HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports ISSN 0783–5396
No. **39**; August 1996 ISBN 951-22-3227-8

HOW TO CALCULATE SYMBOLICALLY SIPHONS AND TRAPS FOR SOME ALGEBRAIC PETRI NETS

Karsten Schmidt
Digital Systems Laboratory
Department of Computer Science
Helsinki University of Technology
Otaniemi, FINLAND

Helsinki University of Technology
Department of Computer Science
Digital Systems Laboratory
Otaniemi, Otakaari 1
SF–02150 ESPOO, FINLAND

HELSINKI UNIVERSITY OF TECHNOLOGY
DIGITAL SYSTEMS LABORATORY
Series **A:** Research Reports
No. **39**; August 1996

ISSN 0783–5396
ISBN 951-22-3227-8

How to Calculate Symbolically Siphons and Traps for some Algebraic Petri nets

KARSTEN SCHMIDT *

Abstract: We present a symbolic approach to the calculation of siphons and traps for algebraic Petri nets where the color domains are specified without equations and with at most unary operation symbols.

Keywords: petri nets, structural analysis, symbolic analysis.

*supported by a scholarship of German Academic Exchange Service(DAAD)

Printing: TKK Monistamo; Otaniemi 1996

Helsinki University of Technology
Department of Computer Science
Digital Systems Laboratory
Otaniemi, Otakaari 1
SF–02150 ESPOO, FINLAND

Phone: $\frac{90}{+358\text{–}0}$ 4511
Telex: 123 889 hutel sf
E-mail: lab@hutds.hut.fi

1 Introduction

One of the purposes of Petri net theory is to predict the behavior of systems from the structure of their Petri net model. For some net structures — siphons and traps — implications to the system's behavior are already known:

- If every non–empty siphon of a net includes a marked trap[1] then no dead marking is reachable in the net;
- A free choice net is live if and only if every non–empty siphon includes a marked trap.

The latter implication, COMMONER's Theorem ([Com72]) is valid for finite nets only. For algebraic nets with infinite color domains (respectively their unfolding according to the standard semantics) a weakened version of this theorem is valid ([Sch96a]):

- If a free choice net is live, then every non–empty siphon includes a marked trap
- If every non–empty siphon of a free choice net includes a *finite* marked trap then the net is live.

These implications leave a grey zone —- nets where some siphons include infinite marked traps but no finite ones; for this constellation both live and non–live examples are known. Nevertheless in many cases implications to liveness can be applied, if we can verify the property "every non–empty siphon includes a marked trap". For checking this property a calculation of a representation of *all* siphons is necessary; a pure verification of siphons predicted by the user ([Sch96b]) is not sufficient for this purpose (it would be sufficient to prove the negation of this property). Unfortunately this calculation turns out to be computationally hard. We can solve the problem only for a restricted class of nets: those where the algebraic specification of the color domains does not contain equations and where all operation symbols are constant symbols or unary symbols (that is operations which have only one argument). Though these restrictions still allow for the specification of natural numbers and thus for modeling counter automata and synchronizations between several of them, they give also a feeling for the limitations of the symbolic approach to structural analysis, since of course they have not been applied voluntarily. Though we cannot prove that for weaker restrictions a symbolic approach to siphon

[1] Throughout the whole paper we restrict ourselves to ordinary nets, otherwise the concept becomes more involved

and trap calculation is not possible in principle, the amount of argumentation with them gives some evidence that there is only few hope for a significantly more general approach. Even under the restrictions applied here there remains a certain danger that a complete representation of all siphons cannot be computed. For instance there are nets which have more than countable many siphons, while with the proposed symbolic approach only countable many of them can be represented.

After presenting briefly the basic definitions for algebraic specifications and algebraic Petri nets, we discuss a concept for a term based set algebra, which forms the kernel of our approach. Then we motivate the symbolic representation of siphons and traps, ending up in a formal definition. We continue with the general outline of the calculation which mainly consists of two major parts. The first part is to partition a symbolic representation of place sets into siphons and non–siphons, while the second part is to append places to symbolic representations of non–siphons. Both parts are studied in detail. We end up with a brief summary.

2 Basic Definitions

First we recall concepts from the theory of abstract data types, restricted to our purposes. For details, please refer to [EM85].

Definition 1 (Equation–free unary Specifications) *A* **signature** $\Sigma = [S, \Omega]$ *consists of a set* S *of* **sorts** *and a family* $\Omega = \{\Omega_{w,s}\}_{w \in S \dot{\cup} \{\epsilon\}, s \in S}$ *of* **operation symbols**. $\Omega_{\epsilon,s}$ *is the set of* **constant symbols** *of sort* s. *A set of* Σ*–***variables** *is a family* $X = \{X_s\}_{s \in S}$ *of* **variables**. *The set* $T_{\Omega,s}(X)$ *of* (Ω, X)*–***terms of sort** s *is inductively defined by 1.* $X_s \cup \Omega_{\epsilon,s} \subseteq T_{\Omega,s}(X)$ *and 2. for* $\omega \in \Omega_{s_1,s}$ *and* $T_1 \in T_{\Omega,s_1}(X)$, $\omega(T_1) \in T_{\Omega,s}(X)$. *The set* $T_{\Omega,s} := T_{\Omega,s}(\emptyset)$ *contains the* **ground terms of sort** s. *An* **equation–free specification** *is a signature* Σ *where for every sort* s $T_{\Omega,s}$ *is not empty.*

We interpret a sort s with the set of ground terms $T_{\Omega,s}$, a constant symbol to stand for itself as an element of its sort and an operation symbol $\omega \in \Omega_{s_1,s}$ to be a mapping from T_{Ω,s_1} to $T_{\Omega,s}$ assigning the element $\omega(T)$ to an element $T \in T_{\Omega,s_1}$.

Definition 2 (Substitutions) *Let* X *and* Y *be two sets of* Σ*–variables. A* **substitution** *on* X *is an assignment* $\sigma : X \to T_\Omega(Y)$, $(X_s \to T_{\Omega,s}(Y))$. *A* **ground substitution** *is a substitution* $\sigma : X \to T_\Omega$. *For a substitution its domain* X *is denoted by* $\mathrm{dom}(\sigma)$ *and the smallest set* Y *such that* σ *maps to* $T_\Omega(Y)$ *is denoted by* $\mathrm{var}(\sigma)$. *An injective substitution* $\sigma : X \to Y$ *is called* **renaming**. *For a term* T *and a substitution* σ *the term* $\sigma(T)$ *(often* *written as* $T\sigma$*) results from simultaneously replacing the variables in* T *by their corresponding* σ*–values. For a set of* Σ*–variables* $\mathrm{Subst}(X)$ *denotes the set of all substitutions and* $\mathrm{GrSubst}(X)$ *the set of all ground substitutions on* X*.*

Definition 3 (Multisets) *For a set* M*, a* **multiset** *over* M *is a mapping from* M *into the integer numbers. A multiset is* **semipositive** *iff all the values are* ≥ 0*. A multiset is* **finite** *iff it has finite support. The* **empty** *multiset over* M*, denoted by* ϑ_M*, assigns* 0 *to every element of* M*. For an element* $m \in M$*, the multiset* $\underline{m}$ *assigns* 1 *to* m *and* 0 *to every other* $m' \in M$*. The multisets* $\mu_1 + \mu_2$ *and* $\mu_1 - \mu_2$ *are defined by* $(\mu_1 + \mu_2)(m) := \mu_1(m) + \mu_2(m)$ *and* $(\mu_1 - \mu_2)(m) := \mu_1(m) - \mu_2(m)$*. This way every finite multiset can be written as a* **formal sum** *of the* $\underline{m}(m \in M)$*. We usually write* m *instead of* $\underline{m}$*. A multiset* μ_1 *is* **less or equal** *to* μ_2 *iff for all* $m \in M$*,* $\mu_1(m) \leq \mu_2(m)$*.*

A **multiterm** is a multiset over a set of terms.

Definition 4 (Algebraic Net on unary equation–free specification) *A tuple* $AN = [\Sigma; P, T, F; \psi, \xi, \lambda; m_0]$ *is an* **algebraic Petri net** *iff 1.)* $\Sigma = [S, \Omega]$ *is an equation–free specification; 2.)* $[P, T, F]$ *is a* **net***, i.e.* P *and* T *are finite and disjoint sets called* **places** *and* **transitions***, respectively, and* F *is a relation* $F \subseteq (P \times T) \cup (T \times P)$*, the elements of which are called* **arcs***; 3.)* ψ *is a* **sort assignment** $\psi : P \to S$*; 4.)* ξ *assigns a set of* Σ*–variables* $\xi(t)$ *to each transition* $t \in T$*; 5.)* λ *is the* **arc inscription** *such that for* $f = [p, t]$ *or* $f = [t, p]$ *in* F*,* $\lambda(f)$ *is a multiterm over* $T_{\Omega,\psi(p)}(\xi(t))$*; 6.)* m_0 *is a* **marking***, i.e. it assigns a finite multiterm over* $T_{\Omega,\psi(p)}$ *to every* $p \in P$*.* m_0 *is called the* **initial marking***.*

For $f \notin F$ we define $\lambda(f) := \vartheta$. It is possible to interpret an algebraic Petri net according to an arbitrary model of the specification Σ. The result is a colored net. This way all the behavioral aspects of an algebraic net can be traced back to colored nets and from there to place transition nets. According to our syntactic interpretation of the specification the place transition net which underlies an algebraic net is constructed as follows:

- The set of places is $\bigcup_{p \in P} \{p\} \times T_{\Omega,\psi(p)}$. We write a place $[p, T]$ of the underlying Petri net as $p.T$.
- The set of transitions is $\bigcup_{t \in T} \{t\} \times \mathrm{GrSubst}(\xi(t))$. We write a transition $[t, \beta]$ of the underlying Petri net as $t.\beta$.
- The multiplicity of an arc from $p.T$ to $t.\beta$ is given by $(\lambda([p, t])\beta)(T)$. (Note that $\lambda([p, t])\beta$ is a multiset of ground terms.);
- Accordingly the multiplicity of an arc from $t.\beta$ to $p.T$ is $(\lambda([t, p])\beta)(T)$.

- The initial marking of the place $p.T$ is given by $m_0(T)$.

An arc with multiplicity 0 is to be interpreted as the absence of this arc. The underlying net can consist of infinitely many nodes. However, only finitely many places can have an initial marking greater than 0 and every transition can only be connected to finitely many places via a direct arc (with multiplicity greater than 0). On the other hand, places can be connected to infinitely many transitions.

For further details according to algebraic Petri nets, please refer to [Rei91].

Definition 5 (Siphon,Trap) *A subset D of P is a* **siphon** *a net $N = [P, T, F]$ iff $FD \subseteq DF$. D is a* **trap** *iff $DF \subseteq FD$. For an algebraic net, siphons and traps are those of the underlying place transition net.*

3 Term based set algebra

One of the major reasons which make our approach work for unary equation–free specifications is that all basic set operations — union, intersection and complement — can be modeled symbolically. Thereby we consider sets of ground terms which are represented by a set of (not necessarily ground) terms.

Definition 6 (Symbolic representation of term sets) *Let Σ be a signature and $\mathcal{T}$ be a finite set of terms, all contained in $T_{\Omega,s}(X)$ for some family of Σ–variables X. $\mathcal{T}$ is called a* **symbolic set of terms**, *it represents the set* $\mathrm{Ground}(\mathcal{T}) = \{T\rho \mid T \in \mathcal{T}, \rho \in \mathrm{GrSubst}(X)\}$.

The union of symbolically represented sets of terms can be represented by fusing the corresponding symbolic representations:

Lemma 1 *For two symbolic sets of terms $\mathcal{T}_1$ and $\mathcal{T}_2$ it holds*

$$\mathrm{Ground}(\mathcal{T}_1) \cup \mathrm{Ground}(\mathcal{T}_2) = \mathrm{Ground}(\mathcal{T}_1 \cup \mathcal{T}_2)$$

For the intersection of symbolically represented term sets the concept of unification can be applied.

Definition 7 (Unification) *Let T_1 and T_2 be two terms contained in some $T_{\Omega,s}(X)$. A substitution over X is called* **unifier** *for T_1 and T_2 iff $T_1\sigma = T_2\sigma$. σ is called* **most general** *iff for all unifiers σ' of T_1 and T_2 there is a substitution τ such that $\sigma' = \sigma\tau$.*

We write $T_1 \,\mathrm{unif}\, T_2$ if there exists a unifier of T_1 and T_2. It is well known that there exists always a most general unifier which is unique up to variable renaming for two terms if there is a unifier at all. This justifies the notation $MGU(T_1, T_2)$ for a most general unifier of T_1 and T_2.

The symbolic implementation of intersection benefits from the observation that the two substitutions which generate an element T in the intersection of two symbolically represented sets form a unifier of the two terms to which they have been applied. Thereby we assume that different terms do not share common variables, that is for T_1 and T_2 appearing in a symbolic set of terms it is always assumed: $\mathrm{var}(T_1) \cap \mathrm{var}(T_2) = \emptyset$ or $T_1 = T_2$.

Lemma 2 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two symbolic sets of terms where no variable appears both in elements of $\mathcal{T}_1$ and $\mathcal{T}_2$. Then it holds*

$$\begin{aligned}\mathrm{Ground}(\mathcal{T}_1) \cap \mathrm{Ground}(\mathcal{T}_2) = \mathrm{Ground}(\{T_1\sigma \mid \\ T_1 \in \mathcal{T}_1, \exists T_2 : T_2 \in \mathcal{T}_2, T_1 \,\mathrm{unif}\, T_2 \text{ and } \sigma = MGU(T_1, T_2)\})\end{aligned}$$

Proof. Let $T \in \mathrm{Ground}(\mathcal{T}_1) \cap \mathrm{Ground}(\mathcal{T}_2)$. Then there exist terms $T_1 \in \mathcal{T}_1$, $T_2 \in \mathcal{T}_2$ and ground substitutions ρ_1 and ρ_2 such that $T = T_1\rho_1 = T_2\rho_2$. Since the variables appearing in T_1 and T_2 are different, the substitution $\rho = \rho_1 \cup \rho_2$ is well defined and is a unifier of T_1 and T_2. Therefore T_1 and T_2 are unifyable and for $\sigma = MGU(T_1, T_2)$, $T_1\sigma$ is contained in the symbolic representation of the intersection. Since σ is most general and ρ is a unifier, there exists a τ such that $\rho = \sigma\tau$. Therefore $T = T_1\rho = T_1\sigma\tau$ is contained in the set on the right hand of the equation.

Let T be an element of of set at the right hand of the equation. Then there is a ρ such that $T = T_1\sigma\rho$. Since σ is a unifier of T_1 and T_2, $T = T_2\sigma\rho$ holds as well. Therefore T is contained in $\mathrm{Ground}(\mathcal{T}_1)$ (as $T_1(\sigma\rho)$) as well as in $\mathrm{Ground}(\mathcal{T}_2)$ (as $T_2(\sigma\rho)$). □

The most difficult operation on term based set representations is the complement of a set. Thereby we can restrict ourselves to sets represented by a single term, since obviously it is $\mathrm{Ground}(\mathcal{T}) = \bigcup_{T \in \mathcal{T}} \mathrm{Ground}(\{T\})$ and $\overline{M_1 \cup \cdots \cup M_n} = \overline{M_1} \cap \cdots \cap \overline{M_n}$.

Definition 8 (Symbolic complement) *Let $\Sigma = [S, \Omega]$ be a signature. For every sort s and every term in $T_{\Omega,s}$, we define a symbolic set of terms* $\mathrm{compl}(T)$ *by induction on the term structure. Thereby we assume that for every sort s there is a fixed variable x_s.*

A. *If* $T = \omega$ *for some* $\omega \in \Omega_{\epsilon,s}$ *then* $\mathrm{compl}(T) = (\Omega_{\epsilon,s} \setminus \{\omega\}) \cup \{\omega'(x_{s'}) \mid s' \in S, \omega' \in \Omega_{s',s}\}$;

B. *If* T *is a variable then* $\mathrm{compl}(T) = \emptyset$;

C. *If* $T = \omega(T')$ *for some* $s' \in S$, $\omega \in \Omega_{s',s}$, $T' \in T_{\Omega,s'}(X)$, *then* $\mathrm{compl}(T) = \{\omega(T^*) \mid T^* \in \mathrm{compl}(T')\} \cup \{\omega'(x_{s'}) \mid s' \in S, \omega' \in \Omega_{s',s}, \omega' \neq \omega\} \cup \Omega_{\epsilon,s}$.

We claim that this operation implements the complement on symbolically represented sets of terms.

Lemma 3 *Let* $T \in T_{\Omega,s}(X)$. *Then*

$$\overline{\mathrm{Ground}(\{T\})} = \mathrm{Ground}(\mathrm{compl}(T))^2$$

Proof. We have to prove that $\mathrm{Ground}(\{T\}) \cap \mathrm{Ground}(\mathrm{compl}(T))$ is empty while their union is the whole $T_{\Omega,s}$. We prove both claims by induction on the term structure.
1. empty intersection
Ad A. The only term contained in $\mathrm{Ground}(\omega)$ is ω itself. This ω is not contained in $\Omega_{\epsilon,s} \setminus \{\omega\}$ and cannot be generated by applying substitutions to any $\omega'(x_{s'})$.
Ad B. It is nothing to show.
Ad C. Let $T_0 \in \mathrm{Ground}(\{T\})$. Then T_0 can be represented as $\omega(T_0')$ for some $T_0' \in T_{\Omega,s'}$. Therefore T_0 can neither be contained in $\Omega_{\epsilon,s}$ nor can it be generated by applying any substitutions to a $\omega'(x)$ where ω' is different from ω. Furthermore, since T_0' is contained in $\mathrm{Ground}(T')$,the induction assumption assures that T_0' cannot be generated by applying substitutions to any element of $\mathrm{compl}(T')$. Therefore $T_0 = \omega(T_0')$ cannot be generated by applying substitutions to any $\omega(T^*)$ with $T^* \in \mathrm{compl}(T')$.
Thus the first claim holds.

2. union covers all ground terms
Ad A. A ground term is either a constant symbol and therefore contained in $\Omega_{\epsilon,s} = (\Omega_{\epsilon,s} \setminus \{\omega\}) \cup \{\omega\}$, or it is a unary operation symbol applied to another ground term and can therefore be generated by substituting some $\omega'(x_{s'})$ appropriately.
Ad B. Every ground term can be generated by substituting it for the variable x.
Ad C. A ground term is either a constant symbol and therefore represented by $\Omega_{\epsilon,s}$ or it is a unary operation symbol applied to another ground term T_0'. If this operation symbol is different from ω then the term can be generated by applying $[x_{s'} \longrightarrow T_0']$ to the fitting $\omega'(x_{s'})$. If the operation symbol in question

[2] A complement of a set $\mathcal{T}$ of terms of a sort s is defined according to s, that is $\overline{\mathcal{T}} = T_{\Omega,s} \setminus \mathcal{T}$.

is ω itself, then by induction assumption T_0' can be generated either from T^* or from an element of compl(T^*). Therefore the term itself can be generated as well. □

Fortunately all symbolic operations replace finite symbolic sets of terms again by finite ones. Therefore Lemma 1, 2 and 3 can be summed up to

Corollary 1 *The family of finitely representable sets of ground terms is closed under union, intersection and complement and therefore as well under set difference. Furthermore inclusion, emptiness and completeness of finitely represented sets can be decided.*

Proof. Set difference can be implemented by $M \setminus N = M \cap \overline{N}$. Obviously, since there are ground terms for every sort, Ground$(\mathcal{T})$ is empty iff $\mathcal{T}$ is. A set is complete iff its complement is empty. Finally, $M \subseteq N$ holds if and only if $M \cap \overline{N} = \emptyset$. All set operations presented above can be processed mechanically and yield finite representations for finite sets of terms. □

In the sequel we denote the symbolic operations representing union, intersection and complement by $\bigsqcup$, $\bigsqcap$ and *compl*, respectively.

However, not all sets of ground terms can be represented symbolically, for example the set $\{s^{2 \cdot n}(0) \mid n \in \mathbb{N}\}$ has no finite representation in the signature consisting of a constant symbol 0 and an unary operation symbol s.

The existence of a finite representation for set complements is the major benefit from the restrictions which we applied to the specification. Without giving a proof here, we claim that in equation free specifications which are not restricted to unary symbols a finite representation of the complement of sets represented by single terms exists if and only if every variables occurs at most once in this term. (Lemma 3 can be considered to be a corollary from this statement since in unary specifications no variable can appear more than once in a term.) Allowing equations in the specification the ability of complementing symbolically represented sets can get lost as well. However for finite data types there exist of course finite representations for arbitrary complements but the question arises whether the approach is then still "symbolic".

4 Informal introduction of symbolic siphons

The purpose of this section is to find a suitable symbolic representation of siphons on a still intuitive level but suitable to calculation. The running example of this section is the database maintenance scheme depicted in Figure 1.

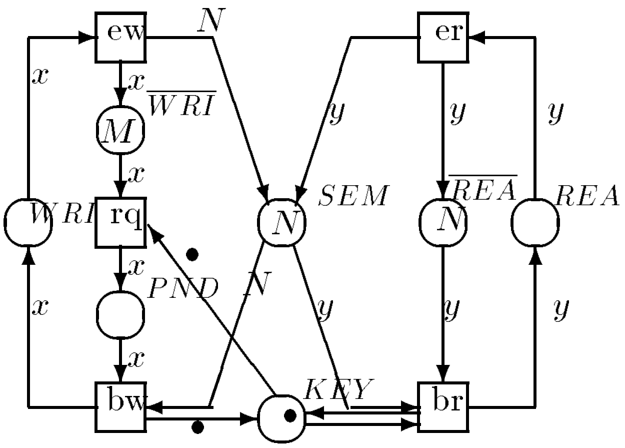

Figure 1: A database maintenance scheme

The underlying specification is

SPECIFICATION processes;

SORT Reading, Writing, Dot;

OPN	A, B, C	: $\longrightarrow$ Reading;
	D, E, F	: $\longrightarrow$ Writing;
	$\bullet$	: $\longrightarrow$ Dot;

The sorts of the places in Figure 1 are `Writing` for WRI, $\overline{WRI}$, and PND, `Reading` for REA, $\overline{REA}$, and SEM, and `Dot` for KEY. The letters N and M are meant to be replaced by $A+B+C$ and $D+E+F$, respectively.

The following list contains all siphons of the unfolded net, one per line.

$\overline{REA}.A$	$REA.A$			
$\overline{REA}.B$	$REA.B$			
$\overline{REA}.C$	$REA.C$			
$\overline{WRI}.D$	$PND.D$	$WRI.D$		
$\overline{WRI}.E$	$PND.E$	$WRI.E$		
$\overline{WRI}.F$	$PND.F$	$WRI.F$		
$WRI.D$	$WRI.E$	$WRI.F$	$SEM.A$	$REA.A$
$WRI.D$	$WRI.E$	$WRI.F$	$SEM.B$	$REA.B$
$WRI.D$	$WRI.E$	$WRI.F$	$SEM.C$	$REA.C$
$PND.D$	$PND.E$	$PND.F$	$KEY.\bullet$	

The horizontal lines are drawn to group similarly structured siphons together. Inside one group the siphons look almost equal, except one color. But this color iterates on the whole sort to generate the different siphons. Therefore it makes sense to represent the set of all siphons in the following way (thereby X is a variable of sort `Reading` and Y a variable of sort `Writing`):

$$\begin{array}{lllll} \overline{REA}.X & REA.X & & & \\ \overline{WRI}.Y & PND.Y & WRI.Y & & \\ WRI.D & WRI.E & WRI.F & SEM.X & REA.X \\ PND.D & PND.E & PND.F & KEY.\bullet & \end{array}$$

Every line in the latter description stands now for a *set* of siphons (the last one for a singleton set). A representation of an arbitrary siphon of the net can be obtained by replacing all occurrences of a variable in one "parameterized" siphon simultaneously by a ground term. In turn, every such replacement does really stand for a siphon. Of course, with a parameterized representation of siphons a lot of memory can be saved. But in addition, at least to the author's view, the parameterized representation increases the readability, since it abstracts from unnecessary details. In fact, it is much more comfortable to address "an arbitrary reading process" than "Process A or B or C". In general the parameterization allows for describing arbitrarily many siphons simultaneously (dependent on the cardinality of the parameterized color set). In principle, if a variable ranges on an infinite color set, even infinitely many siphons can be represented finitely.

But looking at the parameterized descriptions above, another source of parameterization comes into light. In the last two siphons one and the same place (WRI and PND, resp.) appears iterated on all its colors. Therefore we can introduce another kind of variable (written in small letters while capital letters had been used for the former kind) and represent these siphons by:

$$\begin{array}{lll} \forall z \in \texttt{Writing} : WRI.z & SEM.X & REA.X \\ \forall z \in \texttt{Writing} : PND.z & KEY.\bullet & \end{array}$$

In difference to the variables in capital letters (we shall call them *global variables*), assigning different values to variables written in small letter (*local variables* in the sequel) does not yield different siphons, but different places inside one siphon. Thus, while global variables can be used to represent arbitrary *many* siphons, local variables allow for representing arbitrary *large* siphons (of course, "arbitrary" depends on the cardinality of the specified color sets). As well as global variables, local variables are a step towards regarding places of the high level net as a whole instead of talking about the collection of their unfoldings. This way the readability again increases.

Having both kinds of variables (as for instance in the siphon containing $SEM.X$) there is some evidence that a large amount of reduction can be obtained by representing siphons symbolically. For instance, if we introduce arbitrarily many additional reading or writing processes in the above example, the number of symbolic siphons does not change at all when both local and global variables are used. If one kind of variables is omitted, either the size or the number of represented siphons grows when the number of processes is increased. However, for all formal considerations it will be necessary to distinguish carefully between the totally different meanings of global and local variables.

5 Symbolic families of place sets

According to the observations in the previous section we introduce two disjoint sets of variables — one for selecting a set among a set of simultaneously computed sets, the other one for selecting places inside a particular set.

Definition 9 (Symbolic family of place sets) *Let $AN = [\Sigma, P, T, F, \psi, \xi, \lambda, m_0]$ be an algebraic net. A* **symbolic family of place sets** *is a structure $[\{p_i.T_i \mid i \in I\}, X_G, X_L]$ where I is some index set, X_G and X_L are disjoint families of Σ–variables (***global*** and* **local** *variables, resp.) and for all $i \in I$ p_i is a place of AN and $T_i \in T_{\Omega,\psi(p_i)}(X_L \cup X_G)$.*

In order to distinguish different kinds of substitutions (since we shall argue about different variable sets we always use derivates of γ for substitutions on global variables, δ for *ground* substitutions on global variables, φ and χ for substitutions (ground substitutions) of local variables, and τ and β for substitutions (ground substitutions) of transition variables. For substitutions which range on unions of the above variable sets we shall use σ or ρ (for ground substitutions).

Among a symbolic family of places we can select a single set by instantiating the global variables (X_G). The variables in X_L remain unchanged, therefore the selected structure is still a symbolic representation, but of a single set of places.

Definition 10 (Symbolic set of Places) *Let $AN = [\Sigma, P, T, F, \psi, \xi, \lambda]$ be an algebraic net. A* **symbolic set of places** *is a structure $[\{p_i.T_i \mid i \in I\}, X_L]$ where I is some index set, X_L is a family of variables and for all $i \in I$ p_i is a place of AN and $T_i \in T_{\Omega,\psi(p_i)}(X_L)$.*

A symbolic *set* of places $[\{p_i.T_i \mid i \in I\}, X_L]$ can be identified with the singleton *family* of places $[\{p_i.T_i \mid i \in I\}, \emptyset, X_L]$.

Let $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, X_G, X_L]$ be a symbolic family of places, γ a substitution ranging over X_G such that $\mathrm{var}(\gamma) \cap X_L = \emptyset$ and define

$$\mathcal{D}\gamma = \{p_i.T_i\gamma \mid i \in I\}, \mathrm{var}(\gamma), X_L]$$

(Note that the application of a substitution that is defined only for some variables appearing in a term is always meant to be the identity on the remaining variables.)

When a symbolic family of place sets has the form $[M, \emptyset, X_L]$, that is an empty set of global variables, we identify it with the corresponding symbolic *set* of places $[M, X_L]$. Below it will turn out that this corresponds to the identification of a singleton set with the corresponding element. In this sense, for δ being a *ground* substitution of X_G, $\mathcal{D}\delta$ is a symbolic set of places.

Definition 11 (Interpretation of symbolic set/family) *Let $D = [\{p_i.T_i \mid i \in I\}, X_L]$ be a symbolic set of places. Then the set $UNF(D) = \{p_i.T_i\chi \mid i \in I, \chi \in \mathrm{GrSubst}(X_L)\}$ is called the* **interpretation** *of D. For a symbolic family of places $\mathcal{D} = p\{p_i.T_i \mid i \in I\}, X_G, X_L]$ define its* **interpretation** *(or unfolding) $UNF(\mathcal{D}) = \{UNF(D\delta) \mid \delta \in \mathrm{GrSubst}(X_G)\}$.*

The two-step interpretation of symbolic families models global and local variables exactly in the way in which we used them informally in the previous section.

6 General outline of the calculation

For place transition nets the following procedure (a simplified version of the one presented in [Sta90]) computes a family of siphons containing all minimal siphons if it is called `Siphons(`$\emptyset$`,`P`)`:

VAR Siphons: **SET OF SET OF** Places;
(initially empty)

PROCEDURE Siphons(PartialSiphon, CurrentChoice: SetOfPlaces);

BEGIN
    FOR EACH $p \in$ CurrentChoice **DO**
        NewPartialSiphon := PartialSiphon $\cup\{p\}$;
        IF exists t: $t \in$ NewPartialSiphon$F \setminus F$NewPartialSiphon **THEN**
            Siphons(NewPartialSiphon, tF);
        ELSE

Siphons := Siphons $\cup$\{ NewPartialSiphon \};
END IF
END FOR
END Siphons.

The basic idea is to start with a singleton set containing a place (all places of P are traversed in the top level call of `Siphons`) and to expand the set systematically according to violations of the siphon property. In fact every set `CurrentChoice` (except on the top–level) consists exactly of the set of places which are suitable to fix a certain violation of the siphon property (that is they all are input places of a transition which up to this stage of computation has only output places in the siphon). This way the expansion of place sets to siphons is narrowed to a set of "good" candidates. Of course, we shall try to re–use this general idea in the symbolic approach.

The operations to be re–implemented symbolically are two modules:

- *Check:* check whether a given symbolic family of places represents siphons or not; if it does, send it to the output, else send it to the expansion module. If a symbolic representation represents both siphons and non–siphons, split the representation into representations for the siphons and representations of the non–siphons and send the parts to the output or expansion module, respectively.
- *Expand:* for a given symbolic family of places and a transition violating the siphon property, try all possibilities to add an input place of this transition to the place sets and send the results back to the check module.

The procedure is started with $[p.X, \{X\}, \emptyset]$ for all $p \in P$, that is with a symbolic representation of single places, simultaneously for all places belonging to one and the same high–level place.

The module *check*, described in the first item, can be divided into several modules, each of them responsible for checking the siphon property according to one transition. The "ok" output of such a submodule shall be sent to the submodule corresponding to the next transition while the "not ok" output can be sent immediately to the *expand* unit as outlined in the second item. The output of the last *check*–unit forms the output of the whole procedure. Therefore the general outline of the calculation can be depicted as done in Figure 2.

In the next to sections we study the modules in detail, starting with the *check*–part.

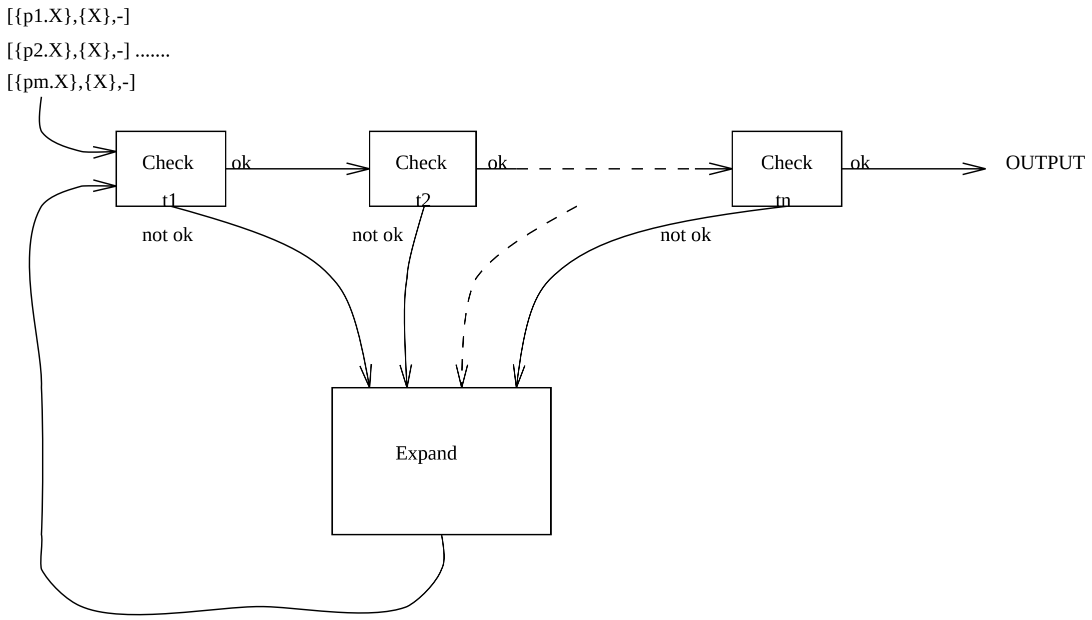

Figure 2: General outline of the calculation

7 Checking the Siphon Property

In this section we study how a symbolic family of place sets can be divided into a family of place sets which satisfy the siphon condition with respect to a single transition t and a family of place sets where this transition violates the siphon property. For both subfamilies symbolic representations must be computed. Thereby we cannot assure that a subfamily can be represented by a single symbolic family of place sets, but we shall always represent it by *finitely many* symbolic families of place sets.

For the whole section we assume that $\mathcal{D}$ contains at most one global variable. Actually this requirement is satisfied for the initial input and we shall see that during computation the number of global variables never increases. In the sequel we address to the global variable (if it exists) always by x_G.

Remembering that different values to global variables stand for different sets represented by a symbolic family of place sets, the partition of a symbolically represented family into siphons and non–siphons is equivalent to a partition of the set of all values of x_G, that is of all ground terms of the sort of x_G. Let $\mathcal{V}^+$ be the set of all ground terms T of the sort of x_G where $\mathcal{D}[x_G \longrightarrow T]$ represents a siphon while $\mathcal{V}^-$ is the set of all ground terms of the sort of x_G where $\mathcal{D}[x_G \longrightarrow T]$ represents a non–siphon. Obviously, we we succeed in representing both $\mathcal{V}^+$ and $\mathcal{V}^-$ symbolically, say by finite sets $\mathcal{P}^+$ and $\mathcal{P}^-$ (that is, $\mathrm{Ground}(\mathcal{P}^+) = \mathcal{V}^+$ and $\mathrm{Ground}(\mathcal{P}^-) = \mathcal{V}^-$) then the set of all siphons represented by $\mathcal{D}$ can be represented by the finite family $\{\mathcal{D}[x_G \longrightarrow T] \mid T \in \mathcal{P}^+\}$ while the set of all non–siphons can be represented by $\{\mathcal{D}[x_G \longrightarrow T] \mid$

$T \in \mathcal{P}^-\}$. The elements of these two sets are to be sent to the "ok" or "not ok" output of the check module, respectively.

Definition 12 (Symbolic partition) *Let $\mathcal{T}$ be a set of ground terms. A* **symbolic partition** *of $\mathcal{T}$ is a pair $[\mathcal{P}^+, \mathcal{P}^-]$ of finite sets of terms such that* $\text{Ground}(\mathcal{P}^+) \cap \text{Ground}(\mathcal{P}^-) = \emptyset$ *and* $\text{Ground}(\mathcal{P}^+) \cup \text{Ground}(\mathcal{P}^-) = \mathcal{T}$.

We use the index "+" always for siphons and "-" for non–siphons.

The section is organized as a sequence of steps. In the first two steps we characterize the siphon property in terms of unification. In the remaining ones we apply this characterization to partition a symbolic family of place sets according to the siphon property for one transition. We perform several transformations of the problem. For the purpose of keeping the overview we summarize every transformation by repeating briefly the studied problem and the approach which solves it or replaces it by other, less complex, problems.

1st step: Siphons and ground unifiers

For the purpose of checking the siphon property remember the definition of siphons. A set D of places forms a siphon iff every transition which has an output arc in D has also in input arc in D. Therefore the first step towards a symbolic check is to characterize symbolically those modes of t which put tokens onto a symbolic family $\mathcal{D} = [M, X_G, X_L]$ of places. Due to the definition of symbolic families of places for every place $p.T$ which is contained in any set appearing in the unfolding of $\mathcal{D}$ there are ground substitutions δ on the global variables and χ on the local variables such that $T = T_\flat \delta\chi$ for some $T_\flat$ such that $p.T_\flat$ appears in M. Thereby δ selects a *set* among those represented by $\mathcal{D}$ and χ selects a *color* of p among those represented by $p.T_\flat\delta$.

Since δ and χ range on disjoint families of variables, $\delta\chi$ can be written as $\delta \cup \chi$.

A transition t puts tokens on $p.T$ in a firing mode β iff $[t, p]$ is contained in F and there is a term $T_O \in \lambda([t, p])$ such that $T_O\beta = T$. Assumed without loss of generality, that the set of global and local variables are both disjoint to $\xi(t)$ (this can be easily achieved by renaming variables), it turns out that $\beta \cup \delta \cup \chi$ is a unifier of $T_\flat$ and T_O.

The other way round, if ρ is a ground substitution ranging on $X_G \cup X_L \cup \xi(t)$ such that $T_\flat\rho = T_O\rho$ (ρ is a unifier of $T_\flat$ and T_O) for some $T_\flat$ with $p.T_\flat$ appearing in $\mathcal{D}$ and $T_O \in \lambda([t, p])$ then $\rho\mid_{\xi(t)}$ is a firing mode of t in which t puts tokens on $p.T_\flat\rho$. Thereby $p.T_\flat\rho$ appears in the unfolding of $\mathcal{D}$ where the global variables are substituted by $\rho\mid_{X_G}$ while the local variables are substituted by $\rho\mid_{X_L}$. Thus $p.T_\flat\rho$ appears in particular in $UNF(\mathcal{D}\rho\mid_{X_G})$.

Therefore the relation between symbolic families of places and output arcs of transitions can be established by means of unification. Accordingly the relation between input arcs of transitions and symbolic families of places can be established by unification as well.

Once having characterized transitions which put tokens on a place appearing in a set represented by $\mathcal{D}$ and those which remove tokens from places appearing in sets represented by $\mathcal{D}$ we can re–formulate the siphon property by means of ground substitution. In fact we have to compare the unifiers between terms appearing in $\mathcal{D}$ and terms appearing at output arcs of a transition with the unifiers between terms appearing in $\mathcal{D}$ and terms appearing at input arcs of a transition. For this purpose we define these two sets of ground substitutions formally:

Definition 13 (Input and output unifiers) *Let AN be an algebraic net, t one of its transitions and $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, X_G, X_L]$ a symbolic family of places of AN for some index space I. The set $\mathcal{U}_I(\mathcal{D}, t)$ of* **input unifiers** *of $\mathcal{D}$ and t is defined by*

$$\mathcal{U}_I(\mathcal{D}, t) = \{\rho\,|_{X_G \cup \xi(t)} \mid \rho \in \mathrm{GrSubst}(X_G \cup X_L \cup \xi(t)) \text{ and } \exists\sharp\exists T_I :$$

$$\sharp \in I, T_I \in \lambda([p_\sharp, t]) \text{ such that } T_\sharp\rho = T_I\rho\}$$

while the set $\mathcal{U}_O(\mathcal{D}, t)$ of **output unifiers** *of $\mathcal{D}$ and t is defined by*

$$\mathcal{U}_O(\mathcal{D}, t) = \{\rho\,|_{X_G \cup \xi(t)} \mid \rho \in \mathrm{GrSubst}(X_G \cup X_L \cup \xi(t)) \text{ and } \exists\flat\exists T_O :$$

$$\flat \in I, T_O \in \lambda([t, p_\flat]) \text{ such that } T_\flat\rho = T_O\rho\}$$

Input and output unifiers can be interpreted as representations of ordered pairs where the first component is a set appearing in the unfolding of $\mathcal{D}$ and the second one is a firing mode of t. According to the above considerations the set of input unifiers collects all those pairs where t, fired in the mode given by the second component, removes tokens from the set given by the first component. Accordingly, the set of output unifiers collects all pairs where t puts tokens on the set represented by the first component, when fired in the mode given by the second one. The first and second components are thereby the restrictions to X_G and $\xi(t)$, respectively. Note, that the substitutions in the $\mathcal{U}_I$ and $\mathcal{U}_O$ range only on $X_G \cup \xi(t)$ while we forget about the $\rho\,|_{X_L}$. This means (informally) that we do not need information about which particular places in $\mathcal{D}$ are touched by t, we only need to know that they exist. In Definition 13 we considered therefore the whole ρ in the equations $T_\sharp\rho = T_I\rho$ and $T_\flat\rho = T_O\rho$ but removed the part concerning X_L when collecting these unifiers to a set. However the kept information is sufficient to check the siphon property:

Lemma 4 *Let AN be an algebraic net. Let I be some index space, and $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, \{x_G\}, X_L]$ be a symbolic family of places. For every ground term T of the sort of x_G, $UNF(\mathcal{D}[x_G \longrightarrow T])$ is a siphon if and only if for all transitions t of AN it holds*

$$\mathcal{U}_O(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]} \subseteq \mathcal{U}_I(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]}$$

Thereby for a set of substitutions M, $M \mid_{[x \longrightarrow T]}$ stands for the set of all elements in M which map x to T.

Of course we assume as usual that global, local and transition variables are disjoint.

Proof.
"if" direction:
Let $\mathcal{U}_O(\mathcal{D}, t) \mid_{[x \longrightarrow T]} \subseteq \mathcal{U}_I(\mathcal{D}, t) \mid_{[x \longrightarrow T]}$ for all transitions t. We show that $UNF(\mathcal{D}[x_G \longrightarrow T])$ is a siphon. Let $t.\beta \in F\ UNF(\mathcal{D}[x_G \longrightarrow T])$ for some transition t and one of its firing modes β. Then, according to the above considerations, there is some $\flat \in I$, a $\chi \in X_L$ and a $T_O \in \lambda([t, p_\flat])$ such that $T_\flat([x_G \longrightarrow T] \cup \chi) = T_O\beta$. Since $\{x_G\}$, X_L and $\xi(t)$ are disjoint, $[x \longrightarrow T] \cup \chi \cup \beta$ is a unifier of T_O and $T_\flat$. Thus it holds $([x \longrightarrow T] \cup \beta) \in \mathcal{U}_O(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]}$ and consequently $([x_G \longrightarrow T] \cup \beta) \in \mathcal{U}_I(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]}$.

This means that there is a χ' such that for some $\sharp \in I$ and some $T_I \in \lambda([p_\sharp, t])$ it holds $T_I([x_G \longrightarrow T] \cup \chi' \cup \beta) = T_\sharp([x_G \longrightarrow T] \cup \chi' \cup \beta)$ which can be rewritten by $T_I\beta = T_\sharp([x_G \longrightarrow T] \cup \chi')$. The latter equation proves that there is a place contained in $UNF(\mathcal{D}[x_G \longrightarrow T])$, namely $p_\sharp.T_\sharp[x_G \longrightarrow T]\chi'$ which is an input place of $t.\beta$.

"only if" direction:
Let $UNF(\mathcal{D}[x_G \longrightarrow T])$ be a siphon, t be a transition and $\rho \in \mathcal{U}_O(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]}$. We show that $\rho \in \mathcal{U}_I(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]}$. For this purpose let $\beta = \rho \mid_{\xi(t)}$. Obviously it holds $\rho = \beta \cup [x_G \longrightarrow T]$. Consider the siphon $UNF(\mathcal{D}[x_G \longrightarrow T])$ and $t.\beta$ which is a firing mode of t. Since by definition of $\mathcal{U}_O(\mathcal{D}, t)$ for some $\flat \in I$, some $T_O \in \lambda([t, p_\flat])$ and some $\chi \in \mathrm{GrSubst}(X_L)$ the equation $T_\flat([x_G \longrightarrow T] \cup \chi) = T_O\beta$ is satisfied, $t.\beta$ puts tokens on $p_\flat.T_\flat[x_G \longrightarrow T]\chi$. This is a place appearing in $UNF(\mathcal{D}[x_G \longrightarrow T])$. Since this set is a siphon, t has an input place in $UNF(\mathcal{D}[x_G \longrightarrow T])$. This input place can be written as $p_\sharp.T_\sharp[x_G \longrightarrow T]\chi'$ for some $\sharp \in I$ and $\chi' \in GrSubst(X_L)$. When this place is an input place of $t.\beta$ there exists a $T_I \in \lambda([p_\sharp, t])$ such that $T_I\beta = T_\sharp p x_G \longrightarrow T]\chi'$. By the assumption that all involved variable sets are disjoint we may conclude $T_I(\beta \cup [x_G \longrightarrow T] \cup \chi') = T_\sharp(\beta \cup [x_G \longrightarrow T] \cup \chi')$. Consequently it holds $(\beta \cup [x_G \longrightarrow T] \cup \chi') \mid_{X_G \cup \xi(t)} = (\beta \cup [x_G \longrightarrow T]) = \rho \in \mathcal{U}_I(\mathcal{D}, t)_{[x_G \longrightarrow T]}$. $\square$

2nd step: Switch to most general unifiers

$\mathcal{U}_I$ and $\mathcal{U}_O$ are sets of ground substitutions. Of course the next step towards a symbolic approach is to replace ground unifiers by most general unifiers. We assume without loss of generality that variables appearing in some most general unifiers are always different from all earlier introduced variables, whether they belong to the net, to the symbolic family of places in question or to earlier computed unifiers. By renaming this restriction can always be satisfied and helps to avoid confusion about mixing up variables.

Definition 14 (Most general input and output unifiers) *Let AN be an algebraic net, t one of its transitions and $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, X_G, X_L]$ a symbolic family of places of AN for some index space I. The set $\mathcal{C}_I(\mathcal{D}, t)$ of* **most general input unifiers** *of $\mathcal{D}$ and t is defined by*

$$\mathcal{C}_I(\mathcal{D}, t) = \{MGU(T_\sharp, T_I) \mid_{X_G \cup \xi(t)} \mid \sharp \in I, T_I \in \lambda([p_\sharp, t])\}$$

while the set $\mathcal{C}_O(\mathcal{D}, t)$ of **most general output unifiers** *of $\mathcal{D}$ and t is defined by*

$$\mathcal{C}_O(\mathcal{D}, t) = \{MGU(T_\flat, T_O) \mid_{X_G \cup \xi(t)} \mid \flat \in I, T_O \in \lambda([t, p_\flat])\}$$

The relation between the $\mathcal{U}_{I/O}$ and $\mathcal{C}_{I/O}$ is obvious:

Corollary 2 *For every symbolic family of places $\mathcal{D}$ and every transition t it holds*

$$\begin{aligned}
\mathcal{U}_I(\mathcal{D}, t) &= \mathrm{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \\
\mathcal{U}_O(\mathcal{D}, t) &= \mathrm{Ground}(\mathcal{C}_O(\mathcal{D}, t) \\
\mathcal{U}_I(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]} &= \mathrm{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]} \\
\mathcal{U}_O(\mathcal{D}, t) \mid_{[x_G \longrightarrow T]} &= \mathrm{Ground}(\mathcal{C}_O(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]}
\end{aligned}$$

Thereby one set inclusion holds since MGU assigns a most general *unifier* to a pair of terms while the reverse inclusion holds since MGU assigns a *most general* unifier. With the concept of most general unifiers we have a symbolic characterization of the siphon property. This is what we wanted to achieve in the first part of this section.

The remaining task for this section is to compute a partition of all possible values to be substituted for the global variable. We obtain this partition by transforming the MGU–based characterization of the siphon property several times, always solving parts of it or dividing the problem into several less complex ones from the solutions of which we can get the solution of the original problem. For keeping overview on what happens it is helpful to keep in mind

1. that we always consider a term to be a symbolic representation of all ground terms which can be obtained by substituting its variables;
2. that a substitution likewise stands for a set of tuples (one component per variable in the domain);
3. that the global variable plays a specific role, since this is the variable which addresses the different place sets represented by $\mathcal{D}$ and for which we finally want to apply restrictions. Therefore this variable plays the role of an arbitrary but fixed parameter and must be carefully separated from the remaining variables. Consequently, if the global variable is substituted by a term still containing a variable, then this new variable takes the role of the parameter since every ground instantiation of this new variable selects a place set among those place sets still representable after applying the substitution. Accordingly every instantiation of the global variable by a ground term appears as the selection of one single place set from $\mathcal{D}$.

3rd step: Simplify siphon property to single output unifier

The task of verifying the inclusion according to Lemma 4 can be split into a sequence of smaller ones. Obviously the set $\mathrm{Ground}(\mathcal{C}_O(\mathcal{D},t))$ can be written as $\bigcup_{\sigma_O \in \mathcal{C}_O(\mathcal{D},t)} \mathrm{Ground}(\{\sigma_O\})$. Furthermore a union of sets M_i is included in another set N if and only if every M_i is included in N. Therefore we can compute a partition of the ground terms T to be substituted for x_G according to the condition $\mathrm{Ground}(\mathcal{C}_O(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]}$ from partitions according to $\mathrm{Ground}(\{\sigma_O\}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]}$. Thereby we have to consider the latter inclusion for every $\sigma_O \in \mathcal{C}_O(\mathcal{D},t)$, that is for every σ_O which is the most general unifier of $T_\flat$ and T_O where T_O is a term appearing in $\lambda([t,p_\flat])$ and $p_\flat.T_\flat$ appears in $\mathcal{D}$.

The considerations of this step can be summarized as follows:

Studied Problem: Get symbolic partition $[\mathcal{P}^+,\mathcal{P}^-]$ for the set all ground terms T of the sort of x_G according to

$$\mathrm{Ground}(\mathcal{C}_O(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]}$$

Approach: Get symbolic partitions $[\mathcal{P}^+_{\sigma_O},\mathcal{P}^-_{\sigma_O}]$ for the set of all ground terms T of the sort of x_G according to

$$\mathrm{Ground}(\{\sigma_O\}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D},t)) \mid_{[x_G \longrightarrow T]}$$

for every $\sigma_O \in \mathcal{C}_O(\mathcal{D},t)$.

$$\mathcal{P}^+ := \bigsqcap\nolimits_{\sigma_O \in \mathcal{C}_O(\mathcal{D},t)} \mathcal{P}^+_{\sigma_O}$$

$$\mathcal{P}^- := \bigsqcup_{\sigma_O \in \mathcal{C}_O(\mathcal{D},t)} \mathcal{P}^-_{\sigma_O}$$

Justification: $\bigcup M_i \subseteq N$ iff $\forall i : M_i \subseteq N$.

Open Problems: Get symbolic partition $[\mathcal{P}^+_{\sigma_O}, \mathcal{P}^-_{\sigma_O}]$ for the set of all ground terms T of the sort of x_G according to

$$\mathrm{Ground}(\{\sigma_O\}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]}$$

for $\sigma_O \in \mathcal{C}_O(\mathcal{D}, t)$

4th step: Align unifiers according to value for global variable

Since our final goal is to partition all the values for x_G, we should sort the different unifiers occurring in the inclusion problem under consideration according to their value for the global variable. That is, we split the inclusion problem again by considering in every part only those parts of the earlier representation which concern equal values to the global variable. Thereby we allow for most general unification to make terms equal.

Definition 15 (Alignment of unifiers) *Let σ_O be an output unifier and $\mathcal{C}$ a subset of $\mathcal{C}_I(\mathcal{D}, t)$. Let x_G be the unique global variable of $\mathcal{D}$. Let $\gamma_\mathcal{C}$ be the most general unifier of $\sigma_O(x_G)$ and all $\sigma(x_G)$ for $\sigma \in \mathcal{C}$. The* **aligned partition problem** *for $\mathcal{C}$ is to partition symbolically all terms T in* $\mathrm{Ground}(\{\gamma_\mathcal{C}(\sigma(x_G))\})$ *into $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ according to the condition*

$$\mathrm{Ground}(\sigma_O \gamma_\mathcal{C}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\{\sigma \gamma_\mathcal{C} \mid \sigma \in \mathcal{C}\}) \mid_{[x_G \longrightarrow T]}$$

$\gamma_\mathcal{C}$ is called the **alignment unifier** *corresponding to $\mathcal{C}$.*

If for some $\mathcal{C}$ the alignment unifier $\gamma_\mathcal{C}$ does not exist, then the whole aligned unification problem for $\mathcal{C}$ is disregarded. Note further, that the application of $\gamma_\mathcal{C}$ leaves all variables except x_G untouched. The aligning operation is justified by

Lemma 5 *Let $\mathcal{D}$ be a symbolic family of place sets and T a ground term of the sort of x_G. The relation*

$$\mathrm{Ground}(\sigma_O) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]}$$

holds if and only if one of the following conditions is true:

1. *$\sigma_O(x_G)$ and T are not unifiable;*

2. *It exists a* $\mathcal{C} \subseteq \mathcal{C}_I(\mathcal{D}, t)$ *such that for the alignment unifier* $\gamma_\mathcal{C}$ *corresponding to* $\mathcal{C}$, $T \in \mathrm{Ground}(\{\gamma_\mathcal{C}(\sigma_O(x_G))\})$ *and*

$$\mathrm{Ground}(\sigma_O \gamma_\mathcal{C}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\{\sigma\gamma_\mathcal{C} : \sigma \in \mathcal{C}\}) \mid_{[x_G \longrightarrow T]}$$

Proof.
"if" direction:
Ad 1. Let $\sigma_O(x_G)$ and T be not unifiable. That is, σ_O has no ground instance ρ where $\rho(x_G) = T$. Therefore $\mathrm{Ground}(\sigma_O) \mid_{[x_G \longrightarrow T]}$ is empty and the inclusion holds.
Ad 2. Since $T \in \mathrm{Ground}(\{\gamma_\mathcal{C}(\sigma_O(x_G))\})$ and the application of $\gamma_\mathcal{C}$ leaves all variables except in $\mathrm{var}(\sigma_O(x_G))$ untouched, it holds

$$\mathrm{Ground}(\sigma_O) \mid_{[x_G \longrightarrow T]} = \mathrm{Ground}(\sigma_O \gamma_\mathcal{C}) \mid_{[x_G \longrightarrow T]}$$

Since furthermore $\mathcal{C}$ is a subset of $\mathcal{C}_I(\mathcal{D}, t)$ and the application of $\gamma_\mathcal{C}$ can apply only further restrictions, it holds

$$\mathrm{Ground}(\{\sigma\gamma_\mathcal{C} : \sigma \in \mathcal{C}\}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]}$$

Therefore the left hand sides of the two inclusions are equal while the right hand side of the assumed inclusion is included in the right hand of the claimed inclusion. Thus the "if" direction holds.

"only if" direction: We may assume that $T \in \mathrm{Ground}(\sigma_O(x_G))$ since otherwise according to the first part of the claim nothing remains to show. Thus there exists a substitution δ on $\mathrm{var}(\sigma(x_G))$ (this set may be empty, then δ is the empty substitution) such that $T = \sigma_O(x_G)\delta$. Let $\mathcal{C}$ be the set of all substitutions σ contained in $\mathcal{C}_I(\mathcal{D}, t)$ where $T \in \mathrm{Ground}(\sigma(x_G))$, say by $T = \sigma(x_G)\delta_\sigma$. Obviously for all substitutions σ' in $\mathcal{C}_I(\mathcal{D}, t)$ which are not contained in $\mathcal{C}$, $\mathrm{Ground}(\sigma')$ does not contain any substitution ρ for which $\rho(x_G) = T$ holds. Moreover all the $\mathrm{var}(\sigma)$ for $\sigma \in \mathcal{C}_I(\mathcal{D}, t) \cup \{\sigma_O\}$ are by general assumption pairwise disjoint, since they result from different applications of unification. Therefore $\delta' = \delta \cup \bigcup_{\sigma \in \mathcal{C}} \{\delta_\sigma\}$ is a unifier for $\sigma_O(x_G)$ and all the $\sigma(x_G)$ for $\sigma \in \mathcal{C}$ since all the instances yield T. Consequently the alignment unifier $\gamma_\mathcal{C}$ corresponding to $\mathcal{C}$ is more general than δ'. But then every ground instance of a substitution where the value of x_G equals $\delta'(x_G)$ is also a ground instance after its concatenation with $\gamma_\mathcal{C}$. Therefore for the particular $\mathcal{C}$ currently under consideration the claimed inclusion as well as the assumed inclusion concern exactly the same sets. □

When we are able to get symbolic partitions of aligned problems, then according to Lemma 5 the partition which remained open in the third step can be easily obtained: the "+" part is formed by the union of all "+" parts of the different aligned problems (second item of the lemma) and the complement of all ground instances of $\sigma_O(x_G)$ (first item of the lemma) while the "-" part can be computed as the complement of the "+" part. Note that the condition of the second item of the lemma is only established for elements in $\text{Ground}(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$. Thus this condition cannot be used to partition *all* values of x_G, it can only serve for a partition of $\text{Ground}(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$. This completes the justification for the definition of alignment. Therefore during the next steps we can focus on aligned inclusion problems only. The benefit from the aligning operation is that all future considerations concern only the values of input and output unifiers for the variables in $\xi(t)$ since the values on x_G are all equal. Here is the summary of this step:

Studied problem: Get symbolic partition $[\mathcal{P}^+_{\sigma_O}, \mathcal{P}^-_{\sigma_O}]$ for the set of all ground terms T of the sort of x_G according to

$$\text{Ground}(\{\sigma_O\}) \mid_{[x_G \longrightarrow T]} \subseteq \text{Ground}(\mathcal{C}_I(\mathcal{D}, t)) \mid_{[x_G \longrightarrow T]}$$

for $\sigma_O \in \mathcal{C}_O(\mathcal{D}, t)$.

Approach: Solve for every $\mathcal{C} \in \mathcal{C}_I(\mathcal{D}, t)$ the aligned problem for $\mathcal{C}$, yielding for every $\mathcal{C}$ a symbolic partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ of the set $\text{Ground}(\gamma_{\mathcal{C}}(\sigma(x_G)))$.

$$\mathcal{P}^+_{\sigma_O} := \bigsqcup_{\mathcal{C} \subseteq \mathcal{C}_I(\mathcal{D},t)} \mathcal{P}^+_{\sigma_O,\mathcal{C}} \sqcup compl(\sigma_O(x_G))$$

$$\mathcal{P}^-_{\sigma_O} := compl(\mathcal{P}^+_{\sigma_O})$$

Justification: Lemma 5.

Open problems: Solve an aligned unification problem for some $\mathcal{C} \in \mathcal{C}_I(\mathcal{D}, t)$.

5th step: Partition of $\mathcal{C}$

In this step we prepare for calculating partially complements of $\text{Ground}(\mathcal{C})$. Thereby the problem is that term based set algebra works only for single terms which is not evident for a corresponding approach to substitutions. However, with additional benefit from the restriction to unary specifications we can reduce the complement problem for unifiers to the complement problem for terms. The major observation for this purpose is that every input and output unifier origins from two terms, one of which appears in $\mathcal{D}$ while the other one appears at a transition arc. Both terms can contain at most one variable. Therefore in every input and output unifier at most for two variables real restriction can apply while all the remaining variables are known to be unbound at all (that is they are only renamed by a new variable which does not appear anywhere else). At most one of the restricted variables is contained in $\xi(t)$ (the one possibly appearing in the arc term used for the unification) while the other one belongs to $\mathcal{D}$ either as local variable (then it is not present in the unifier any more) or as global variable.

That is, among the variables in $\xi(t)$ there is at most one variable for which we can not assure that its value is a totally unbound variable. Therefore in the set of ground substitutions which are represented by an input or output unifier *all* value combinations occur for all variables except x_G and the one variable which appeared in the arc term used for unification.

For the purpose of benefiting from this special structure of unifiers we sort the substitutions representing the right hand side of the inclusion which forms the condition for the aligned partition problem for $\mathcal{C}$. These substitutions are exactly the $\gamma_{\mathcal{C}}$–instances of the most general input unifiers appearing in $\mathcal{C}$.

With $\mathcal{C}_x$ we denote the set of all those $\sigma\gamma_{\mathcal{C}}$ where σ appears in $\mathcal{C}$ and origins from a unification problem $\{T_I, T_\sharp\}$ with T_I being the arc term, $T_\sharp$ being the term appearing in $\mathcal{D}$ and satisfying $\text{var}(T_I) = \{x\}$. The set $\mathcal{C}_\epsilon$ is defined correspondingly for those unifiers which origin from an arc term which is a ground term. Obviously the family of all non–empty $\mathcal{C}_x$ (for $x \in \xi(t) \cup \{\epsilon\}$) is a partition on $\mathcal{C}\gamma_{\mathcal{C}}$ (the representation of the right hand side of the inclusion under consideration). In every $\mathcal{C}_x$ x is the only variable of $\xi(t)$ which is possibly restricted. In order to distinguish between the occurrence of the parameter variable and other variables we split every $\mathcal{C}_x$ into $\mathcal{C}_{x,p}$ and $\mathcal{C}_{x,\overline{p}}$. An element $\sigma\gamma_{\mathcal{C}}$ of $\mathcal{C}_x$ is contained in $\mathcal{C}_{x,p}$ iff $\text{var}(\gamma_{\mathcal{C}}(\sigma(x)))$ is not empty and equals $\text{var}(\gamma_{\mathcal{C}}(\sigma(x_G)))$ (which implies that $\text{var}(\gamma_{\mathcal{C}}(\sigma(x_G)))$ is not empty). Otherwise $\sigma\gamma_{\mathcal{C}}$ is put into $\mathcal{C}_{x,\overline{p}}$.

Informally, with these notations we have divided the input unifiers into those containing stuff which depends on the particular set included in $\mathcal{D}$ (via the value of the parameter variable) and those which contain stuff which is totally independent on the particular set in $\mathcal{D}$ (expressed through the absence of the parameter variable). Here is the summary:

Studied problem: Given a $\mathcal{C} \in \mathcal{C}_I(\mathcal{D}, t)$, solve the aligned partition problem for $\mathcal{C}$.

Approach/Open problem: Compute a symbolic partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ of all $T \in Ground(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\text{Ground}(\sigma_O\gamma_C) \mid_{[x_G \longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x\in\xi(t)} \mathcal{C}_{x,p} \cup \bigcup_{x\in\xi(t)} \mathcal{C}_{x,\overline{p}} \cup \mathcal{C}_\epsilon) \mid_{[x_G \longrightarrow T]}$$

Justification: The open problem differs from the studied one only by writing the right hand side as a union.

6th step: Examine $\mathcal{C}_\epsilon$

Assume, $\mathcal{C}_\epsilon$ is not empty. Since the unification of a ground term at an arc of t does not influence any variable binding on $\xi(t)$, a substitution contained in this set maps *every* variable of $\xi(t)$ to a free variable all of which are different. Therefore this part of the substitution represents *all* firing modes of t. Consequently in this case the inclusion $\mathrm{Ground}(\sigma_O\gamma_\mathcal{C}) \subseteq \mathrm{Ground}(\mathcal{C}_\epsilon)$ holds without making further considerations necessary. Since $\mathcal{C}_\epsilon$ represents a subset of the right hand side of the problem which is still open after the 5th step, the "-" part of the partition to be obtained there is empty while the "+" part consists of the whole $\mathrm{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$.

Studied problem: Compute a symbolic partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ of all $T \in$ $Ground(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\mathrm{Ground}(\sigma_O\gamma_C)\mid_{[x_G\longrightarrow T]}\subseteq \mathrm{Ground}(\bigcup_{x\in\xi(t)} \mathcal{C}_{x,p} \cup \bigcup_{x\in\xi(t)} \mathcal{C}_{x,\overline{p}} \cup \mathcal{C}_\epsilon)\mid_{[x_G\longrightarrow T]}$$

in the case that $\mathcal{C}_\epsilon \neq \emptyset$.

Approach:

$$\mathcal{P}^+_{\sigma_O,\mathcal{C}} = \{\gamma_\mathcal{C}(\sigma_O(x_G))\}$$
$$\mathcal{P}^-_{\sigma_O,\mathcal{C}} = \emptyset$$

Justification $\mathcal{C}_\epsilon$ represents *all* value tuples on $\xi(t)$.

Open problem: Compute a symbolic partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ of all $T \in$ $Ground(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\mathrm{Ground}(\sigma_O\gamma_C)\mid_{[x_G\longrightarrow T]}\subseteq \mathrm{Ground}(\bigcup_{x\in\xi(t)} \mathcal{C}_{x,p} \cup \bigcup_{x\in\xi(t)} \mathcal{C}_{x,\overline{p}} \cup \mathcal{C}_\epsilon)\mid_{[x_G\longrightarrow T]}$$

in the case that $\mathcal{C}_\epsilon = \emptyset$.

7th step: Intersect the inclusion problem with complements of some input unifiers

The next parts of $\mathcal{C}$ we want to get rid of are the $\mathcal{C}_{x,\overline{p}}$. As already mentioned, these substitutions apply restrictions to one variable in $\xi(t)$ only.

If we regard only one element σ of $\mathcal{C}_{x,\overline{p}}$, then the set $\mathrm{Ground}(\sigma)$ has a specific structure. Except x all variables of σ are renamings, where every variable is renamed by a unique new variable. That is, on $\xi(t) \setminus \{x\}$ the ground instances of σ cover *all* combinations of ground terms of the corresponding sorts. More abstract, Ground(σ), considered only on $\xi(t)$, has the form $M \times E$ where M is a set of those ground terms which can be assigned to the only restricted variable x in $\xi(t)$ and E the set of *all* value combinations on the remaining variables. For such a set it holds: $\overline{M \times E} = \overline{M} \times E$. Fortunately there is no correlation between M and the value of σ on x_G, since otherwise σ would be contained in $\mathcal{C}_{x,p}$ rather than in $\mathcal{C}_{x,\overline{p}}$. Therefore the whole set Ground(σ) can be written as $G \times M \times E$ where $G = \text{Ground}(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$. Accordingly the whole Ground$(\mathcal{C}_{x,\overline{p}})$ can be represented as $G \times N \times E$ where $N = \text{Ground}(\{\sigma(x) \mid \sigma \in \mathcal{C}_{x,\overline{p}}\})$ and G and E are the same as above. When we denote Ground$(\sigma_O\gamma_{\mathcal{C}})$ (the left hand of the inclusion under consideration) by A and the set of all ground substitutions which are represented by any elements of the $\mathcal{C}_y$ $(y \in \xi(t))$ except those in $\mathcal{C}_{x,\overline{p}}$, then the inclusion to be studied up to now is

$$A \subseteq G \times N \times E \cup B$$

(Actually we still have to study the projection of this inclusion to particular ground values of x_G but the whole inclusion represents of course all these projections simultaneously. Therefore we present the following arguments in the easier notation, without projections.) According to set algebraic laws we can rewrite the inclusion in the following way:

$$\begin{array}{ll}
 & A \subseteq G \times N \times E \cup B \\
\text{iff} & A \cap \overline{G \times N \times E} \subseteq B \\
\text{iff} & A \cap \overline{G \times N} \times E \subseteq B \\
 & \text{(since } E \text{ represents } \textit{all} \text{ values)} \\
\text{iff} & A \cap (\overline{G} \times E \times E \cup E \times \overline{N} \times E) \subseteq B \\
\text{iff} & A \cap E \times \overline{N} \times E \subseteq B \\
 & \text{(since the 1st component of } A \text{ is contained in } G) \\
\text{iff} & A \cap G \times \overline{N} \times E \subseteq B \\
 & \text{(since the 1st component of } A \text{ is contained in } G)
\end{array}$$

The last two transformations benefit from the alignment which we have done on unifiers. In the new representation of the inclusion the right hand of the inclusion is formed by all the $\mathcal{C}_{y,p}$ and $\mathcal{C}_{y,\overline{p}}$ as it was before, except $\mathcal{C}_{x,\overline{p}}$ for the particular x which is currently under consideration. The operations to be performed on the left hand are to find a symbolic representation for $G \times \overline{N} \times E$ and to intersect the result with A which is represented by $\sigma_O\gamma_{\mathcal{C}}$. Due to the considerations in Section 3 $G \times \overline{N} \times E$ can be built by taking an arbitrary σ from $\mathcal{C}_{x,\overline{p}}$ and replacing its image on x by the elements of compl$(\{\sigma'(x) \mid \sigma' \in \mathcal{C}_{x,\overline{p}}\})$. This operation results in a set of substitutions which we denote by compl($\mathcal{C}_{x,\overline{p}}$). Obviously this construction implements the operation that transforms $G \times N \times E$ into $G \times \overline{N} \times E$.

Before intersecting compl($\mathcal{C}_{x,\overline{p}}$) with $\{\sigma_O \gamma_\mathcal{C}\}$ symbolically, we can apply the above transformation for *all* $\mathcal{C}_{y,\overline{p}}$ ($y \in \xi(t)$). Then the right hand of the remaining inclusion is formed only by the $\mathcal{C}_{y,p}$ while at the left hand side the intersection of $\sigma_O \gamma_\mathcal{C}$ and all the *compl*($\mathcal{C}_{x,\overline{p}}$) (actually the intersection of their ground instances) must be implemented.

For this purpose remember that for every element ρ of the intersection of symbolically represented sets (even of substitutions) there is an element in the representation of every set such that ρ is a ground instance of it. Thus ρ appears in the intersection of symbolically represented sets of substitutions if and only if it is a unifier of a set of substitutions which consists of exactly one element per symbolic representation of a set to be intersected. This way the left hand of the transformed inclusion is formed by a set $\{\sigma_O \gamma_C \sigma^* \mid \sigma^* = MGU(\{\sigma_O \gamma_\mathcal{C}\} \cup \bigcup_{x \in \xi(t)} \{\sigma_x\})$ and $\forall x \in \xi(t) : \sigma_x \in compl(\mathcal{C}_{x,\overline{p}})\}$

Of course the resulting set of substitutions can be empty, either totally empty or at least empty for some of its projections to ground values on x_G. Especially it is totally empty if one of the *compl*($\mathcal{C}_{x,\overline{p}}$) is empty since in this case no unification problem can be built. However, this emptiness means that the inclusion for those (or all, respectively) values of x_G can be proven only taking into consideration the $\mathcal{C}_{x,\overline{p}}$ which is nothing wrong. We can put those values immediately into the "+" part of our current partition. Therefore the summary of this step is:

Studied Problem: Compute a symbolic partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ of all $T \in Ground(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\text{Ground}(\sigma_O \gamma_C) \mid_{[x_G \longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x \in \xi(t)} \mathcal{C}_{x,p} \cup \bigcup_{x \in \xi(t)} \mathcal{C}_{x,\overline{p}} \cup \mathcal{C}_\epsilon) \mid_{[x_G \longrightarrow T]}$$

in the case that $\mathcal{C}_\epsilon = \emptyset$.

Approach: Compute $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ according to

$$\text{Ground}(\mathcal{S}) \mid_{[x_G \longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x \in \xi(t)} \mathcal{C}_{x,p}) \mid_{[x_G \longrightarrow T]}$$

where

$$\mathcal{S} = \{\sigma_O \gamma_C \sigma^* \mid \sigma^* = MGU(\{\sigma_O \gamma_\mathcal{C}\} \cup \bigcup_{x \in \xi(t)} \{\sigma_x\}) \text{ and } \forall x \in \xi(t) :$$
$$\sigma_x \in compl(\mathcal{C}_{x,\overline{p}})\}$$

If $\mathcal{S} = \emptyset$, then

$$\mathcal{P}^+_{\sigma_O,\mathcal{C}} := \{\gamma_\mathcal{C}(\sigma_O(x_G))\}$$
$$\mathcal{P}^-_{\sigma_O,\mathcal{C}} := \emptyset$$

Justification: Set algebraic arguments for soundness of transformation; symbolic implementation on partial complement for $\mathcal{C}_{x,\overline{p}}$ and intersection of symbolically represented sets of unifiers.

Open problem: Compute $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ according to

$$\mathrm{Ground}(\mathcal{S}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\bigcup_{x \in \xi(t)} \mathcal{C}_{x,p}) \mid_{[x_G \longrightarrow T]}$$

8th step: Realignment

After having again a *set* of substitutions representing the left hand of the inclusion now, we can once more split the problem into smaller ones where the left hand is represented only by a single unifier. Thereby we can notice the possible instantiations of the global variable part on the left hands of the inclusion. Since for all projections according to the value of x_G which are no ground instances of $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ (σ^* is one of the most general unifiers implementing the intersection of the previous step) the left hand of the inclusion is empty, we can apply $\sigma^* \mid_{\{x_G\}}$ to the substitutions on the right hand side as well. This instantiation represents the removal of all elements at the right hand side of the inclusion for which the projection to x_G yields a value which does not appear as a value of x_G at the left hand of the inclusion. Therefore this is only to replace for some of the projections the inclusion $\emptyset \subseteq M$ by $\emptyset \subseteq \emptyset$ which does not influence the truth value of this inclusion. For all other projections (those T which are ground instances of $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G))))$ the particular inclusion is not changed at all by this instantiation. However the instantiation has the advantages that 1.) again the values on x_G are identical at both hands of the inclusion and 2.) it might appear that the value on x_G becomes a ground term which allows us to apply already discussed methods to solve the remaining problems (see next step).

Studied problem: Compute a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C}}, \mathcal{P}^-_{\sigma_O,\mathcal{C}}]$ for the ground terms $T \in \mathrm{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\mathrm{Ground}(\mathcal{S}) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\bigcup_{x \in \xi(t)} \mathcal{C}_{x,p}) \mid_{[x_G \longrightarrow T]}$$

where

$$\mathcal{S} = \{\sigma_O \gamma_C \sigma^* \mid \sigma^* = MGU(\{\sigma_O \gamma_\mathcal{C}\} \cup \bigcup_{x \in \xi(t)} \{\sigma_x\}) \text{ and } \forall x \in \xi(t):$$

$$\sigma_x \in compl(\mathcal{C}_{x,\overline{p}})\}$$

Approach: Compute for every $\sigma^* \in \mathcal{S}$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*}]$ for all $T \in \text{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\text{Ground}(\sigma_O\gamma_\mathcal{C}\sigma^*)\mid_{[x_G\longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x\in\xi(t)} \{\sigma(\sigma^*\mid_{\{x_G\}}) \mid \sigma \in \mathcal{C}_{x,p}\})\mid_{[x_G\longrightarrow T]}$$

$$\mathcal{P}^+_{\sigma_O,\mathcal{C}} := \bigsqcap_{\sigma^*\in\mathcal{S}} \mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*}$$
$$\mathcal{P}^-_{\sigma_O,\mathcal{C}} := \bigsqcup_{\sigma^*\in\mathcal{S}} \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*}$$

Justification: $\bigcup M_i \subseteq N$ iff all $M_i \subseteq N$; Instantiation of right hand sides does not influence truth value of inclusions.

Open problem: Compute for a substitution σ^* a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*}]$ for all $T \in \text{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\text{Ground}(\sigma_O\gamma_\mathcal{C}\sigma^*)\mid_{[x_G\longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x\in\xi(t)} \{\sigma(\sigma^*\mid_{\{x_G\}}) \mid \sigma \in \mathcal{C}_{x,p}\})\mid_{[x_G\longrightarrow T]}$$

9th step: Handling ground instances

If it happens that $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ is a ground term, then the parameter variable has disappeared in the $\mathcal{C}_{x,p}$ as well. Therefore we can solve the remaining inclusion problem by complementing the right hand in exactly the same way we did it for the $\mathcal{C}_{x,\overline{p}}$ in step 7 (now the $\mathcal{C}_{x,p}$ play the role of $\mathcal{C}_{x,\overline{p}}$). The only difference is, that after this process the right hand of the inclusion problem is guaranteed to be empty. That is, if after intersecting the output unifier with the complements there is still a substitution left, then the inclusion is finally proved to be not valid for the particular ground term addressed by the ground value of the global variable.

Studied Problem: Compute for a substitution σ^* a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*}]$ for all $T \in \text{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\text{Ground}(\sigma_O\gamma_\mathcal{C}\sigma^*)\mid_{[x_G\longrightarrow T]} \subseteq \text{Ground}(\bigcup_{x\in\xi(t)} \{\sigma(\sigma^*\mid_{\{x_G\}}) \mid \sigma \in \mathcal{C}_{x,p}\})\mid_{[x_G\longrightarrow T]}$$

in the case that $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ is a ground term.

Approach: Proceed as in step 7 by considering the $\mathcal{C}_{x,p}$ to play the role of the $\mathcal{C}_{x,\overline{p}}$. If the set $\mathcal{S}$ obtained for the left hand of the inclusion is empty then set

$$\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*} := \{\gamma_\mathcal{C}(\sigma_O(x_G))\}$$

$$\mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*} := \emptyset$$

and if $\mathcal{S}$ is not empty then set

$$\mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*} := \{\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))\}$$

$$\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*} := \{\gamma_\mathcal{C}(\sigma_O(x_G))\} \sqcap \mathrm{compl}(\{\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))\})$$

Justification: As in step 7, since parameter variable disappeared;
if $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ is a ground term then the inclusion concerns only the projection to this particular term; final evaluation of inclusion with empty right hand.

Open problem: As above, in the case that $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ is not a ground term.

10th Step: Project the problem to a single $x \in \xi(t)$

The right hands of the inclusion problems still to be considered are formed by the $\mathcal{C}_{x,p}\sigma^*$ for $x \in \xi(t)$ (with $\mathcal{C}_{x,p}\sigma^*$ we denote the set of all $\sigma^* \mid_{\{x_G\}}$–instances of elements of $\mathcal{C}_{x,p}$). With the next observation we simplify the problem once more where in every part of the original problem only one single $\mathcal{C}_{x,p}\sigma^*$ need to be considered. For this purpose assume the parameter variable which appears by the current case assumption in $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ to be fixed, say to one of its ground instances T. This means that the variable in $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$ is considered as ground substituted, say by σ_T. When we assume this σ_T to be applied wherever it appears in the currently considered inclusion, then the instantiated inclusion problem appears to be exactly the projection of the original inclusion problem to all substitutions which map x_G to T. Therefore we may focus further considerations mainly on $\xi(t)$.

Since according to previous observations every substitution in some $\mathcal{C}_{x,p}\sigma^*$ restricts only one single variable of $\xi(t)$, and $\gamma_\mathcal{C}$ as well as σ^* concern only variables which have been previously restricted, the set of substitutions (on $\xi(t)$) represented by a $\mathcal{C}_{x,p}\sigma^*$ can be written as $C_i = E_1 \times E_2 \times \cdots \times E_{i-1} \times M_i \times E_{i+1} \times \cdots \times E_n$ where we assume that the variables contained in $\xi(t)$ are numbered consecutively, x is the i–th one and E_j are the sets of *all* ground terms of the sort fitting to the j–th variable. M_i is some subset of E_i expressing the possible restriction of this variable value (of course M_i is different for every σ_T). The right hand of the current inclusion problem is formed by the union of the C_i for $1 \leq i \leq n$.

The left hand of the inclusion origins from unifying the output unifier σ_O after its instantiation with $\gamma_\mathcal{C}$ with elements of $\mathcal{C}_{x,\overline{p}}$. The elements of $\mathcal{C}_{x,\overline{p}}$ map elements of $\xi(t)$ to ground terms or terms containing a unique variable (we have discussed this earlier). Therefore in $\sigma_O \gamma_{\mathcal{C}} \sigma^*$ there is still at most one variable in $\xi(t)$ in the value of which the same variable appears as in its value of x_G. The other variables are still bound to ground terms or terms containing a unique variable different from the one appearing in the value on x_G. Thus the set of ground substitutions on $\xi(t)$ forming the left hand of the inclusion can be written as $N_1 \times \cdots \times N_n$, where all N_j except the one belonging to the variable which appeared in T_O equal the corresponding E_j while the remaining N_i is some subset of E_i, also dependent on σ_T. For sets representable in this way the following lemma can be applied:

Lemma 6 *Let $n \leq 1$. Let E_i $(1 \leq i \leq n)$ be the set of* all *ground terms of a sort s_i. Consider further sets M_i and N_i (both subsets of E_i correspondingly for $1 \leq i \leq n$). Then it holds $N_1 \times \ldots N_n \subseteq \bigcup_{i=1}^{n} E_1 \times \cdots \times E_{i-1} \times M_i \times E_{i+1} \times \cdots \times E_n$ if and only if $N_i \subseteq M_i$ for at least one i $(1 \leq i \leq n)$.*

Proof. Let $N_{i_0} \subseteq M_{i_0}$ for some i_0. If $a_j \in N_j$ for $1 \leq j \leq n$ then $[a_1, \ldots, a_n]$ is obviously contained in $E_1 \times \cdots \times M_{i_0} \times \cdots \times E_n$ and thus in the right hand of the inclusion. If, the other way round, no N_i is included in M_i, then there is for every i a b_i such that $b_i \in N_i \setminus M_i$. The tuple $[b_1, \ldots, b_n]$ differs in the i–th component from every element of $E_1 \times \cdots \times M_i \times \cdots \times E_n$ and therefore it is not contained in any of these sets and neither in their union. □

Note that this lemma can be applied whenever we fix a certain projection of the current inclusion. Therefore we may reduce the current inclusion problem for unifiers to the following inclusions:

$$\mathrm{Ground}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))) \subseteq \mathrm{Ground}(\{\sigma(x) \mid \sigma \in \mathcal{C}_{x,p}\sigma^*\})$$

Since in all these new inclusions there is still the variable of $\mathrm{var}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G))))$ appearing at the right hand, the inclusion forms still a restriction on the ground instances of x_G. If one of these inclusions holds for some $x \in \xi(t)$ under the assumption that the value of x_G is fixed to T then the projection to $[x_G \longrightarrow T]$ of the original inclusion holds as well. Otherwise, if the inclusion holds for none of the $x \in \xi(t)$, then the original inclusion does not hold as well. Therefore we have now an inclusion problem for *terms* rather then for substitutions.

Studied problem: Compute for a substitution σ^* a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*}]$ for all $T \in \mathrm{Ground}(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\mathrm{Ground}(\sigma_O \gamma_{\mathcal{C}} \sigma^*) \mid_{[x_G \longrightarrow T]} \subseteq \mathrm{Ground}(\bigcup_{x \in \xi(t)} \{\sigma(\sigma^* \mid_{\{x_G\}}) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{[x_G \longrightarrow T]}$$

in the case that $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ is not a ground term.

Approach: Compute for all $x \in \xi(t)$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}]$ for all $T \in \text{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\text{Ground}(\sigma^*(\gamma_\mathcal{C}(\sigma_O(x)))) \mid_{\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

$$\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*} := \bigsqcup_{x\in\xi(t)} \mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}$$

$$\mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*} := \bigsqcap_{x\in\xi(t)} \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}$$

Justification: Lemma 6

Open problem: Compute for all $x \in \xi(t)$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}]$ for all $T \in \text{Ground}(\gamma_\mathcal{C}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\text{Ground}(\sigma^*(\gamma_\mathcal{C}(\sigma_O(x)))) \mid_{\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

11th step: Handling the remaining inclusions

The remaining problem is an inclusion of terms. However, as the restriction $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G))) = T$ suggests, this problem cannot be approached by the usual symbolic set algebra, since this set algebra cannot be applied to set representations which contain arbitrary but fixed parameters (which is exactly the semantics of the above restriction). Nevertheless the problem is small enough to be solved now. Note that *all* terms representing the right hand of the inclusion contain the variable which also appears in $\text{var}(\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G))))$, since this is exactly the sorting criterion to form $\mathcal{C}_{x,p}$. There is really such a variable; this is assured by the current case assumption. Denote this variable by x^*. That means that every projection of the inclusion according to $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G))) = T$, fixes the right hand of the inclusion to a *finite* set of ground terms, which in the case that T is a ground instance of $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x)))$, say by a ground substitution ρ^* for x^*, equals $\{\rho^*(\sigma^* \mid_{\{x_G\}} (\sigma(x))) \mid \sigma \in \mathcal{C}_{x,p}\}$ and which is empty when T is not a ground instance of $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$.

On the other hand, the projection according to $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G))) = T$ transforms the left hand of the inclusion to at most one single ground term if $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x)))$ is a ground term or it contains x^*. Thereby there remains a term iff T is a ground instance of $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x_G)))$.

Otherwise, if $\sigma^*(\gamma_\mathcal{C}(\sigma_O(x)))$ contains any other variable, it represents a *set* of terms which is finite if and only if the sort of this variable is finite.

Consider three cases according to the variable appearing in $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$. A) This variable does not exist or equals x^*; B) This variable is different from x^* but has a finite sort; C) This variable is different from x^* but has an infinite sort.

Ad A.) As discussed above, every projection according to $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G))) = T$ transforms the left hand of the inclusion to a singleton or empty set. The set is empty iff T is not a ground instance of $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$. Those T shall belong to the "+" part of our partition.

If T is a ground instance of $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ then we get the substitution ρ^* for x^* as above. The inclusion concerns then the ρ^*–instance of $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ on the left hand of the inclusion and the ρ^*–instances of the terms appearing at the right hand. The inclusion holds if and only if the term at the left hand is equal to one of the terms at the right hand. That is, the inclusion holds if and only if ρ^* is a unifier of $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ and one of the $(\sigma^* \mid_{\{x_G\}})(\sigma(x))$ where $\sigma \in \mathcal{C}_{x,p}$. The set of *all* such ρ^* is consequently represented by the most general unifiers of the pairs of considered terms. For terms T which are ground instances of $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ where the corresponding ρ^* is not a unifier of one of the unification problems, the inclusion which is currently under consideration does consequently not hold. Therefore case A can be summarized as follows:

Studied problem: Compute for some $x \in \xi(t)$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}]$ for all $T \in \text{Ground}(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\text{Ground}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))) \mid_{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

in the case that $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ is a ground term or a term containing the variable in $\text{var}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G))))$.

Approach:

$$\begin{aligned} \mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x} &:= (\{\gamma_{\mathcal{C}}(\sigma_O(x_G))\} \sqcap \mathit{compl}(\{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))\})) \\ &\sqcup \{\mathit{MGU}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x))), (\sigma^* \mid_{\{x_G\}})(\sigma(x))) \mid \sigma \in \mathcal{C}_{x,p}\} \\ \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x} &:= \{\gamma_{\mathcal{C}}(\sigma_O(x_G))\} \sqcap \mathit{compl}(\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}) \end{aligned}$$

Justification: First part of $\sqcup$ in "+" part: left hand of inclusion is empty; Second part of $\sqcup$ in "+" part: x^* is the only variable still appearing and every ground instance of x^* corresponds to one of the projections of the inclusion.

Open problems: Cases B and C.

Ad B: Due to the finiteness of the sort of x Ground$(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x))))$ is finite as well and we can study the inclusion by considering every element of this set on its own. Every of the resulting problems fits to case A of this step. Therefore we can trace back the partition in case B to several partitions according to case A.

Studied Problem: Compute for some $x \in \xi(t)$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}]$ for all $T \in$ Ground$(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\text{Ground}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))) \mid_{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

in the case that $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ contains a variable y which is different from the one in var$(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G))))$ and corresponds to a finite sort.

Approach: Compute for every ground term T' of the sort of y a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x,T'}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x,T'}]$ for all $T \in$ Ground$(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))[y \longrightarrow T']\} \mid_{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

which can be done as in case A;

$$\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x} := \bigsqcap_{T'} \mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x,T'}$$

$$\mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x} := \bigsqcup_{T'} \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x,T'}$$

Justification: A set is included in another one iff all elements are.

Open problem: Case C.

Ad C. This may be the simplest case. If the sort of the variable appearing in $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ is infinite and differs from the one in $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$, then the set of its ground instances is infinite, whatever restrictions apply for $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$. As discussed earlier, the right hand of the inclusion represents for every projection a finite set. Therefore we are considering an inclusion of an infinite set in a finite one which does never hold.

Studied Problem: Compute for some $x \in \xi(t)$ a partition $[\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x}, \mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x}]$ for all $T \in$ Ground$(\gamma_{\mathcal{C}}(\sigma_O(x_G)))$ according to the condition

$$\begin{aligned} &\text{Ground}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))) \mid_{\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G)))=T} \\ \subseteq\ &\text{Ground}(\{(\sigma^* \mid_{\{x_G\}})(\sigma(x)) \mid \sigma \in \mathcal{C}_{x,p}\}) \mid_{\sigma^*(\sigma(x_G))=T} \end{aligned}$$

in the case that $\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x)))$ contains a variable y which is different from the one in $\mathrm{var}(\sigma^*(\gamma_{\mathcal{C}}(\sigma_O(x_G))))$ and corresponds to an infinite sort.

Approach:

$$\mathcal{P}^+_{\sigma_O,\mathcal{C},\sigma^*,x} := \emptyset$$
$$\mathcal{P}^-_{\sigma_O,\mathcal{C},\sigma^*,x} := \{\gamma_{\mathcal{C}}(\sigma_O(x_G))\}$$

Justification: An infinite set can never be included in a finite one.

No open problem.

After several transformations the initial partition can be divided into a certain number of smaller ones which finally can be solved. It might appear to the reader that this approach is far too complicated to compete with the low–level approach. In the case of finite specifications we want to leave this question open. However in the case of specifications for infinite data types there is no low–level approach at all. In contrast, all manipulations which have been done throughout the past 11 steps are always finite, even if the specified data type is infinite. This is a strong argument in favor of the symbolic approach. For finite specifications we express a certain hope that in many cases only a small amount of the transformations is to be done for which we have prepared ourselves above. That is, with a certain probability one part of the partition is empty and the whole symbolic family can pass the check procedure without splitting.

It might also appear that the distinction of cases B and C during the last step cannot be done automatically. However, in the case of equation–free unary specifications we can decide the finiteness of a sort using the signature graph of the involved specification. The nodes of this graph are the sorts of the specification while for every unary operation symbol mapping from s_1 to s_2 there is an arc from s_1 to s_2.

If in this graph a sort is not reachable from a sort for which constant symbols are specified, then there are no ground terms for this sort at all. We have excluded this case in Definition 2. Thus a sort has finitely many ground terms if and only if it is *not* reachable from *any* circle in this graph. Since the signature graph is finite the decision can be drawn automatically.

8 Expanding symbolic families of place sets

In this section we study, how new places are added to a symbolic family $\mathcal{D}$ of place sets. For the purpose of narrowing the number of choices we add only pre–places of a transition which has firing modes violating the siphon property for some of the represented siphons. This transition can be recognized by the particular check–module which sent the symbolic family of non–siphons to the expand module. Let t be this transition. Furthermore, every place we add to a place set D which appears in the interpretation of a symbolic family of place sets should be an input place of a firing mode of t which has output places in $\mathcal{D}$. Since one of these input places *must* belong to a siphon which includes D, we can assure that we compute all minimal siphons when we are able to traverse all input places of t. In order to process an expansion simultaneously, it would be comfortable to have a symbolic description τ of firing modes of t such that for every set represented by $\mathcal{D}$ this representation provides firing modes which put tokens on this particular set. Thereby, to establish a dependency between the selected set and the symbolic representation of the firing mode, an appearance of the global variable of $\mathcal{D}$ could be accepted.

Fortunately such a τ exists and we can generate it from calculations which we did in the *check*–part of our algorithm. Remember the approach of steps 3 and 4 there. In step 4 we considered a specific most general output unifier σ_O of $\mathcal{D}_{old}$ and t, say according to $T_\flat$ such that $p_\flat.T_\flat$ appears in $\mathcal{D}$ and a T_O which appears in $\lambda([t, p_\flat])$. Consider the symbolic partition of the ground values of x_G for $\mathcal{D}_{old}$ computed there. Since all the elements of $compl(\sigma_O(x_G))$ are contained in $\mathcal{P}^+_{\sigma_O}$, all ground terms which are represented by terms appearing in $\mathcal{P}^-_{\sigma_O}$ are ground instances of $\sigma_O(x_G)$. However, if all ground instances of a term T are also ground instances of another term T' then (at least for our restricted class of specifications) T itself is an instance of T'. Therefore for every term T appearing in $\mathcal{P}^+_{\sigma_O}$ there is a substitution γ on $\mathrm{var}(\sigma_O(x_G))$ such that $T = \sigma_O\gamma$. Since $\mathcal{P}^-$ is simply the union of all $\mathcal{P}^-_{\sigma_O}$ this property still holds for every element in $\mathcal{P}^-$. But according to the procedure outlined in the previous section a symbolic family of place sets $\mathcal{D}$ which is passed to the *expand*–module, equals $\mathcal{D}_{old}[x_G \longrightarrow T]$ for some $T \in \mathcal{P}^-$, that is $\mathcal{D}_{old}[x_G \longrightarrow \gamma(\sigma_O(x_G))]$ for a most general output unifier σ_O of $\mathcal{D}_{old}$ and a substitution γ. Keeping in mind the definition of most general output unifiers, there is a $p_\flat.T_\flat$ appearing in $\mathcal{D}_{old}$ and a $T_O \in \lambda([t, p_\flat])$ such that $T_O\sigma = T_\flat\sigma$ for some σ such that σ_O is its restriction to $\{x_G\} \cup \xi(t)$. Consequently it holds $T_O\sigma\gamma = T_\flat\sigma\gamma$. Since $T_\flat\sigma\gamma$ is an instance of $T_\flat[x_G \longrightarrow \gamma(\sigma(x_G))]$ which appears in $\mathcal{D}$ this means

Corollary 3 *For every ground substitution δ of the global variable of $\mathcal{D}$, every ground instance β of $(\sigma_O\gamma)\mid_{\xi(t)} \delta$ is a firing mode of t which puts tokens on $UNF(\mathcal{D}\delta)$.*

In the sequel we denote $(\sigma_O\gamma)\mid_{\xi(t)}$ by τ and assume only the property of Corollary 3 for it. This substitution may contain the global variable of $\mathcal{D}$, since it origins from instantiating the global part of the output unifier σ_O to the term which substituted the global variable of $\mathcal{D}_{old}$. For all other variables in $\mathrm{var}(\tau)$ we may assume that they are different from variables appearing in $\mathcal{D}$.

All we have to do is to traverse all choices of choosing a pre–place $p_\sharp$ of t and a term $T_\sharp$ appearing in $\lambda([p,t])$ for which we have to append $p_\sharp.T_\sharp\tau$ to the symbolic family $\mathcal{D}$ which is currently under consideration.

Definition 16 (Expansion of symbolic families) *Let* $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, \{x_G\}, X_L]$ *be a symbolic family of place sets,* t *a transition and* τ *a substitution on* $\xi(t)$ *where* $\mathrm{var}(\tau) \cap X_L = \emptyset$*, while the global variable* x_G *may be contained in* $\mathrm{var}(\tau)$*. Then for every* $p_\sharp \in Ft$ *and every* $T_\sharp \in \lambda([p_\sharp, t])$ *the symbolic family of places* $Expand(\mathcal{D}, t, \tau, p_\sharp, T_\sharp)$ *is defined as* $[\{p_i.T_i \mid i \in I\} \cup \{p_\sharp.T_\sharp\tau\}, \{x_G\}, X_L \cup (\mathrm{var}(\tau) \setminus \{x_G\})$*.*

Of course variables may be removed when they do not appear in the first component of the new symbolic family. If a symbolic family of place sets is passed to the *expand*–module of our global algorithm, then all obtainable expanded families (traversing the places and terms) are passed back to the *check*–module.

The following lemma justifies this expansion.

Lemma 7 *Let* $\mathcal{D} = [\{p_i.T_i \mid i \in I\}, \{x_G\}, X_L]$ *be a symbolic family of places,* t *a transition and* τ *be a substitution over* $\xi(t)$ *where* $\mathrm{var}(\tau) \cap X_L = \emptyset$*, while the global variable* x_G *may be contained in* $\mathrm{var}(\tau)$*.*

For every ground substitution δ *of* $\{x_G\}$*,* $UNF(Expand(\mathcal{D}, t, \tau, p_\sharp, T_\sharp)\delta)$ *includes* $UNF(\mathcal{D}\delta)$ *and the elements not contained in* $UNF(\mathcal{D}\delta)$ *are input places of transition modes which are ground instances of* $\tau\delta$*. Furthermore for every firing mode of* t *which is a ground instance of* $\tau\delta$ *an input place is contained in* $UNF(Expand(\mathcal{D}, t, \tau, p_\sharp, T_\sharp)\delta)$*.*

Proof. The inclusion of a place set and its expansion is obvious. Furthermore the new elements are input places of t, since they are made from input places of t and terms appearing at the corresponding arc. Since we applied τ to the input terms the elements are in fact input places of firing modes which are ground instances of τ, and due to the unfolding rule for symbolic families of place sets they are even ground instances of $\tau\delta$. For a ground instance $\tau\delta\rho$ of $\tau\delta$, of course the place $p_\sharp.T_\sharp\delta\rho$ which is contained in the unfolding of the expanded set is an input place of $t.\tau\delta\rho$. □

This lemma does only partly justify the symbolic approach. Though it assures that after the expansion input places for all firing modes addressed by τ are contained in the new set, it does not give evidence that 1.) this expansion is minimal and 2.) all combinations for expanding a set of places which would be tried at the low level are covered.

If the term $T_\sharp\tau$ is a ground term or the global variable appears in it, then for every place set represented by $\mathcal{D}$ only one single place is appended by the expansion procedure. Therefore in these two cases the low level procedure is copied in its very detail and nothing remains to show. In the remaining case the situation is more difficult: If we append a term which contains a variable different from the global one (and can usually be substituted by more than one ground term), then the expansion appends more than one place color at once to every set represented by $\mathcal{D}$, namely all place colors of $p_\sharp$ which are ground instances of the appended term $T_\sharp\tau$.

The next lemma provides some minimality criterion. It states that certain sets of input places must be contained in every siphon which includes a place set represented by $\mathcal{D}$.

Lemma 8 *Let $\mathcal{D}$ be a symbolic family of place sets and δ a ground substitution of the global variable. Let t be a transition and $\tau\delta$ a substitution on $\xi(t)$ such that every ground instance of $\tau\delta$ is a firing mode where t has output places in $UNF(\mathcal{D}\delta)$ Consider for every $x \in \mathrm{var}(\tau\delta)$ the set $\mathcal{T}_x$ of those terms T which appear at input arcs of t and where $x \in \mathrm{var}(T\tau\delta)$. Let $\mathcal{T}_\epsilon$ be the set of terms which appear at input arcs of t where $T\tau\delta$ is a ground term.*

Then every siphon D which includes $UNF(\mathcal{D}\delta)$ contains some element $p.T\tau\delta$ where p is an input place of t and $T \in \mathcal{T}_\epsilon$, or there exists an element $x \in \mathrm{var}(\tau\delta)$ such that D contains for every ground substitution ρ of x an element $p.T\tau\rho$ with $T \in \mathcal{T}_x$.

Proof. According to the siphon property D contains an input place for every firing mode of every transition t which puts tokens on D. Especially it contains input places of all firing modes of t represented by τ. The input places of a firing mode β are given by the $p.T\beta$ where p is an input place of t and T appears in $\lambda([p,t])$.

Assume that D does not contain a $p.T\tau\delta$ for $T \in \mathcal{T}_\epsilon$ and for every $x \in \mathrm{var}(\tau\delta)$ there is at least one ground term T_x such that D does not contain $p.T\tau\delta[x \longrightarrow T_x]$ for any $T \in \mathcal{T}_x$ and the corresponding place p. Consider the firing mode $\tau\delta(\bigcup_{x\in\mathrm{var}(\tau\delta)}[x \longrightarrow T_x])$ which is obviously a ground instance of $\tau\delta$.

By construction none of the input places of the mode β of t are contained in D. This contradicts the assumption that D is a siphon. $\square$

The only case in this lemma which is not covered by the expansion procedure proposed above is that some $\mathcal{T}_x$ contains more than one element and D contains instances of both. In fact, the case that a siphon which includes D contains instances of T_ϵ is covered by the expansion of $\mathcal{D}$ with terms whose τ–instances are ground terms or contain the global variable of $\mathcal{D}$. In the case that the siphon contains all instances of some T_x, and this T_x contains only one element T, then the siphon must contain all ground instances of T. This is covered by the expansion of $\mathcal{D}$ with T, and by appending the variable of T to the set of local variables.

If one of the $\mathcal{T}_x$ contains more than one element, then a siphon including an element represented by $\mathcal{D}$ does not necessarily include elements of either of the computed expansions. With our expansion procedure we simply forget about the cases where we remove the violations of the siphon property by appending input places corresponding to some non singleton $\mathcal{T}_x$ and where these input places are not represented by one single term in $\mathcal{T}_x$, but spread over different of them. Figure 3 exhibits an example where this situation happens.

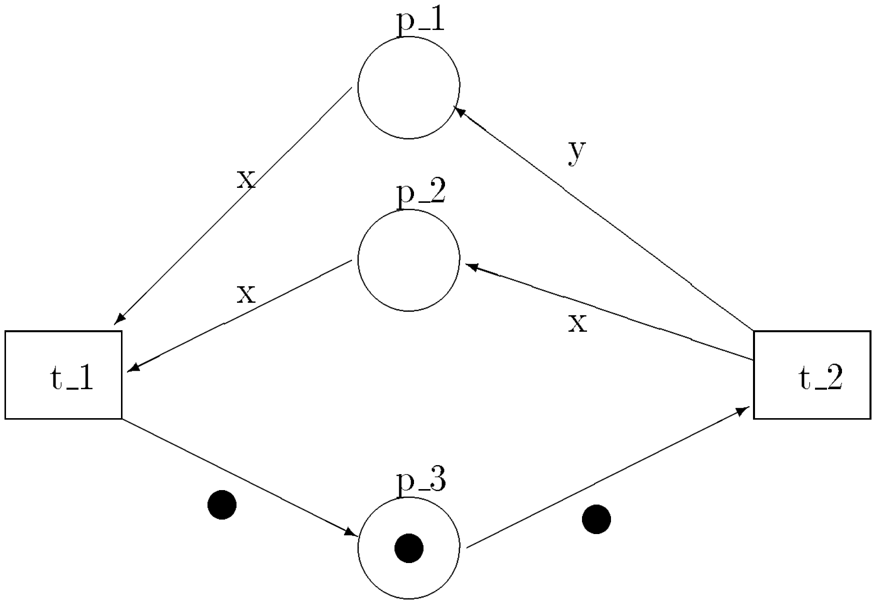

Figure 3: Net with more than countable many minimal siphons

If we take $\mathcal{D} = [\{p_3.\bullet\}, \emptyset, \emptyset]$ then obviously t_1 puts tokens into the only place represented by $\mathcal{D}$. A substitution representing the violations is for instance $[x \longrightarrow x_1]$. Therefore $\mathcal{T}_{x_1}$ contains two elements, x at the arc from p_1 and x at the arc from p_2. The two possible expansions of $\mathcal{D}$ are $[\{p_3.\bullet, p_1.x_1\}, \emptyset, \{x_1\}]$ and $[\{p_3.\bullet, p_2.x_1\}, \emptyset, \{x_1\}]$. Both of them represent siphons: $\{p_3.\bullet\} \cup \{p_1.T \mid T \in T_{\Omega,\psi(p_1)}\}$ and $\{p_3.\bullet\} \cup \{p_2.T \mid T \in T_{\Omega,\psi(p_1)}\}$. Unfortunately there are more of them: for every set M of terms of sort $\psi(t)$ the set $\{p_3.\bullet\} \cup \{p_1.T \mid T \in M\} \cup \{p_2.T \mid T \in T_{\Omega,\psi(p_2)} \setminus M\}$ is a minimal siphon of the exhibited net. If there are infinitely many ground terms of sort $\psi(p_1)$ then the net has more than countable many different minimal siphons, since the number of different subsets of an infinite set is not countable. Only two of them are computed by the proposed expansion procedure. However, no algorithm can compute a symbolic representation of the minimal siphons of this net, since every symbolic family of place sets can represent at most countable many siphons. Even if we restrict ourselves to an enumeration of the set of minimal siphons, we can compute at most a countable number of representations of countable many siphons, which still covers only countable many siphons. Therefore there cannot be any complete computation of the set of siphons, of course unless we use a different symbolic representation for siphons which probably would make the approach even more difficult.

However, there are certain ways to avoid this problem as long as possible. In the case of a finite sort the problem can be solved. We only have to expand the symbolic siphon by all ground instances of an input arc term rather than by a term containing a local variable. This way a huge amount of combinatorics comes into play, but the same combinatorics would happen in the low–level procedure as well. Furthermore the symbolic family of place sets may still contain variables elsewhere which keep the possibility of processing place sets simultaneously (in fact, the global variable is never touched by the reported problem).

In the case of infinite sorts we can try to expand the symbolic family first according to other violations of the siphon property. This way it may happen that enough input places of t are appended to $\mathcal{D}$ to make the troubling violation disappear. In fact, if some input places of the troubling transition are forced to be contained in a siphon anyway, by arguing with other violations then we do not need to try all the other combinations of input places, since we are mainly interested in minimal siphons.

If the problem cannot be solved by either of these strategy, the only thing left to do is unfortunately to switch off the application of the implications which can be drawn from the property "every minimal siphon contains a marked trap". Without a complete representation of the minimal siphons we have no chance to verify this property.

9 Conclusion

Though for conceptual reasons not always a finite representation of all siphons of a net can be achieved, the failure rate of the calculation should be acceptable, when we take into consideration that we deal with high–level nets inscribed with infinite color domains. Since the restricted class of algebraic specifications allows to model counter automata, most behavioral properties, among them liveness, are undecidable. From this point of view the presented approach is a considerable success, even taking into consideration its incompleteness as discussed for the expansion module, but also the possible non–termination which has not been mentioned so far and finally the grey zone left by the weaker implications from siphons and trap to the system's behavior.

However, compared with previous expectations, at least to the author's view, the restrictions which are necessary to make the approach work are surprisingly hard. Thereby there are partially conceptual reasons for the restrictions. At least the ability of computing complements of term based set representations is essential for the approach. Consider for example the net depicted in Figure 4.

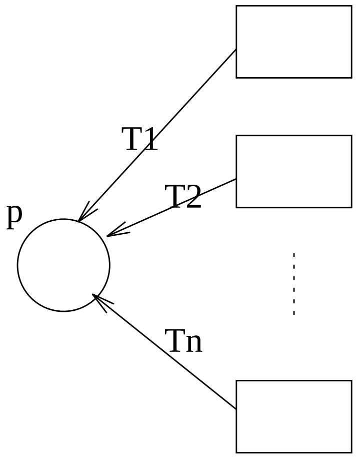

Figure 4: Difficult situation for siphon calculation

The siphons of this net are exactly those colors of places which are *not* contained in Ground($\{T_1, \ldots, T_n\}$). Therefore every complement problem on terms can be represented as a siphon calculation problem. Thus the ability of calculating symbolically complements of sets represented by terms is unavoidable for any symbolic siphon calculation approach.

Unfortunately the ability of calculating siphons might not necessarily be sufficient for a siphon calculation procedure, though we cannot demonstrate this expectation with another simple example. However, particularly the justifications for the approach to partition families of place sets into families of siphons and families of non–siphons involved a great deal of arguing with the restrictions of the net class. Especially we made use of the unarity of the specification when we simplified the inclusion of input and output unifiers based on the observation that in every term at most one variable can appear. In arbitrary specifications global and local variables can mix up in one and the same term and a totally different approach would be necessary to partition the set of global variables (there is no longer at most one global variable) according to the inclusion of unifiers.

The absence of equations made it possible to decide whether a sort consists of finitely or infinitely many elements. Having equations, this decision can no longer be done only regarding the signature graph.

Thus both parts of the restrictions appear in addition to the minimum requirement to be able to compute complements symbolically. Therefore there is only few hope that the approach can be generalized significantly. Of course we can start the same sort of calculation as presented here for a net specified over an arbitrary specification and hope that we do not run into a troubling case. But the failure rate is then much greater than for unary equation–free specifications.

An interesting observation can be done by comparing the two parts of the calculation procedure. Though the check–part is much more involved than the expand–part, as the structure of the corresponding sections might suggest, and though this is the part where we benefit most from the restrictions applied to the specifications, this is the part which is always able to compute satisfactory results. In comparison, the expand–part would also work for more general specification (therefore it can fortunately be used to propose fixing of violation of user–predicted siphons in the verification calculus ([Sch96b])). Nevertheless it is the expand–part where principle constraints to the feasibility apply even for the restricted class of specification. This might serve as an example, that the symbolic approach to high level Petri net analysis is always full of surprises.

References

[Com72] F. Commoner. Deadlocks in Petri nets. Technical report, Applied Data Research Inc. Wakefield, Massachussetts, 1972.

[EM85] H. Ehrig and B. Mahr. *Fundamentals of Algebraic Specifications*, volume 1 of *EATCS Monographs on Theoretical Computer Science 6*. Springer, Berlin, 1985.

[Rei91] W. Reisig. Petri nets and algebraic specifications. *Theoretical Computer Science*, 80:1–34, 1991.

[Sch96a] K. Schmidt. Siphons, traps and high–level nets with infinite color domains. *In preparation*, 1996.

[Sch96b] K. Schmidt. Symbolic verification of siphons and traps for algebraic Petri nets. *In preparation*, 1996.

[Sta90] P. Starke. *Analyse von Petri–Netz–Modellen.* B.G. Teubner Stuttgart, 1990.