\rangle_{D_A}$ does not have a solution.
Then by 5. Lemma 9 there does not exist an adapter and using 5. Lemma 11, there does not exist a solution for $\langle s=t\rangle_{H10}$. Hence the Diophantine equation will not have a solution.

(1.2) Yes: i.e. $\langle s=t\rangle_{D_A}$ has a D_A-unifier σ.
Define $\tau=\{u \to 1 \mid \text{for all } u \in \mathbb{V}(\sigma s,\sigma t)\}$. Now since σ is a D_A-unifier, so is $\tau\circ\sigma$. Hence by 5. Lemma 9 there exists an adapter and by 5. Lemma 10 there exists a Diophantine solution for $\langle s=t\rangle_{H10}$.

Thus in either case there would be a decision procedure for Hilbert's Tenth Problem.

(2.) Suppose Hilbert's Tenth Problem is decidable.
Let $\langle s=t\rangle_{D_A}$ be any D_A-unification problem and consider the problem $\langle s=t\rangle_{H10}$. Then we have immediately: if $\langle s=t\rangle_{H10}$ has a (no) solution then there exists a (no) adapter for s and t by 5.Lemma 9 and hence by 5.Lemma 11 $\langle s=t\rangle_{D_A}$ is (is not) unifiable. ■

Suppose now the D_A-unification problem was decidable, then the restricted D_A-unification problem for compatible terms in EXT1 would be decidable too. With the above theorem Hilbert's Tenth Problem would be decidable, hence:

5. Corollary 13: The D_A-unification problem is undecidable.

Closing this section we state our final observation:

5. Theorem 14: Let Ψ be any set of axioms with $D_A \subseteq \Psi \subseteq H10$. Then the Ψ-unification problem

$$\langle s=t\rangle_{\Psi} \text{ for } s,t\in T1$$

is undecidable

which can be shown for any fixed Ψ with the same proof technique as in the previous paragraphs.

6.CONCLUSION

The unification problem in ω-order logics is undecidable for $\omega \geq 2$ [Hu 73], [Go 81] even if no additional axioms are present. It is also well-known that this is not the case for first order logics, where algorithms are known that compute a most general unifier of a given pair of free first order terms if it exists and terminate with failure otherwise [Ro 65]. Such algorithms are a basic cornerstone in most work on computational deduction and provide yet another characterization of the gulf between first and higher order logics.

Although not entirely obvious, it comes without much surprise that it should be possible to find *first order equational* logics such that the associated unification problem is again undecidable, provided there are enough equations to code into. For intellectual curiosity as well as practical interest in applications of unification theory, however, there is the problem to isolate minimal equational theories such that the unification problem in that theory is undecidable.

D. Hilbert's Tenth Problem presented to the International Congress of Mathematicians in his 1900 speech in Paris [Hi 19] poses the question if it is decidable whether a given polynomial equation is solvable in positive integers; a problem that was finally shown to be undecidable. Let H10 be the axioms formalizing number theory, possibly augmented with axioms that are useful in expressing polynomial equations. Hilbert's Tenth Problem is then the H10-unification problem, that requires among others a higher order induction axiom (or infinitely many axioms or...), hence is not equational.
However H10 will contain the distributivity and associativity axioms for integer multiplication and addition. We have shown that it is possible to eliminate all axioms in H10 except the D_A-axioms (and hence return to first order) and still maintain the celebrated undecidability result.

There is the natural question then whether D_A is really the minimal axiomatic subset of H10 such that undecidability can be shown.

The associativity axiom alone (the A-unification problem) was the subject of intensive research: provided an identity element the A-unification problem is the problem of solving equations in a free monoid. This problem became known as the string unification problem [Si 75], [Ja 85] in the field of automated deduction, as Markov's problem [Hm 64] among most Eastern Europeans in semigroup theory and as Löb's problem in the West. Its decidability remained open for almost twenty years: it was finally shown to be decidable [Ma 77]. Hence the associativity axiom on its own is excluded as a possible candidate for a minimal subset of H10 (and thus of D_A).

The race is now open for the (un)decidability of the D-unification problem i.e. the problem whether one or both of the D-axioms in itself are sufficient for an undecidable unification problem. Some first results in this direction are presented in [Sz 82], [AT 85], [Mz 86]. If it is undecidable it is a smallest subset of H10, if it is decidable why is it that the combination of two decidable theories (D) and (A) pose an undecidable D_A-problem?

Acknowledgements: This paper was improved (and the length of it substantially reduced) by technical suggestions and contributions from N.Eisinger, F.Baader, H.J.Bürckert, A.Herold and M.Schmidt-Schauß.
The observation that Huet´s theorem can be used in section 4 (instead of a lengthy technical and direct proof) is due to F.Baader.

7.REFERENCES

[AT 85] A.Arnberg, E.Tiden: Unification Problems with one-sided Distributivity, Proc. of Conf. on Rewriting Techniques, Springer LNCS, vol 202,1985.

[Ba 86] A.Baader: Unification in Varieties of Idempotent Semigroups, Internal Report, Universität Erlangen, Institut für Mathem. und Datenverarbeitung, 1986.

[Bu 85] B.Buchberger: Basic Features and Development of the Critical-Pair Completion Procedure, Proc. of the First Conf. on Rewriting Techniques,Springer LNCS, vol 202, 1985.

[BS 81] S.Burris, H.P.Sankappanavar: A Course in Universal Algebra, Springer Verlag, 1981.

[Da 73] M.Davis: Hilbert´s Tenth Problem is unsolvable, Amer.Math.Monthly,vol 80,1973.

[FH 86] F.Fages, G.P.Huet : Complete Sets of Unifiers and Matchers in Equational Theories, J. of Theoretical Comp. Science, 43, pp 189-200, 1986.

[Go 81] D.Golfarb: The Undecidability of the Second Order Unification Problem, J. of.Theor. Comp. Science, 13, pp 225-230, 1981.

[Gr 79] G.Grätzer: Universal Algebra,Springer Verlag, 1979.

[Hi 19] D.Hilbert: Mathematische Probleme; Vortrag gehalten auf dem Internationalen Mathematiker Kongress in Paris, Nachrichten Akad. Wiss. Göttingen,Math.-Phys. KL, 1900, pp253-297.

[Hm 64] J.I.Hmelevskij: The Solution of certain Systems of Word Equations, Dokl. Akad. Nauk SSSR,1964;749 Soviet Math. Dokl. 5,1964,724.

[Hu 73] G.P.Huet: The Undecidability of the Unification Problem in Third Order Logic, Information and Control , vol 22, no. 3, pp 257-267, 1973.

[Hu 80] G.P.Huet: Confluent Reductions: Abstract Properties and Applications to Term Rewriting Systems, JACM, vol 27, no. 4, pp 797-821, 1980.

[HO 80] G.P.Huet, D.C.Oppen : Equations and Rewrite Rules, in: Formal Languages: Perspectives and Open Problems, R.Book (ed), Academic Press 1976.

[Ja 85] J.Jaffar: Minimal and Complete Word Unification, Internal Report, Monash University, Australia, 1985.

[KB 70] D.B.Knuth, P.B.Bendix: Simple Word Problems in Universal Algebras, in: Computational Problems in Abstract Algebra, J.Leech (ed), Pergamon Press, Oxford, 1970.

[Ki 87] J.C.Kirchner (guest editor), Special Issue on Unification Theory, Journal of Symbolic Computation, to appear 1987.

[Ma 77] G.S.Makanin: The Problem of Solvability of Equations in a Free Semigroup, Soviet Acad. Nauk SSSR, Tom 233, no.2, 1977.

[Ma 70] Y.Matiyasevic: Enumerable Sets are Diophantine, Dokl. Akad.Nauk SSSR, 191, 1970, pp 279-282.

[Mz 86] J.Mzali: Matching with Distributivity, Proc. of 8th Conf. on Autom. Deduction, Springer LNCS, vol 230, pp 496-505, 1986.

[Ro 65] J.A.Robinson: A Machine Oriented Logic Basd on the Resolution Principle, JACM, vol 12, 1965.

[Si 86] J.Siekmann: Unification Theory, Proc. of Europ. Conf. on Artificial Intelligence, Brighton,1986.

[Si 75] J.Siekmann: Stringunification, Essex University Memo, 1975.

[Sc 86] M.Schmidt-Schauß: Unification under Associativity and Idempotency is of Type Nullary, J. of Automated Reasoning, vol 2, no.3, pp 277-282, 1986.

[Sz 82] P.Szabo: Theory of First Order Unification (in German), Thesis University of Karlsruhe, W. Germany, 1982.

[WR 67] L.Wos, G.A.Robinson, D.Carson, L.Shalla: The Concept of Demodulation in Automated Theorem Proving, JACM, vol 14, no.4, 1967.