Systems for open terms: an overview

Citation for published version (APA):

Document status and date:
Published: 01/01/2001

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Systems for Open Terms: An Overview

by

G.I. Jojgov

ISSN 0926-4515

All rights reserved

editors: prof.dr. J.C.M. Baeten
        prof.dr. P.A.J. Hilbers

Reports are available at:
http://www.win.tue.nl/win/cs

Computer Science Reports 01/03
Eindhoven, April 2001
Systems for Open Terms: An Overview

G. I. Jojgov
(Георги И. Жожгов)
Eindhoven University of Technology
GJojgov@win.tue.nl

March, 2001
Abstract

In this paper we make an overview of some existing systems of open (incomplete) terms including ALF, Typelab, OLEG, \( \lambda \Pi \ell \), Automath, \( \lambda c \) and \( \lambda s \).
Chapter 1

1.1 Introduction

The increasing complexity and scope of problems being solved using formal tools combined with the state-space explosion observed in many of them has led to increased interest in computer-assisted interactive reasoning for problems for which full automation is unfeasable or impossible. One particular area of interest is interactive theorem proving. The idea is that a human guides the computer by providing information which cannot be automatically deduced or which will help prune the search tree. The computer may still be allowed to automatically solve easy subproblems for which efficient automation is possible.

When talking about (interactive) theorem proving we always have in mind some formal framework in which we express our goals and the input data. Usually this is some elaborate theory involving notions like formula, deduction, proof, etc. This can be for example propositional logic, predicate calculus, $\lambda$-calculus or type theory. The predominant part of the currently available systems for interactive theorem proving are based on type theory. There we present formulas as types and their proofs as terms of these types using the Curry-Howard isomorphism.

Due to its incremental nature, in the interactive process of proof construction we need to handle incomplete proofs which in a type theory framework correspond to incomplete terms. These are terms that contain "holes" corresponding to (the parts of) the proofs that are not constructed yet. We call such terms also open.

An interactive theorem prover needs to guarantee at least soundness of the process of proof construction. This means that the terms constructed correspond to real proofs in the underlying formal system. The failure of soundness when open terms are being treated in a naive way was the reason why research in the field of open terms was initiated. This theory helps us model the intermediate steps in the process of interactive theorem proving thus allowing us to reason about this process in a formal way. Such a theory also gives theoretical justification to the implementations of different tools which in many cases in the past were based on intuitive ideas only.

Another potential benefit of the theory of open terms is that it may give rise to new applications by giving an abstract overview of the interactive process which is not being obscured by implementation details and decisions.

This report is organized in several parts. First we discuss the points of interest for a system of open terms which can also play the role of classification criteria.

Then we discuss in brief the following systems: ALF, Typelab, OLEG, AII$\zeta$, the context calculus $\lambda$c, Automath and the $\lambda$S$\tau$-calculus.
We conclude with a discussion and some observations.
We assume the reader is acquainted with basic type theory and logic.

1.2 Issues of Interest and Classification Criteria

In this section we briefly discuss what we will be interested in when considering a system for open terms, why a particular issue is of importance and how the systems can be classified according to it. Of course, there may be dependencies and not all the topics discussed below are relevant to all systems.

**Representation** The most important issue is how open terms are represented in the formal system. Generally there are two main approaches: Historically first appeared those based on some kind of *ad-hoc datastructures*. These can be different kind of trees or graphs, lists of the tactics applied and so on. The advantages of such representations are usually in the field of performance and ease of implementation. However this makes the representations of open terms tool-dependent (in some cases even worse: version-dependent).

On the other hand we have the second approach, which calls for extending the language of terms with so-called *metavariables*. These variables are intended as placeholders for terms that are not yet constructed. This gives us a tool-independent presentation of open terms which is based on formal rules. As a consequence the developments in a tool can be compared to the theoretical results and expectations which improves the reliability of the tool. Currently metavariables are de facto the standard for representing open terms.

There is however no widely agreed upon way to introduce metavariables in general. There are numerous solutions for various systems, but they are quite different and in most cases incompatible.

**Explicit Substitutions** Furthermore, systems based on metavariables can be subdivided into two classes based on the way they treat computations with open terms. One of them is a large class of systems of metavariables and *explicit substitutions*. In the other class we find systems which in some way avoid the complications incurred by the cross-influence between metavariables and computation.

**Underlying System** We may classify the systems according to their *underlying system* (i.e. the system for which we want to introduce open terms). This system can be for example propositional logic, predicate logic, higher-order logic. For calculi based on type theory this classification can be done in terms of the systems of the \( \lambda \)-cube, as PTS specification and so on. There are also adaptable systems which allow the user to specify the formal framework that he or she wants to use.
Metatheoretical Properties For systems based on type theory we can investigate various relevant properties like for example confluence, strong/weak normalization, subject reduction, decidability of conversion, type inference and type checking and these properties can be used to classify the systems accordingly.

Use of de Bruijn indices Very often systems which are intended for implementation purposes differ from the ones that are purely theoretical. One such difference is for example the use of de Bruijn indices instead of named variables or a combination of the two.

1.3 Informal Use of the Metavariabes in Interactive Theorem Proving

Let us illustrate the use of metavariables in an interactive session with several small examples:

1.3.1 Metavariables standing for (proof)terms

Step 1: Initial state We are looking for term of type $A \rightarrow A$. The unknown term is denoted by $?_1$:

$$\vdash ?_1 : A \rightarrow A$$

Step 2: We decide that the unknown term is an abstraction and solve $?_1$ by $\lambda x : A.?_2$ introducing a new metavariable $?_2$. Note that $?_2$ is typed in the context $x : A$.

$$\vdash \lambda x : A.?_2 : A \rightarrow A$$
$$x : A \vdash ?_2 : A$$

Step 3: Now the task has become easy, because we have to find a term of type $A$ in context $x : A$. Of course we can choose $?_2 = x$.

$$\vdash \lambda x : A.x : A \rightarrow A$$
$$x : A \vdash x : A$$

Step 3': Note that instead of choosing to solve $?_2$ by $x$, we could have chosen another path. Alternative solution is for example $?_2 = ?_3 ?_4$. This non-determinism is usually imposed by the complexity of the type inhabitation problem, which in most of the cases is undecidable. Another reason is completeness - we want to be able to derive a judgement in any possible way, because in some cases (interactive program refinement for example) the actual term inhabiting the type is of importance$^1$.

$$\vdash \lambda x : A.?_3 ?_4 : A \rightarrow A$$

$^1$As anyone aware of the difference between the quicksort algorithm and the one checking all permutations for a sorted one will agree.
In several steps the metavariables can be solved by $\beta_3 = \lambda y : A. y$, $\beta_4 = x$ or $\beta_3 = \lambda y : A. x$, $\beta_4 = x$.

### 1.3.2 Metavariables as witnesses

Another typical case when we need to use metavariables is when we want to delay the moment when we specify the witness for an existential quantifier. Consider the following arithmetic formula:

$$\exists x \in N \forall y \in N (\forall z \in N (x + z \leq z)) \rightarrow x \leq y$$

A proof of such a formula involves finding a witness $t$ for $x$ and proving $\forall y (\forall z (t + z \leq z)) \rightarrow t \leq y$. In this case it is easy to see that $0$ can be used as a witness, but we want to have the option to delay this to a later moment. This is good for example for didactic reasons, because it usually helps to identify the reasons for choosing the specific witness by making explicit its definitional properties. Delaying the instantiation of an existential quantifier gives us also the opportunity to possibly find the witness automatically at a later stage by means of unification for example. In our case from $\forall z (x + z \leq z)$ we can deduce $x \leq 0$ and this is a clear hint that we can actually use the value $0$ for $x$.

Because a metavariable can stand for a witness in an existential statement, sometimes metavariables are also called existential variables.

**The Transitivity Example** Note however that using metavariables as means to delay the specification of witnesses occurs not only with existential statements. Consider the transitivity property for a relation $R$ on a set $A$:

$$(\text{transitive } R) : \forall x \forall y \forall z. xRy \rightarrow yRz \rightarrow xRz$$

Suppose $a$ and $c$ are elements of $A$ and we want to prove $aRc$ using the transitivity of $R$. Hence we must instantiate the universally quantified statement by taking $x := a$ and $z := c$. But what to take for $y$? At this point we may not know that and a sensible strategy would be to postpone this choice for a later moment when we can find a witness $b$ for $y$ such that $aRb$ and $bRc$. This postponement can be done by introducing a metavariable for $y$.

### 1.3.3 Computing with metavariables

In this report we will also be looking at the problem of computations with open terms. Let us give an example when metavariables are involved in a computation.

Suppose we want to prove $\exists f : N \rightarrow N \forall n : N (fn = n)$ for some type $N$. Having forgotten that identity functions exist, we introduce a metavariable $\beta_4$. 
as a witness for the existential quantifier:

\[ \vdash ?_1 : \forall n : N(\neg f(n) = n) \]

\[ n : N \vdash ?_1 : N \to N \]

for the sake of the example, we decide to refine \( \neg f \) by \( \lambda x : N. ?_2 \) introducing a new metavariable:

\[ \vdash ?_1 : \forall n : N((\lambda x : A. ?_2)n = n) \]

\[ n : N, x : A \vdash ?_2 : N \]

but now we have created a redex \( (\lambda x : N. ?_2)n \). If we decide to reduce this redex we obtain \( ?_2\{x := n\} \). Now we see that in order to have \( ?_2\{x := n\} = n \) we can choose either \( ?_2 = x \) or \( ?_2 = n \) because in both cases after the substitution is executed\(^2\) we get \( n \).

This example once again shows the benefit of postponing the moment when we actually specify the witness.

### 1.4 Main Problem to Solve When Introducing Metavariables

As we saw in the previous section, it is nice to be able to compute with open terms. In systems which have some built-in notion of computation this turns into necessity. Take for example the conversion rule in the Calculus of Constructions. It allows a type to be replaced by an equal one. In this case the introduction of metavariables forces us to define computation on open terms.

The fundamental problem is that we need to compute with something which is not known yet. Let us consider the most common form of computation – \( \beta \)-reduction and look at the following diagram:

\[
\begin{array}{c}
\frac{(\lambda x : T. ?_n )t}{?_n := x} \\
\end{array}
\]

\[
\begin{array}{c}
\frac{(\lambda x : T. x) t}{\beta} \\
\end{array}
\]

\[
\begin{array}{c}
\frac{(\lambda x : T. ?_n) t}{?_n := x} \\
\end{array}
\]

\[
\begin{array}{c}
\frac{?_n := x}{\beta} \\
\end{array}
\]

\[
\begin{array}{c}
\frac{x}{t} \\
\end{array}
\]

\(^2M\{x := N\} \) is notation for the (meta)operation of substitution of the free occurrences of \( x \) in \( M \) by the term \( N \). Here the expression \( ?_n\{x := N\} \) can only be given meaning after \( ?_n \) is instantiated by a term. Later we will see systems which give meaning to this expression before the instantiation.
We see that ignoring the effects of substitutions on open terms leads to loss of confluence or other strongly desirable properties (as subject reduction in some cases, see section 2.4).

Similar problems occur if the system has local definitions. Then computations involve expansion of definitions and elimination of dummy definitions.
Chapter 2

Open Term Systems for Theorem Proving

In this chapter we will discuss several systems that are specifically designed to facilitate interactive theorem proving.

2.1 ALF

ALF [18] is an interactive proof editor based on Martin-Löf's monomorphic type theory. It employs placeholders and explicit substitutions in order to represent open terms. The theory of open terms of ALF was developed in order to support the main feature of the tool which is allowing the user to directly work with the proof objects. This theory reduces the process of editing proof-objects to typechecking of incomplete terms.

2.1.1 Syntax of the Calculus

The terms of the target calculus (the one to which the completely constructed term should belong to) are given by the following grammar:

\[
\text{Term ::= } \text{Var} | \text{Const} | [\text{Var}] \text{Term} | (\text{Term} \text{Term}) | \text{Term Subst}
\]

Note that in an abstraction term \([x].M\) the abstracted variable \(x\) is not given a type and thus terms do not depend on types. As we see, in a term we may have explicitly given substitutions which have the following syntax:

\[
\text{Subst ::= } \{\} | \{\text{Subst, Var := Term}\} | \text{Subst Subst}
\]

Substitutions are composed as follows:

\[
\{\} \beta = \beta \\
\{\gamma, x := e\} \beta = \{\gamma \beta, x := (e \beta)\}
\]
On the level of types we distinguish between types and families of types which are given as follows:

\[
\text{Type ::= Set} | \text{Type} \rightarrow [\text{Var}]\text{Family} | \text{Family Term} | \text{Type Subst}
\]

\[
\text{Family ::= E} | [\text{Var}]\text{Type} | \text{Family Subst}
\]

So a type is either Set or a function type or an element of a family or a type with a substitution applied to it. Note that when we select a member of a family, we do this by application to a term. Therefore types may depend on terms and hence we are in a dependently typed setting.

Contexts are lists of declarations:

\[
\text{Context ::= } [] | [\text{Context}, \text{Var}: \text{Type}]
\]

The calculus has judgements dealing with terms, types, contexts and substitutions. Each of these categories may have either of the following kinds of judgements: formation judgements, typing judgements and equality judgements. Contexts as lists of declarations are used to type substitutions which are lists of variable assignments: \(\Delta \vdash \gamma : \Gamma\). The equality judgements internalize the different notions of reductions present in the system. Take for example the \(\beta\)-conversion on types. The corresponding rule is:

\[
\frac{\Gamma, x : \alpha' \vdash \alpha : \text{Type} \quad \Delta \vdash \gamma : \Gamma \quad \Delta \vdash a : \alpha' \gamma}{\Delta \vdash (\{x\} \alpha \gamma)a = \alpha\{\gamma, x := a\} : \text{Type}} \quad \beta\text{Subst}
\]

Note the presence of the substitution \(\gamma\) just outside the scope of the abstraction binder. The rule has this form because according to the rules for substitutions, a substitution cannot be distributed over abstraction and therefore substitutions generated by other redexes tend to accumulate at abstractions. As a side effect of this approach we do not have to be concerned about \(\alpha\)-conversion because terms containing free variables in a substitution will never go into the scope of possibly conflicting binders\(^1\).

### 2.1.2 Adding incomplete (open) terms

ALF supports incomplete terms by introducing the so-called placeholders. These correspond to metavariables in other systems. The placeholders are denoted by identifiers starting with '?'\(^2\). So incomplete terms are terms containing placeholders:

\[
\text{Term ::= } \ldots [?]_n
\]

Incomplete types are types containing incomplete terms and incomplete contexts are contexts containing incomplete types.

Each placeholder has an unique expected type (possibly incomplete) and an unique local context (also possibly incomplete). This means that the placeholder

\(^1\)Nevertheless, there are some other problems with \(\alpha\)-conversion, discussed in section 4.4 of [18]

\(^2\)
may be refined by a term whose free variables are from the local context and in this context this term should have a type identical to the expected type of the placeholder.

The typechecking in this system poses specific problems, some of which are not encountered in other systems because of the nature of the underlying type system. One of them is caused by the lack of type annotations of the bound variables in λ-abstraction. Consider for example the term

\[ [x]x \]

It can be inferred using the derivation rules that in the empty context this term is both of type \( \text{Set} \to \text{Set} \) and \( \text{Bool} \to \text{Bool} \) for example. Therefore the following problem

\[ f(?_1, ?_1) : \text{Set} \]

where

\[ f : (\text{Set} \to \text{Set}) \to (\text{Bool} \to \text{Bool}) \to \text{Set} \]

can be solved by instantiating ?₁ with the term \([x]x\). But then we see that the first occurrence of ? has expected type \( \text{Set} \to \text{Set} \) and the second \( \text{Bool} \to \text{Bool} \) which contradicts the condition for uniqueness of the expected type imposed on the placeholders.

Similar problem occurs if we are not careful about the local contexts. Consider for example

\[ g : (A \to A) \to A \to \text{Set} \]

and the problem

\[ g([x]?, ?_2) : \text{Set} \]

in the empty context. Considering the first occurrence of ?₂ which is in context \([x : A]\), we may decide to solve ?₂ by \( x \) obtaining

\[ g([x]x, x) : \text{Set} \]

which is not typable in the empty context, because \( x \) occurs as a free variable.

ALF solves these problems by disallowing the introduction of multiple occurrences of the same placeholder by the user. Possible duplications that may appear in the process of unification will then refer to the "major" occurrence when the placeholder was introduced and must therefore have the same type and same context (or extension of it).

The typechecking of complete terms is done by an algorithm which has two main phases [18]. In the first part a list of equations is being generated and in the second we check if these equations hold by simplifying them. These equations involve two terms of given type and in given context. Magnusson [18] shows that if \( \alpha \) is a proper type in a context \( \Gamma \) then \( \Gamma \vdash e : \alpha \) holds if and only if the generated list of equations \( GE(e, \alpha, \Gamma) \) holds. The second part of the typechecking algorithm takes a list of equations and simplifies them, eliminating the trivial ones. It turns out that the list of equations holds if and only if they
can all be eliminated by this procedure and this gives us a criterion whether or not the judgement we wanted to typecheck holds.

The list of term equations is generated in two steps. First we generate a list of type equations which are then simplified into term equations in the second step. In both cases simplification means application of the appropriate conversion rules of the calculus. Note that this requires the calculus to be normalizing.

We have to point out that as result of the lack of type annotations of the bound variables, terms containing a $\beta$-redex cannot be typechecked as we have to be able to compute the type of the head of an application. This is impossible since the type inference problem in this system is not decidable. Therefore the typechecking algorithm works only for terms which are in normal form. This is not a big theoretical problem, since the calculus is (assumed to be) weakly normalizing, but it may affect the performance of an implementation.

The reason why the typechecking algorithm is split in two phases and the reason to use lists of equations becomes evident when we try to typecheck incomplete terms. In this case the first phase produces a list of declarations and typing constraints which are of the form

\[ ?_n : \alpha_n \Gamma_n \]
\[ \alpha_1 = \alpha_2 \Gamma \]

and the equations may involve placeholders. A placeholder declared in the list can be used in the declaration of other placeholders or in equations which appear after the declaration of the placeholder. This ensures non-circularity in the dependencies between placeholders.

There are no special derivation rules for placeholders. Instead, when first a placeholder is encountered, its context and expected type are computed and a declaration for it is output in the list. If later, due to dependencies, the type of the same placeholder needs to be determined, it can be looked up in the list.

This list constitutes a high-order typed unification problem and as we are in dependently-typed setting there is no algorithm to solve it in general. Nevertheless, we may try to find partial solution. The equations that are left unsolved are constraints on the future instantiations of the placeholders.

The set of placeholder declarations in a list is closely related to the signatures in AllC (see section 2.4) or proof-problems in TypeLab (see section 2.2).

The corresponding correctness property for incomplete terms states that if $\mathcal{C}$ is the list of equations generated for $\Gamma \vdash e : \alpha$ then for any instantiation $\sigma$ of the placeholders of $\mathcal{C}$ by complete terms

\[ \Gamma \sigma \vdash e \sigma : \alpha \sigma \] if and only if $\mathcal{C} \sigma$ holds

This guarantees that the resulting term at the end of the session will be type-correct and we need not typecheck it.

The meta-theoretical properties of the system are not studied in [18] and the results obtained are relative to some meta-theory assumptions such as that constructors are one-to-one, that equal ground terms have the same outermost constructor and that computation on provably equal typable terms terminates and leads to equal normal forms.
2.1.3 The Transitivity Example

Let us consider how the transitivity example can be represented in ALF. Suppose we have defined \( A, R, a \) and \( c \) of their appropriate types and that we have a term \( \text{trans} \) of type \((x,y,z:A) \rightarrow (R \ x \ y) \rightarrow (R \ y \ z) \rightarrow (R \ x \ z)\). Then our goal would be to create a term of type \((R \ a \ c)\):

\[
\text{theorem-aRb}: \quad (R \ a \ c) = \ ?
\]

We specify that we want to use \( \text{trans} \):

\[
\text{theorem-aRb}: \quad (R \ a \ c) = \text{trans} \ a \ ?2 \ c \ ?4 \ ?5
\]

Now we have 3 goals: to find a witness \( ?2 \) of type \( A \) and to prove that \((R \ a \ ?2)\) and \((R \ ?2 \ c)\). At this point we may choose to provide the witness directly, or try to prove the two other goals. This may be a good idea if we cannot directly see a suitable candidate for \( ?2 \). Imagine for example that somehow we manage to prove \((R \ a \ b)\) for some \( b \). Then we can instantiate \( ?2 \) and only the goal \((R \ b \ c)\) will remain.

On the level of the theory of ALF the first state is represented by a list of equations which is roughly:

\[
[?0 : (R \ a \ c) \ \Gamma]
\]

where \( \Gamma \) is a context declaring the appropriate variables. When we refine \( ?0 \) by \( \text{trans} \) and use the witnesses \( a \) and \( c \), we get approximately:

\[
[?2 : A \ \Gamma]
[?4 : (R \ a \ ?2) \ \Gamma]
[?5 : (R \ ?2 \ c) \ \Gamma]
[?0 = \text{trans} \ a \ ?2 \ c \ ?4 \ ?5 \ \Gamma]
\]

At a later stage, if we succeed in proving \( p : (R \ a \ b) \) for some \( b \), we can solve \( ?4 \) by \( p \):

\[
[?2 : A \ \Gamma]
[?5 : (R \ ?2 \ c) \ \Gamma]
[?0 = \text{trans} \ a \ ?2 \ c \ p \ ?5 \ \Gamma]
[(R \ a \ ?2) = (R \ a \ b) \ \Gamma]
\]

which leads to adding the last condition. By simplifying it we can (automatically) infer \( ?2 = b \).

2.1.4 Classification Summary: ALF

<table>
<thead>
<tr>
<th>Properties</th>
<th>MTA</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>MTA</td>
<td>Uses MV</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>?</td>
<td>Explicit Subst.</td>
</tr>
<tr>
<td>Typechecking</td>
<td>decidable</td>
<td>yes</td>
</tr>
<tr>
<td>Type inference</td>
<td>not decidable</td>
<td></td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>MTA</td>
<td>Based on Martin-Löf</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>no</td>
<td>monomorphic TT</td>
</tr>
</tbody>
</table>

(MTA – Meta-Theory Assumption)
2.2 Typelab

The Typelab system is a software tool based on the theory described in [25, 24]. It was designed as a tool aimed at specification and refinement of programs. Specifications are represented in type theory as dependently typed \( \Sigma \)-types and finding programs meeting these specifications is reduced to finding inhabitants of the corresponding \( \Sigma \)-type.

The target system in Typelab is the Extended Calculus of Constructions or ECC [17]. ECC is an extension of the Calculus of Constructions [7] with dependent \( \Sigma \)-types and universes.

This means that Typelab employs a dependently-typed language with metavariables, explicit substitutions and dependent product and \( \Sigma \)-types. It has an infinite hierarchy of type universes \( Prop : Type_0 : Type_1 : \ldots : Type_n : \ldots \).

2.2.1 Description of the Calculus

The terms of the language are given by the following grammar:

\[
\begin{align*}
Term :&= Var[Prop]Type_1 \\
& | \Pi Var : Term.Term | \lambda Var : Term.Term | \\
& | \Sigma Var : Term.Term | (Term \cdot Term) | \\
& | pair_{term}(Term, Term) | \pi_1(Term) | \pi_2(Term) | \\
& | MVar Subst
\end{align*}
\]

\[
Subst : = [[|] [Var := Term]] :: Subst
\]

where \( Var \) is the set of variables, \( MVar \) is the set of metavariables and :: denotes concatenation.

Note that in contrast to other systems based on simpler type theories, here we do not distinguish between terms and types on syntax level.

In Typelab open terms are represented by means of terms which contain metavariables. These metavariables are intended to be placeholders or "holes" for terms that will be constructed at a later stage. Each metavariable has attached to it explicit substitution \( \sigma \). Actually, metavariables without substitutions are not valid terms, but we will identify a metavariable with an empty substitution with the metavariable itself.

We will also assume that we are given a function \( svars : MVar \rightarrow List(Var) \) which assigns a list of variables to each metavariable. The intended meaning is that \( svars(\beta n) \) is the list of variables which may be substituted in \( \beta n \). In other words, this is the list of the variables that may occur free in a term that will instantiate the metavariable. So, if \( svars(\beta n) = \langle x, y \rangle \), then \( \lambda z.x \) is eligible to be considered as a candidate to instantiate \( \beta n \), but \( \lambda z.tz \) is not.

In order to prevent uniqueness problems for the normal forms, we assume a linear order on the variables as given by \( svars \) and require that the variables in an explicit substitution be ordered respectively. For example, if \( svars(\beta n) = \langle x, y \rangle \), then \( \beta n[x := M, y := N] \) is a well-formed term, but \( \beta n[y := N, x := M] \) is not. Note that the order may be different for different metavariables.
As we have already mentioned, Typelab is using dependent types and therefore according to the syntax of the calculus metavariables are allowed to appear in types. But then they may be involved in computations by the conversion rule:

$$\frac{\Gamma \vdash M : A \quad \Gamma \vdash B : \text{Type}}{\Gamma \vdash M : B} \quad \text{(conv)}$$

As we have already noticed, computations with open terms are not so easy:

$$\frac{?n \quad \text{x}}{\text{t}} \rightarrow_{\beta} \frac{\lambda x : ?n \ 	ext{t} \quad ?n \quad ?n \ 	ext{x}}{\text{t} \quad \text{x}}$$

It is clear that the substitution generated by the $\beta$-reduction should not be ignored, since this leads to a term $x$ in which the variable $x$ has escaped its scope and is now free. The solution to this problem is the introduction of explicit substitutions attached to the metavariables: $?n[x := t, y := v]$. This requires additional clauses in the definition of normal (external) substitutions dealing with the case of metavariables.

We still have the usual definition of $\beta$-reduction in terms of (external) substitution denoted by curly brackets ($\sigma = \{ x := t \}$). We need to extend the definition of this metaoperation with the case when it is applied to a metavariable. So $(?n[y_1 := t_1, \ldots, y_k := t_k])\{ x := s \}$ is defined as:

If $x \in \text{svars}(?n)$, $x \not\in \text{\{y_1, \ldots, y_k\}}$ and $x$ is between $y_i$ and $y_{i+1}$ in $\text{svars}(?n)$:

$$?n[y_1 := t_1 \sigma, \ldots, y_i := t_i \sigma, x := s, y_{i+1} := t_{i+1} \sigma, \ldots, y_k := t_k \sigma]$$

If $x \not\in \text{svars}(?n)$:

$$?n[y_1 := t_1 \sigma, \ldots, y_k := t_k \sigma]$$

This means that in effect computations are not completely internalized. The meta-part corresponds to the usual meta-operation of substitution and the explicit substitutions record the metasubstitutions which reach metavariables.

The calculus has the following two base reduction relations:

- $\beta$-reduction:
  $$\lambda x : T. M \rightarrow_{\beta} M\{ x := N \}$$

- $\pi$-reduction:
  $$\pi_i(\text{pair}_T(M_1, M_2)) \rightarrow_{\pi} M_i$$

The compatible, reflexive and transitive closure of these two relations is denoted by $\rightarrow$. This relation is shown to have the Church-Rosser property. It gives rise to a corresponding covertibility = relation, which is further extended to the cumulativity relation $\leq$ which is transitive, symmetric and furthermore:

2The case $x \in \text{\{y_1, \ldots, y_k\}}$ is not possible if we conform with the usual variable conventions.
Prop \leq \text{Type}_0$ and $\text{Type}_j \leq \text{Type}_{j+1}$
If $A_1 = A_2$ and $B_1 \leq B_2$ then $\Pi x : A_1.B_1 \leq \Pi x : A_2.B_2$
If $A_1 \leq A_2$ and $B_1 \leq B_2$ then $\Sigma x : A_1.B_1 \leq \Sigma x : A_2.B_2$

The typing rules of the calculus are an extension of those of ECC and include three new rules for typing expressions with metavariables. When we consider typing of metavariables, we always have in mind some proof problem.

A proof problem $P$ is a triple $(M_P, \text{ctxt}_P, \text{type}_P)$ where $M_P$ is a finite set of metavariables, $\text{ctxt}_P$ is a function which for every $?n$ in $M_P$ produces a context such that $\text{dom}(\text{ctxt}(?n)) = \text{vars}(?n)$. The function $\text{type}_P$ assigns a term to each metavariable in $M_P$. We will drop the index $P$ when this does not lead to confusion.

Put differently, $P = \{\{?, \ldots, ?, k\}, \{\Gamma_1, \ldots, \Gamma_k\}, \{T_1, \ldots, T_k\}\}$ means that we are looking for $k$ terms $?_1, \ldots, ?_k$ and each $?_i$ in context $\Gamma_i$ should have type $T_i$. It is allowed for some $T_i$ or $\Gamma_i$ to contain $?_j$ in case this does not lead to circular dependencies between the metavariables. More precisely, the relation defined by $?_i < ?_j$ if and only if $?_i \in \text{MV}(T_j) \cup (\Gamma_j)$ should be an irreflexive partial order. For a fixed proof problem $P$ this is a finite relation and consequently this question is decidable.

The typing rules involving metavariables in Typelab are:

\[
\frac{\text{ctxt}(?n) \vdash \text{type}(?n) : \text{Type}_j}{\text{ctxt}(?n) \vdash ?n \vdash \text{type}(?n)} \quad \text{(MV-base)}
\]

If $z \notin \text{FV}(\Gamma) \cup \text{FV}(\Delta) \cup \text{dom}(\sigma)$:

\[
\frac{\Gamma \vdash T : \text{Type}_j \quad \Gamma, z : T, \Delta \vdash ?n \sigma : N}{\Gamma, z : T, \Delta \vdash ?n \sigma : N} \quad \text{(MV-weak)}
\]

\[
\frac{\Gamma \vdash t : T \quad \Gamma, x : T, \Delta \vdash ?n \sigma : N}{\Gamma, \Delta\{x := t\} \vdash \{?n\sigma\}\{x := t\} : N\{x := t\}} \quad \text{(MV-\beta-Red)}
\]

For this typing system one can show that the following properties hold:

**Principal Type** If $M$ is well-typed in a context $\Gamma$ then there exists a type $A$ (principal type) such that $\Gamma \vdash M : A$ and for every $A'$

$$\Gamma \vdash M : A' \text{ if and only if } A \leq A' \text{ and } A' \text{ is type in } \Gamma$$

**Subject Reduction** If $\Gamma \vdash M : A$ and $M \rightarrow N$ then $\Gamma \vdash N : A$

**Decidability** Type inference and hence type checking (using the Principal Type property) are decidable.

**Strong normalization** The calculus is strongly normalizing.

The proof construction proceeds by creating metavariables for unknown terms and solving them by means of instantiations. Instantiations are defined in such way that guarantees the well-typedness of the resulting proof problem. This on its behalf is the basis for the soundness of the proof method.
2.2.2 The Transitivity Example

Our transitivity example can be expressed also in TypeLab. Suppose $\Gamma$ is a context containing the declarations for a set $A$, a binary relation $R$, two elements $a$ and $c$ of $A$ and the assumption $\text{trans}$ that $R$ is transitive:

$$\Gamma = (A : \text{Type}, R : A \rightarrow A \rightarrow \text{Prop}, a : A, c : A, \text{trans} : \forall x, y, z : A \; xRy \rightarrow yRz \rightarrow xRz)$$

In this context we want to prove $aRc$:

$$\Gamma \vdash \gamma : xRy$$

Using $\text{trans}$ we obtain three goals:

$$\Gamma \vdash \gamma_2 : A$$
$$\Gamma \vdash \gamma_3 : aR\gamma_2$$
$$\Gamma \vdash \gamma_4 : \gamma_2 R\gamma$$

which we are free to solve in any order we want.

2.2.3 Classification Summary: Typelab

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>Uses MV yes</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>Explicit Subst. yes</td>
</tr>
<tr>
<td>Typechecking</td>
<td>MV only</td>
</tr>
<tr>
<td>Type inference</td>
<td>Based on Expanded Calculus</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>de Bruijn Ind. no</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td></td>
</tr>
</tbody>
</table>

2.3 OLEG

OLEG [19] is a calculus of open terms for the ECC with local definitions but without $\Sigma$-types. It consists of two layers - OLEG core and OLEG development calculus. The core is essentially ECC with definitions and the development calculus is wrapped around it adding metavariable support.

OLEG was introduced as a tool to study dependently typed functional programs and is partly implemented as an extension of the proof-assistant LEGO [23].

The OLEG calculus has a couple of features which distinguish it from the rest of the calculi that we consider here. The first one is the introduction of explicit binders for holes which allow a metavariable to be treated as a normal one in many respects. The second is the way OLEG handles computations (reductions) involving metavariables. Instead of using functional coding or explicit substitution, OLEG forbids occurrences of metavariable declarations at trouble spots.

Let us look more closely at the calculus:

$^3$To be precise these three goals form the following proof problem: $P = \langle (\gamma_2, \gamma_3, \gamma_4), (\Gamma, \Gamma, \Gamma), (A, aR\gamma_2, \gamma_2 R\gamma) \rangle$. 

15
2.3.1 The syntax

The syntax and typing of OLEG are given by a number of rules. We will not present all of them here, they can be found in [19]. A core OLEG term can be

- Variable $x$
- Application $MN$
- Abstraction $\lambda x : T.M$
- Dependent Type $\forall x : T.M$
- Definition $!x = M : T.N$

There are three contraction schemes:

(13) $\Gamma \vdash (\lambda x : T.N)M \rightarrow_\beta !x = M : T.N$

(15) $\Gamma, x = M : T, \Delta \vdash x \rightarrow_\delta M$

(17) $\Gamma \vdash !x = M : T.N \rightarrow_\iota N$ if $x \notin FV(N)$

Note that (13) does not involve externally defined substitution, it just converts a $\beta$-redex into a $\delta$-redex:

$$(\lambda x : T.N)M \rightarrow_\beta !x = M : T.N \rightarrow_\delta \ldots \rightarrow_\delta !x = M.N[x/M] \rightarrow_\iota N[x/M]$$

The core calculus enjoys the usual properties of ECC as Church-Rosser, subject reduction, strong normalization and cut (if $\Gamma, !x = s : S, \Delta \vdash M : A$ then $\Gamma, \Delta\{x := s\} \vdash M\{x := s\} : A\{x := s\}$). The typing rules for the core calculus mimic those for ECC, with substitution replaced by appropriate definitions, for example:

$$(\text{app}) \quad \begin{align*}
\Gamma \vdash F : \forall x : S.T & \quad \Gamma \vdash t : S \\
\Gamma \vdash Ft : !x = t : S.T
\end{align*}$$

The development calculus is an extension of the core calculus with open terms. The term language is extended with a new binder for metavariables “?”. So, if $M$ is a term, then $?x : T.M$ is also a term. The binder declares a metavariable and this metavariable can be used as a normal variable in its scope: $\lambda f : A.?x : B.fx$. Another extension of the term language is the introduction of “guesses” which represent a term that may possibly be a solution of a metavariable: $?x \approx P : S.M$. The $\approx$ sign is used to emphasize that this expression cannot be treated as a definition and $x$ cannot be replaced by $P$ in $M$.

The term construction process goes like this: first we introduce a metavariable of the appropriate type $?x : A.M$ then we try to solve $x$ by $P_1$ resulting in $?x \approx P_1.M$. However, this may be a wrong guess, so we can change it to $P_2$ and so on, until we have found the right one $?x \approx P.M$. Once this has happened, we can turn the guess into local definition $!x = P.M$. All these operations can be done by means of appropriate admissible rules.
For every metavariable we have to be able to uniquely identify its context and type. The type is explicitly given in its declaration and the context can be determined by the position of its declaration (i.e. the context is formed by all variables in whose scope is the declaration of the metavariable).

The treatment of metavariables in such a way is not enough to avoid problems. For example it is quite possible to obtain terms like

\[(?x : A.x).f\]

whose typability is problematic, especially if we want to be able to solve metavariables in arbitrary order and to have commuting hole filling and computation.

McBride in [19] relates those problems to metavariable declarations leaking into types and imposes restrictions on the system which guarantee that declarations of metavariables will not occur in the types involved in a derivable judgement.

These restrictions are strong enough to ensure that the resulting development calculus has the replacement property which states that a guess of given type can safely be replaced by any other of the same type; a property that is very useful in applications (for precise formulation and proof, cf. [19]).

The rules involving metavariable declarations are given below:

\[
\begin{align*}
(\text{declare}) & \quad \frac{\Delta \vdash \text{valid}}{\Delta, ?x : T \vdash \text{valid}} \\
(\text{construct}) & \quad \frac{\Delta \vdash \text{p : T}}{\Delta, ?x \equiv \text{p : T} \vdash \text{valid}} \\
(\text{hole}) & \quad \frac{\Delta, ?x : S \vdash \text{p : T}}{\Delta \vdash ?x : S, \text{p : T} \quad \text{if} \ x \not\in T} \\
(\text{guess}) & \quad \frac{\Delta, ?x \equiv q : S \vdash \text{p : T}}{\Delta \vdash ?x \equiv q : S, \text{p : T} \quad \text{if} \ x \not\in T}
\end{align*}
\]

Note the side conditions in (hole) and (guess). They disallow metavariable bounded in a term to occur in its type. If this condition were not present, we would have to accept metavariable declarations in types.

The side conditions in these two typing rules are not enough to guarantee that \(\gamma\)-binders do not appear in types. This is achieved by disabling applications and \(\lambda\)-bounded values containing \(\gamma\)-binders (i.e. if \(\lambda \gamma \text{x = s : S.M} \) and \(\text{f s} \) are terms, then \(\text{s} \) is not allowed to contain \(\gamma\)-binders) and removing \(\beta\) and \(\delta\) reduction from the contraction schemes for partial constructions. Using the embedding of core terms into the development calculus the system is still allowed to do computations on expressions with metavariables (declared in the context), but computations on terms with \(\gamma\)-binders are disallowed.

The lack of \(\delta\)-reduction on terms with \(\gamma\) binders prevents us from getting inconsistencies like:

For every metavariable we have to be able to uniquely identify its context and type. The type is explicitly given in its declaration and the context can be determined by the position of its declaration [i.e. the context is formed by all variables in whose scope is the declaration of the metavariable].

The treatment of metavariables in such a way is not enough to avoid problems. For example it is quite possible to obtain terms like \(\lambda f : (\gamma x : A.x).f\) whose typability is problematic, especially if we want to be able to solve metavariables in arbitrary order and to have commuting hole filling and computation.

McBride in [19] relates those problems to metavariable declarations leaking into types and imposes restrictions on the system which guarantee that declarations of metavariables will not occur in the types involved in a derivable judgement.

These restrictions are strong enough to ensure that the resulting development calculus has the replacement property which states that a guess of given type can safely be replaced by any other of the same type; a property that is very useful in applications (for precise formulation and proof, cf. [19]).

The rules involving metavariable declarations are given below:

\[
\begin{align*}
(\text{declare}) & \quad \frac{\Delta \vdash \text{valid}}{\Delta, ?x : T \vdash \text{valid}} \\
(\text{construct}) & \quad \frac{\Delta \vdash \text{p : T}}{\Delta, ?x \equiv \text{p : T} \vdash \text{valid}} \\
(\text{hole}) & \quad \frac{\Delta, ?x : S \vdash \text{p : T}}{\Delta \vdash ?x : S, \text{p : T} \quad \text{if} \ x \not\in T} \\
(\text{guess}) & \quad \frac{\Delta, ?x \equiv q : S \vdash \text{p : T}}{\Delta \vdash ?x \equiv q : S, \text{p : T} \quad \text{if} \ x \not\in T}
\end{align*}
\]

Note the side conditions in (hole) and (guess). They disallow metavariable bounded in a term to occur in its type. If this condition were not present, we would have to accept metavariable declarations in types.

The side conditions in these two typing rules are not enough to guarantee that \(\gamma\)-binders do not appear in types. This is achieved by disabling applications and \(\lambda\)-bounded values containing \(\gamma\)-binders (i.e. if \(\lambda \gamma \text{x = s : S.M} \) and \(\text{f s} \) are terms, then \(\text{s} \) is not allowed to contain \(\gamma\)-binders) and removing \(\beta\) and \(\delta\) reduction from the contraction schemes for partial constructions. Using the embedding of core terms into the development calculus the system is still allowed to do computations on expressions with metavariables (declared in the context), but computations on terms with \(\gamma\)-binders are disallowed.

The lack of \(\delta\)-reduction on terms with \(\gamma\) binders prevents us from getting inconsistencies like:

\[
\begin{align*}
\text{\(\lambda x : T. ?n : A. n \quad ?x : x\)} & \quad \text{\(\lambda x : A. ?n : x\)} \\
\text{\(\lambda x : A. ?n : x\)} & \quad \text{\(\lambda x : A. ?n : x, n\)}
\end{align*}
\]
Of course, a term may exploit metavariables in the context, but then they are outside the scope of the \(!\)-binder and cause no trouble:

\[
\text{x is not in the scope of n}
\]

Although the restrictions imposed on the use of metavariables in OLEG are very strict, the calculus is still expressive enough to capture all the cases in the intended use of metavariables in the context of [19]. The metatheoretic properties of the core are also present in the development calculus: confluence, Subject Reduction, Cut, Strong Normalization, etc.

### 2.3.2 The Transitivity Example

The declarations we need for introducing the problem, using the notation for states introduced in section 2.4 of [19] are:

\[
\begin{align*}
\lambda \ A & : Type \\
\lambda \ R & : A \rightarrow A \rightarrow Prop \\
\lambda \ a, c & : A \\
\lambda \ trans & : \forall x, y, z : A. (Rxy) \rightarrow (Ryz) \rightarrow (Rxz) \\
?x_1 & : (Rac)
\end{align*}
\]

Then refining \(x_1\) by \(\text{trans}\) and filling in \(a\) and \(c\) for \(x\) and \(z\) we get

\[
\begin{align*}
?x_1 & \Rightarrow ?y : A \\
?r_1 : (R a y) \\
?r_2 : (R y c) \\
\text{trans a y c r_1 r_2} \\
:(R a c)
\end{align*}
\]

and we are free to solve \(r_1\) or \(r_2\) first.

### 2.3.3 Classification Summary: OLEG

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>yes</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>yes</td>
</tr>
<tr>
<td>Typechecking</td>
<td>decidable</td>
</tr>
<tr>
<td>Type inference</td>
<td>decidable</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>yes</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>yes</td>
</tr>
<tr>
<td>Uses MV</td>
<td>yes</td>
</tr>
<tr>
<td>Explicit Subst.</td>
<td>no</td>
</tr>
<tr>
<td>local definitions</td>
<td></td>
</tr>
<tr>
<td>Based on</td>
<td>Extended Calculus</td>
</tr>
<tr>
<td>de Bruijn Ind.</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>Based on Constructions</td>
</tr>
</tbody>
</table>
2.4 The $\lambda\Pi_L$-calculus

The $\lambda\Pi_L$-calculus is a calculus of explicit substitutions using de Bruijn indices. It has dependent types and the metavariables and the substitutions are first-class objects in the calculus. Its origins can be traced back to $\lambda\sigma$, but it essentially extends it, since adding support for dependent types is not straightforward. $\lambda\Pi_L$ is based on dependently typed lambda calculus $\lambda\Pi$. It has a version for the stronger Calculus of Constructions called $CC_L$, but the solutions to the main problems can be found in $\lambda\Pi_L$.

2.4.1 Description of $\lambda\Pi_L$

The well-formed expressions in $\lambda\Pi_L$ are given by the following grammar:

- natural numbers $n$ :: $\emptyset|n+1$
- metavariables $\mathcal{M}$ ::= $X[Y]...$
- terms $A, B, M, N$ ::= $\text{Kind}[\text{Type}]\uparrow\lambda A.M[\Pi A.M](MN)[M|S]\mathcal{M}$
- substitutions $S, T$ ::= $\uparrow^n[M \cdot \mathcal{A}]S \circ T$

It has the following reduction rules:

- **(Beta)** $(\lambda A.M)N \rightarrow M[N'A]$  
- **(Lambda)** $(\lambda A.M)[S] \rightarrow \lambda A[S].(M[\langle 1 \cdot A (S \circ \uparrow^1) \rangle])$
- **(Pi)** $(\Pi A.B)[S] \rightarrow \Pi A[S].(B[\langle 1 \cdot A (S \circ \uparrow^1) \rangle])$
- **(App)** $(MN)[S] \rightarrow (M[S]N[S])$
- **(Clos)** $(M[S])[T] \rightarrow M[S \circ T]$
- **(VarCons)** $1[M \cdot \mathcal{A}]S \rightarrow M$
- **(Id)** $M[\uparrow^0] \rightarrow M$
- **(Map)** $(M \cdot \mathcal{A}) \circ T \rightarrow M[T] \cdot (S \circ T)$
- **(IdS)** $\uparrow^0 \circ S \rightarrow S$
- **(ShiftCons)** $\uparrow^{n+1} o (M \cdot \mathcal{A}) S \rightarrow \uparrow^n oS$
- **(ShiftShift)** $\uparrow^{n+1} o \uparrow^m \rightarrow \uparrow^n o \uparrow^{m+1}$
- **(Shift0)** $1 \uparrow^1 \rightarrow \uparrow^0$
- **(ShiftS)** $[\uparrow^n \cdot \mathcal{A}] \uparrow^{n+1} \rightarrow \uparrow^n$
- **(Type)** $\text{Type}[S] \rightarrow \text{Type}$

In order to be able to type metavariables, judgements in the system are extended to contain signatures (denoted by $\Sigma$) which are lists of metavariable declarations: $\Sigma;\Gamma \vdash M : A$. The judgements expressing the validity of a signature have the form $\vdash \Sigma$, the judgement expressing that a context $\Gamma$ is valid in a signature $\Sigma$ is $\vdash \Sigma;\Gamma$. The type $A_i$ and the context $\Gamma_i$ of a metavariable $X_i$ in a signature $\Sigma = (X_1 : \Gamma_1, A_1, \ldots, X_n : \Gamma_n, A_n)$ may depend only on metavariables with index $j$ where $i < j \leq n$. The rules for the metavariables are the following:

- **(MV-decl)** $\vdash \Sigma;\Gamma \vdash A : s. \Sigma$-fresh

$$\Sigma;\Gamma \vdash \Delta = \lambda\Pi_L \vdash X : A \vdash \Sigma;\Gamma \vdash X : A \in \Sigma$$
The full set of rules can be found in [21]. Since substitutions by themselves are objects in the calculus, they can be typed and the type of a substitution is a context.

One specific feature of $\lambda\Pi_{\mathcal{C}}$ is that the substitutions are annotated with types. These annotations are created by the rules (Beta), (Lambda) and (Pi) and are preserved during the life-cycle of the substitution.

There are however some subtle details that need attention. Blou [5] presents an example by Geuvers for a term (namely $(\lambda x : N.\lambda f : (Pz) \rightarrow N.\lambda y : (Pz).fy)z$ in the context $\Gamma_N = N : Type, P : N \rightarrow Type$ and $z : N$) which is typable, but reduces to a term which is not. This is due to the fact that there might be a difference between the type in the declaration of a variable after the substitution and the type of the variable at the place where it is used, but the substitution is still not executed. This happens if we distribute the substitution over the binder for $f$:

$$\lambda f : (Pz) \rightarrow N.\lambda y : (Pz).fy\{x := z\}$$

$f$ expects an argument of type $(Pz)$ but since the substitution is not yet propagated over the binder for $y$, the variable $y$ has type $(Pz)$ which is not convertible to $(pz)$. This problem is avoided by introducing the type annotations in substitutions. Take for example the (Lambda) rule. Written using names for the variables it reads

$$(\lambda x : A.M)[S] \rightarrow (\lambda x : A[S].M[x := x : A, S])$$

In this way passing through abstraction the substitution is extended by a "dummy" substitution which records the original type of the variable. In this way using the (Clos) derivation rule

$$(\text{Clos}) \quad \frac{\Sigma; \Gamma \vdash S : \Delta \quad \Sigma; \Delta \vdash M : A \quad \Sigma; \Delta \vdash A : s}{\Sigma; \Gamma \vdash M[S] : A[S]}$$

the correct types can be computed.

Note that this problem is avoided in Typelab which also has explicit substitutions by allowing them only at metavariables and using the (meta)operation substitution.

The metatheoretic properties of $\lambda\Pi_{\mathcal{C}}$ have been studied and proofs can be found for example in [21]. The calculus is confluent on well-typed terms and has the properties of subject reduction, type uniqueness, weak normalization, etc.

2.4.2 The Transitivity Example

$\lambda\Pi_{\mathcal{C}}$ is strong enough to provide the functionality needed to facilitate the transitivity example. Let$^4$ $\Gamma$ be the context $\text{trans} : \Pi x, y, z : A.(Rx\neg y) \rightarrow$

$^4$Using names for variables
(Ryz) \rightarrow (Rxz).a : A.c : A.R : A \rightarrow A \rightarrow Type.A : Type. Then the goal of proving (Rac) can be expressed by the following judgement:

\[(X : \Gamma (Rac)); \Gamma \vdash X : (Rac)\]

refining by \textit{trans} results in

\[(P_2 : \Gamma (RYc).P_1 : \Gamma (RaY).Y : \Gamma A); \Gamma \vdash trans a Y c P_1 P_2 : (Rac)\]

### 2.4.3 Classification Summary: \(\lambda\Pi_C\)

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>yes</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>yes</td>
</tr>
<tr>
<td>Typechecking</td>
<td>decidable</td>
</tr>
<tr>
<td>Type inference</td>
<td>decidable</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>yes</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>no</td>
</tr>
<tr>
<td>Uses MV</td>
<td>yes</td>
</tr>
<tr>
<td>Explicit Subst.</td>
<td>yes</td>
</tr>
<tr>
<td>Based on (\lambda\Pi)</td>
<td></td>
</tr>
<tr>
<td>de Bruijn Ind.</td>
<td>yes</td>
</tr>
</tbody>
</table>
Chapter 3

Other Calculi of Open Terms

In this chapter we consider several other systems of open terms which are not explicitly developed or suited for theorem proving, but still deserve attention.

3.1 The Context Calculus $\lambda c$

3.1.1 Contexts in untyped $\lambda$-calculus as open terms

In [6] a calculus of (untyped) $\lambda$-calculus contexts is presented. In this section we will use the term context in the sense of this paper. Note that it denotes a different notion, related more to open terms than to contexts as lists of variable declarations.

So, the central notion is context, "an expression with special places called holes, where other expressions can be placed". We will denote the holes by the symbol $\Box$. Examples for contexts are

$$(\lambda x.\Box)z \quad y\Box$$

The main motivation for introducing this calculus is to formalize the notion of context used in many formal discussions just as a notational convention. This formalization has to take into consideration the effect of variable capturing which occurs when the term filling a hole has a free variable which becomes bound by a binder in whose scope the hole is situated. Suppose in the context $(\lambda x.\Box)z$ we want to put the term $xz$. Then the result is

$$(\lambda x.xz)z$$

and the variable $z$ becomes bound.

Another potential difficulty of the communication between a context and the expressions being put in its holes is again the one described in section 1.4.
This means that the informally defined \(\lambda\)-contexts are not considered modulo \(\alpha\)-conversion and cannot be subject to \(\beta\)-reduction which is limiting their use. The \(\lambda\)-calculus described below formalizes the notion of context and solves these problems.

### 3.1.2 Description of the Calculus

In \(\lambda\) we have new constructors: \(\Lambda\) for multiple abstraction, \(\otimes\) for multiple application, \(\circ\) for composition of contexts and also the corresponding multiple (simultaneous) substitution:

\[
T ::= V \\
(\lambda V.T) \\
(TT) \\
(\Lambda n V_1, \ldots, V_n.T) \\
(\otimes T, T_1, \ldots, T_n) \\
(\circ(T, T_1, \ldots, T_n))
\]

There are two main operations defined on contexts: hole filling and composition. In the first case we are given a context and we put a term in a hole of the context. This operation is represented by multiple application. In the case of composition we fill the hole with another context.

The system has the following reduction rules:

\[
(\lambda x.U)V \rightarrow_\beta U\{x := V\} \\
\otimes(\Lambda \bar{x}.U)V \rightarrow_\beta U\{\bar{x} := V\} \\
\circ(\bar{B}g.U, (\Lambda \bar{h}^1 \bar{B} \bar{z}^1.V_1), \ldots, (\Lambda \bar{h}^n \bar{B} \bar{z}^n.V_n)) \rightarrow_\circ \\
\rightarrow_\circ \Lambda \bar{h}^1 \bar{B} \bar{z}^1, \ldots, \bar{z}^n.U\{g_1 := \Lambda \bar{h}^1.V_1, \ldots, g_n := \Lambda \bar{h}^n.V_n\}
\]

where \(\bar{B}\)s are either all sequences of \(\lambda\) or all sequences of \(\Lambda\) (not necessarily of the same length). Note also the movement of the binders for \(\bar{x}\) outside the scope of the binders for \(\bar{h}\).

Here is an example illustrating the use of the composition rule:

\[
\circ((Ag, g'.U), (A\bar{z}Ah.V), (A\bar{y}Ak, k'.W)) \rightarrow \Lambda h, k, k'.U\{g := \Lambda \bar{z}V, g' := \Lambda \bar{y}W\}
\]

The main result in [6] states that this calculus is confluent.

It is a framework for representing different kinds of contexts. The exact structure of the contexts can be fine-tuned using pretyping. Here we will discuss the example in [6] where pretyping yielding untyped \(\lambda\)-contexts is presented.

In the untyped \(\lambda\)-calculus the terms and contexts are given by the following rules:

\[
M ::= x | \lambda x.M | MM \\
C ::= x | \Box | \lambda x.C | CC
\]
with the assumption that each occurrence of $\square$ is different and the holes are ordered from left to right.

The translation of these terms into $\Lambda c$ is just the identity and the translation for contexts is defined as follows:

$$|C| = A\tilde{h}.C[(\overline{\otimes h_1}, \tilde{x^1}), \ldots, (\overline{\otimes h_n}, \tilde{x^n})]$$

where $C[M_1, \ldots, M_n]$ means the context $C$ with its first hole filled by $M_1$, the second by $M_2$, etc. The vector $\tilde{x^i}$ contains all variables in whose scope is the $i$th hole. This means that a hole $\square$ in a $\lambda$-context is represented by a hole variable $h$ applied to all the variables that may occur free in the hole $\overline{\otimes h x_1, \ldots, x_n}$. In this way, any subsequent substitution passing through the hole will be "remembered" because the appropriate variable would be substituted in the argument list.

We can extend this translation to the hole-filling operation $hf(C, \overline{\otimes M})$ which fills the holes of $C$ by the terms $\overline{\otimes M}$ and the composition operation $comp(C, \overline{\otimes D})$ which fills the holes of $C$ with other contexts from $\overline{\otimes D}$. One has to distinguish between $hf(C, M)$ and $C[M]$ as the first denotes the operation and the second denotes the resulting term. The same holds for $comp(C, D)$ and $C[D]$.

$$|hf(C, \overline{\otimes M})| = \overline{\otimes M} |C|, A\tilde{x^1}.|D_1|, \ldots, A\tilde{x^n}.|D_n|$$

where $\tilde{x^i}$ is the list of bound variables in whose scope is the $i$th hole of $C$. The translation of the composition is not so straightforward, because the holes in $\overline{\otimes D}$ have to be "lifted":

$$|comp(C, \overline{\otimes D})| = c(|C|, (A\tilde{x^1}.LIFT(|D_1|, \tilde{x^1})), \ldots, (A\tilde{x^n}.LIFT(|D_n|, \tilde{x^n})))$$

$$LIFT(D, \tilde{x}) = A\tilde{y}.(\overline{\otimes D}(\overline{\otimes y^\cdots, \overline{\otimes y^\cdots}})), \ldots, (A\tilde{y^n}.(\overline{\otimes y^n, \overline{\otimes y^n}}))$$

where $\tilde{y^i}$ is the list of the bound variables in whose scope is the $i$th hole of $D$.

The pre-typing rules given in [6] guarantee the correctness of this translation of untyped $\lambda$-contexts into $\Lambda c$. The derivation system has only one base type $t$ for terms, hole variables have type $[\overline{\otimes t}]$, contexts have types $[\overline{\otimes t}] \times \ldots \times [\overline{\otimes t}] \Rightarrow t$ and contexts, which are to be put in the holes of another context have type $[\overline{\otimes t}]([\overline{\otimes t}] \times \ldots \times [\overline{\otimes t}] \Rightarrow t)$.

The derivation rules impose discipline on the term formations so that the above translation can be implemented. The resulting calculus $\Lambda c$ has the reduction rules of $\Lambda c$, but the set of terms is limited to the well-pretyped terms of $\Lambda c$.

Since $\Lambda c$ was not aimed at theorem proving and the pre-typing is only used to separate the well-formed terms and does not support dependent types, we have no way to express the problem in the transitivity example in $\Lambda c$. 

24
3.1.3 Classification Summary: \( \lambda c \)

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>yes</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>not applicable</td>
</tr>
<tr>
<td>Type checking</td>
<td>not applicable</td>
</tr>
<tr>
<td>Type inference</td>
<td>not applicable</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>?</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>?</td>
</tr>
<tr>
<td>Uses MV</td>
<td>yes</td>
</tr>
<tr>
<td>Explicit Subst.</td>
<td>Functional encoding</td>
</tr>
<tr>
<td>Based on</td>
<td>Untyped ( \lambda )-calculus</td>
</tr>
<tr>
<td>de Bruijn Ind.</td>
<td>no</td>
</tr>
</tbody>
</table>

3.2 Automath

3.2.1 The calculus of segments

The calculus was introduced by de Bruijn [9] and later further studied in [2]. It provides facilities to introduce abbreviations for parts of terms containing a hole which can be filled with another term. Let us consider a small example:\(^1\):

\[
\lambda x \cdot \lambda y \cdot \lambda z \cdot \lambda w
\]

This diagram is a tree representation of the term \( \lambda x \lambda y \lambda z \cdot \lambda w \cdot x \cdot y \cdot z \).

There are several things that at first look strange and we need to explain them: First, the expression \( \lambda z \cdot \lambda w \cdot x \cdot y \cdot z \) represents an abbreviation and the symbol \( w \) denotes the place where we need to plug in a term in order to construct a valid instantiation of the segment. One segment may only have one open end denoted by \( w \). It can only be at the far right end of the tree representation of the segment also known as the "heart" position [10].

An abbreviated expression may be used by means of a special variable called segment variable (in this case \( \sigma \)). In contrast to normal variables, segment variables do not appear on the leaves, but inside the tree. Segment variables stand for a segment. The binding between the segment and the segment variable is accomplished by an artificially created redex \((\lambda \sigma \cdot \ldots ) (\lambda z \cdot \lambda y \cdot w)\).

Second, there is a problem with the bound variables: in the original expression the variables \( x \) and \( y \) are bound and in the abbreviated expression they

\(^1\) In [9] and [2] the system is presented as a namefree calculus, but here we will illustrate it using named variables.
seem to be outside the scope of the binders. This situation is resolved by adding annotations to the segment variables:

\[
\begin{align*}
\lambda_x & \quad \lambda_y \quad \omega \\
\sigma(x_1, y_1) & \quad \sigma(x_2, y_2) \\
\lambda_y & \quad x_2
\end{align*}
\]

Things now seem to be looking fine, but there is a third problem, hidden in the \(\beta\)-reduction. Of course we want our abbreviations to really behave like ones and not affect \(\beta\)-reduction which means that the unfolding of abbreviations (which is in fact done by a sort of \(\beta\)-reduction) must commute with the execution of substitutions.

The problem lies in the fact that as a result of a \(\beta\)-reduction we may need to apply a substitution to a segment. Then we have to define how the substitution works on the \(\omega\)-terminator at the end of a segment. We adapt an example from [2]:

\[
\begin{align*}
\lambda_x & \quad \delta \quad \lambda_y \quad \lambda_z \quad \omega \\
\sigma(x, y, z) & \quad y
\end{align*}
\]

We see that there is a redex \((\lambda y \ldots)u\). If we decide to reduce it, we need to propagate the substitution of \(u\) for \(y\) in \(\lambda z.\omega\). This confronts us with two problems. First, how does the substitution work on \(\omega\) and second, what does the variable \(y\) in the main branch of the tree refer to after the reduction. The naive definition of substitution which does nothing on \(\omega\) leads to:

\[
\begin{align*}
\lambda_x & \quad \lambda_z \quad \omega \\
\sigma(x, y, z) & \quad y
\end{align*}
\]

And this is of course unacceptable as we can see after unfolding the abbreviation. Instead, when we reach \(\omega\), we create a redex which "remembers" the substitution, effectively postponing it until the abbreviation is unfolded. This however leads to reordering of the binders in the segment. Consequently we need to reorder the variable references in the segment variable\(^2\):

\[
\begin{align*}
\lambda_x & \quad \lambda_z \quad \delta \quad \lambda_y \quad \omega \\
\sigma(x, z, y) & \quad y
\end{align*}
\]

\(^2\)At first sight this means that substitution is not a local operation anymore because we need to update the segment variables. This is however avoided by explicitly recording a permutation of the variables in the \(\omega\) instead of the segment variable occurrence and applying this permutation to the variable list later at the moment of unfolding.
3.2.2 Properties and Extensions of the System

In [2] several properties of the segment calculus are proved. The most important one is that the calculus is Church-Rosser.

A typed version of the calculus also exists. It is an extension of the simply typed lambda calculus in the Church style. A special sort of types, called $\rho$-types are introduced in order to type the segments. For the typed version of the calculus it is proved that under some assumptions, the closure property holds. It states that the type of a term is preserved under $\beta$-reduction.

3.2.3 Classification Summary: Automath

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>Uses MV</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>Explicit Subst.</td>
</tr>
<tr>
<td>Typechecking</td>
<td>Create</td>
</tr>
<tr>
<td>Type inference</td>
<td>Based on</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>Untyped</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>de Bruijn Ind.</td>
</tr>
</tbody>
</table>

3.3 The $\lambda_{se}$-calculus

The $\lambda_{se}$-calculus [14] was introduced as extension of the $\lambda_s$-calculus [13, 22] with open terms. This is a calculus of open terms with metavariables and explicit substitutions. It employs nameless variables in the style of de Bruijn. Unfortunately no dependently-typed version of this calculus has been developed so far, which prevents its wider use in theorem proving.

The first thing that separates $\lambda_s$ and $\lambda_{se}$ from the other calculi of explicit substitutions in de Bruijn notation is the way the explicit substitutions are introduced. Instead of creating a new syntactic category of substitutions (compare with section 2.4), this calculus internalizes them by means of two families of operators:

- Explicit substitution operators $\sigma^j$, $1 \leq j$. The expression $a \sigma^j b$ means the term $a$ where the free occurrences of the variable with index $j$ are to be replaced by $b$.

- Updating functions $\phi^i_k$, $0 \leq k$, $1 \leq i$. These functions are used to update the indices of a term subject to substitution. The intuitive meaning is that $\phi^i_k t$ is the term $t$ in which all indices greater than $k$ are incremented by $i - 1$.

The terms of the system are given by the following grammar:

$$M, N ::= V | N | (M N) | (\lambda M) | (M \sigma^j N) | \phi^i_k M$$

where $V$ is the set of metavariables.
\( \lambda s_e \) is a rewrite system with the following reduction rules:

\[
\begin{align*}
\sigma\text{-generation} & \quad (\lambda a)b \rightarrow a\sigma^1 b \\
\sigma\text{-\lambda-transition} & \quad (\lambda a)\sigma^i b \rightarrow \lambda(\sigma^{i+1} b) \\
\sigma\text{-app-transition} & \quad (a_1 a_2)\sigma^i b \rightarrow (a_1 \sigma^i b)(a_2 \sigma^i b) \\
\sigma\text{-destruction} & \quad n\sigma^i b \rightarrow \begin{cases} 
\varphi^i_b & \text{if } n = i \\
\varphi^i_n & \text{if } n < i \\
n - 1 & \text{if } n > i 
\end{cases} \\
\varphi\text{-\lambda-transition} & \quad \varphi^i_k(\lambda a) \rightarrow \lambda(\varphi^i_{k+1} a) \\
\varphi\text{-app-transition} & \quad \varphi^i_k(a_1 a_2) \rightarrow \varphi^i_k(a_1)\varphi^i_k(a_2) \\
\varphi\text{-destruction} & \quad \varphi^i_k(n) \rightarrow \begin{cases} 
n + i - 1 & \text{if } n > k \\
n & \text{if } n \leq k 
\end{cases} \\
\sigma\text{-\sigma-transition} & \quad (a\sigma^i b)\sigma^j c \rightarrow (a\sigma^{i+j+1} c)\sigma^i (b\sigma^j c) \\
\sigma\text{-\varphi-transition} & \quad (\varphi^i_k a)\sigma^i b \rightarrow \varphi^i_{k+1} a \\
\sigma\text{-\varphi-transition} & \quad (\varphi^i_k a)\sigma^i b \rightarrow \varphi^i_k (a\sigma^{i-j} b) \\
\varphi\text{-\sigma-transition} & \quad \varphi^i_k (a\sigma^j b) \rightarrow \varphi^i_{k+1} (a)\varphi^j_{k-j+1} b \\
\varphi\text{-\varphi-transition} & \quad \varphi^i_k (\varphi^j_l a) \rightarrow \varphi^i_{k+1-j} a \\
\varphi\text{-\varphi-transition} & \quad \varphi^i_k (\varphi^j_l a) \rightarrow \varphi^i_{k+l+1-j} a
\end{align*}
\]

Considered as a rewrite system, \( \lambda s_e \) has the following properties:

**\( \beta \)-reduction** It simulates \( \beta \)-reduction, which means that if \( M \) \( \beta \)-reduces in one step to \( N \) then \( M \) reduces to \( N \) using the rules of \( \lambda s_e \).

**Confluence** The calculus is confluent on open terms.

The calculus does not preserve strong normalization In [11] Guillaume gives an example for a strongly normalizable \( \lambda \)-term which is not \( \lambda s_e \)-strongly normalizable.

The system \( \lambda s_e \) has been used successfully in areas connected to theorem proving, namely higher order unification [12].

Another interesting feature is that it supports explicit substitutions attached to any subterm (unlike Typelab for example) which would give efficiency advantages to a theorem proving system based on an eventual extension of \( \lambda s_e \) with dependent types. The results obtained so far for \( \lambda s_e \) compare very well to similar results in systems based on the \( \lambda \sigma \)-calculus, which is the basis for \PiC which we discussed earlier. Therefore it will not be surprising if an extension of \( \lambda s_e \) soon joins the systems of open terms for theorem proving.
### 3.3.1 Classification Summary: $\lambda_s$$_e$

<table>
<thead>
<tr>
<th>Properties</th>
<th>Technical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>Uses MV</td>
</tr>
<tr>
<td>Subject Reduction</td>
<td>Explicit Subst.</td>
</tr>
<tr>
<td>Typechecking</td>
<td>Based on Untyped/Simply-typed</td>
</tr>
<tr>
<td>Type inference</td>
<td>$\lambda$-calculus</td>
</tr>
<tr>
<td>Weak Normalization</td>
<td>de Bruijn Ind.</td>
</tr>
<tr>
<td>Strong Normalization</td>
<td>PSN fails</td>
</tr>
</tbody>
</table>
Chapter 4

Discussion

In this chapter we want to compare the different systems, to discuss the problems they face and the ways to solve them.

4.1 Common features

Although the systems discussed are quite different from each other, there are several issues on which they agree:

- **Computational effects** It seems that all calculi of open terms must in some way deal with the computational issues of open terms. All systems discussed address the issue of performing a substitution on an open term. Although the approaches are different, they all find some answer to this problem.

- **Representation of holes by (meta)variables** A common trend is to represent unknown terms by (meta)variables. A notable agreement between all the systems is that the main unknown object is a term and therefore metavariables are usually introduced to stand for terms. For some calculi having other syntactic categories (substitutions, types, etc.) metavariables can also be introduced for objects of these categories, but not always successfully.

- **Uniqueness of the type and context of the holes** If allowed, multiple occurrences of a metavariable must agree on their types and contexts. We have seen that in examples where this condition was broken (see e.g. 2.1.2), usually the soundness of the instantiation of the metavariables is compromised.

The uniqueness condition brings into the game some additional structure which describes the contexts and the types of the metavariables. In Typelab this is the proof problem, in OLEG these are the metavariable
binders, in $\lambda\Pi_C$ we have signatures, etc. Except for OLEG where this information is internalized, this external structure parameterizes the whole system – note for example that the type inference rules in Typelab depend on a proof problem that is fixed in advance. Another use of such a structure is to contain the additional information as for example constraints or restrictions on the instantiations of the MVs.

4.2 Open Terms and Computation

As we can see from the descriptions of the different systems, the main issue when introducing open terms is how to handle interactions between them and the computation rules of the system. It seems that the two approaches adopted in the systems discussed are either to avoid troublesome computations or to introduce extra facilities which help solve the problem.

Avoiding troublesome computations in dependently-typed systems is done by imposing restrictions on the reduction and/or the typing rules of the system which guarantee behaviour of the open terms close to the case of simply-typed systems avoiding the problems of dependent types.

On the other hand, systems which need to have a more liberal use of metavariables introduce a mechanism to offset the negative effects of computations on open terms. These negative effects may be loss of confluence or non-commuting of metavariable instantiations and computation.

We can identify two main approaches to achieving this goal:

- Explicit substitutions. They are used to delay the execution of a substitution over an open term, so that it is not lost.

- Functional encoding of the scope. The holes in this case are considered as functions applied to the variables which are allowed to be free in the instantiations of a hole. Then a substitution is recorded by its effect on the arguments of the hole. Its effect on the hole variable is ignored because it stands for a closed term and therefore is not affected by substitutions.

4.3 Computations in ALF, Typelab, OLEG and $\lambda\Pi_C$: Comparison

In this section we want to compare directly the ways computations are done in these systems.

In all of them the classical definition of $\beta$-reduction

$$(\lambda x : A.M)N \rightarrow_\beta M[N/x]$$

where $M[N/x]$ is the term $M$ in which the free occurrences of $x$ are replaced by $N$ is no longer atomic, but is divided into several phases.
Then the second phase formalizes the intermediate steps of distributing the value of \( x \) down the term structure (except for Typelab). ALF and \( \lambda\Pi_C \) try to mimic the propagation of the (meta)substitution by providing rules for computing with explicit substitutions. Unfortunately this very much complicates these calculi and leads to a loss of termination. On the other hand, OLEG takes an intermediate position between Typelab and ALF/\( \lambda\Pi_C \) converting \( \iota \)-redexes to local definitions. Then using the \( \delta \)- and \( \iota \)-contraction schemes it propagates the value of \( x \) in \( M \).

So the original \( \beta \) rule is simulated in the following ways:

<table>
<thead>
<tr>
<th>System</th>
<th>( \beta )-reduction expressed by means of</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALF</td>
<td>Explicit substitutions + substitution rules</td>
</tr>
<tr>
<td>Typelab</td>
<td>(meta)substitution + expl. subst at metavariables</td>
</tr>
<tr>
<td>OLEG</td>
<td>Local definition and a sequence of ( \delta )-reductions</td>
</tr>
<tr>
<td>( \lambda\Pi_C )</td>
<td>Explicit substitutions + substitution rules</td>
</tr>
</tbody>
</table>

These approaches are very different and have their own specifics. For example, full-power explicit substitutions seem to cause more problems than the approach adopted in Typelab, because they allow too much freedom in handling the intermediate steps in substitution reductions. Apart from the termination problem, this causes also consistency problems because of the danger of loss of the Subject Reduction property. Keeping the metasubstitution allows us to discard the intermediate steps which are not essential for the purposes of interactive theorem proving at the expense of worsened implementation efficiency.

### 4.4 Explicit Substitutions

The idea of delaying substitutions can be found in many systems, even ones which do not call themselves explicit substitutions calculi. Explicit substitutions have different forms and names - take for example the redex created by a substitution when it passes over an \( \omega \) in Automath or the explicit substitutions in \( \lambda\omega \), Typelab, etc. Even the definitions in OLEG can be viewed as a tool to delay substitutions (cf. \( (\beta) \)-rule).

In each different case of introduction of explicit substitutions we have different goals and different tools at our disposal. This has led to the creation of a large variety of calculi of explicit substitutions. This paper does not aim at presenting an overview of those systems, one could consult for example [16]. Instead, we are interested in explicit substitutions only as far as they are related to the problems of open terms.

Nevertheless, there are many choices to be made when using explicit substitutions. Allowing too much liberty leads to the loss of properties as confluence, strong normalization or preservation of strong normalization. It seems that
these problems occur more frequently in systems of open terms and explicit substitutions which use de Bruijn indices.

Although using de Bruijn notation leads to some technical difficulties, for systems with full-power explicit substitutions, the fundamental problem of computation in open term is translated into the fundamental problem of all explicit substitution calculus – namely the interaction between substitutions.

Typelab however shows us that maybe this is not an intrinsic feature of open terms since it uses no internal substitution reductions and is still able to perform better in terms of properties of the calculus than the other systems. We may conclude that most of the problems in systems of open terms with computations on explicit substitutions are due to the complexity of the underlying explicit substitution calculus without open terms.

4.5 Typechecking, Constraints and Unification

We conclude our discussion with a look at the connection between open terms and unification. Let us introduce the problem by the following example:

Suppose

\[ \vdash B : Type \]
\[ \vdash ?A : Type \]
\[ \vdash ?X : ?A \rightarrow ?A \]

and consider the instantiation \( ?X = \lambda x : B.x \). The question is do we accept this to be a valid instantiation or not.

ALF and \( \lambda \Pi C \) give a positive answer to this question while for OLEG and Typelab this is not the case. In general, in order to check the validity of an instantiation, we will have to solve an equation \( M = N \) where both \( M \) and \( N \) may contain metavariables. In the case of dependently-typed calculi we clearly will be confronted with a higher-order unification problem which is known to be undecidable in general. But ALF and \( \lambda \Pi C \) have means to delay the checking of these kind of equations by using them as constraints on the future instantiations of the metavariables involved. This leads to a natural incorporation of unification algorithms solving decidable equations into the typechecker.

How does this compare to the situation in OLEG and Typelab – there unification is not explicitly incorporated into the typechecker, but rather applied on typechecked proof problems.

By careful inspection one may conclude that given the same unificational power the two approaches are not fundamentally different and the choice of one or the other depends on external factors.

4.6 Further Work

As we have already mentioned earlier, the focus of all considered systems is the unknown term represented by a metavariable. One possible abstraction of this
principle one level up - namely shifting the attention from terms to \textit{judgements} as abstract objects.

It was already pointed out that a metavariable is always considered together with its type and context. This of course means that we are actually given a judgement

\[ \Gamma_M \vdash M : A_M \]

which can be viewed either as goal or assumption in an interactive derivation session.

On the other hand the derivation rules can be viewed as relations or, with some assumptions, as functions on the set of judgements. In this way of course we obtain the calculus describing derivability of judgements.

If we view the judgements as basic atomic objects, we may then try to describe what an abstract instantiation is and investigate the problem of axiomatizing it. This axiomatization should of course be adequate with respect to the existing systems of open terms that may act as model for such a theory.

The most interesting point, which may turn out to be also the biggest problem with such a calculus is that it totally hides the computational aspect of the underlying system for open terms. Whether this is an advantage that allows generalization and useful abstraction from the underlying machinery or terminal invalidity is an interesting question.

\textbf{Acknowledgements}

The author would like to thank my supervisors for their comments which led to many improvements in this report.
## Appendix A

### Classification Summary

<table>
<thead>
<tr>
<th>Technical Issues</th>
<th>ALF</th>
<th>Typelab</th>
<th>OLEG</th>
<th>$\lambda \Pi_c$</th>
<th>$\lambda_c$</th>
<th>Automath</th>
<th>$\lambda \delta_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses MV</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Explicit Subst.</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>functional</td>
<td>create</td>
<td>yes</td>
</tr>
<tr>
<td>Based on</td>
<td>Martin-Löf</td>
<td>ECC</td>
<td>ECC</td>
<td>$\lambda \Pi$</td>
<td>not decidable</td>
<td>untyped</td>
<td>untyped/full $\lambda$-calculus</td>
</tr>
<tr>
<td>de Bruijn Ind.</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Properties</th>
<th>ALF</th>
<th>Typelab</th>
<th>OLEG</th>
<th>$\lambda \Pi_c$</th>
<th>$\lambda_c$</th>
<th>Automath</th>
<th>$\lambda \delta_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>MTA</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Subject red.</td>
<td>?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Type checking</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
</tr>
<tr>
<td>Type inference</td>
<td>not decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
<td>decidable</td>
</tr>
<tr>
<td>Weak Norm.</td>
<td>MTA</td>
<td>yes</td>
<td>yes</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>yes</td>
</tr>
<tr>
<td>Strong Norm.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>PSN fails</td>
</tr>
</tbody>
</table>
## Contents

1.1 Introduction .............................................. 1
1.2 Issues of Interest and Classification Criteria ............ 2
1.3 Informal Use of the Metavariabes in Interactive Theorem Proving 3
   1.3.1 Metavariabes standing for (proof)terms ............. 3
   1.3.2 Metavariabes as witnesses .......................... 4
   1.3.3 Computing with metavariables ...................... 4
1.4 Main Problem to Solve When Introducing Metavariabes ..... 5

2 Open Term Systems for Theorem Proving ........................ 7
2.1 ALF ............................................. 7
   2.1.1 Syntax of the Calculus ........................... 7
   2.1.2 Adding incomplete (open) terms ................... 8
   2.1.3 The Transitivity Example ........................ 11
   2.1.4 Classification Summary: ALF ........................ 11
2.2 Typelab ........................................... 12
   2.2.1 Description of the Calculus ....................... 12
   2.2.2 The Transitivity Example ........................ 15
   2.2.3 Classification Summary: Typelab ................... 15
2.3 OLEG ............................................ 15
   2.3.1 The syntax .................................. 16
   2.3.2 The Transitivity Example ........................ 18
   2.3.3 Classification Summary: OLEG ..................... 18
2.4 The $\lambda\Pi_X$-calculus ................................ 19
   2.4.1 Description of $\lambda\Pi_X$ ....................... 19
   2.4.2 The Transitivity Example ........................ 20
   2.4.3 Classification Summary: $\lambda\Pi_X$ ............... 21

3 Other Calculi of Open Terms .................................. 22
3.1 The Context Calculus $\lambda c$ ............................ 22
   3.1.1 Contexts in untyped $\lambda$-calculus as open terms .... 22
   3.1.2 Description of the Calculus ....................... 23
   3.1.3 Classification Summary: $\lambda c$ .................. 25
3.2 Automath ............................................ 25
   3.2.1 The calculus of segments .......................... 25
   3.2.2 Properties and Extensions of the System ............. 27
3.2.3 Classification Summary: Automath ........................................ 27
3.3 The \( \Lambda s_e \)-calculus ......................................................... 27
3.3.1 Classification Summary: \( \Lambda s_e \) ..................................... 29

4 Discussion ................................................................................. 30
4.1 Common features ................................................................. 30
4.2 Open Terms and Computation ............................................... 31
4.3 Computations in ALF, Typelab, OLEG and \( \Lambda \Pi_e \): Comparison . 31
4.4 Explicit Substitutions ............................................................. 32
4.5 Typechecking, Constraints and Unification .............................. 33
4.6 Further Work ......................................................................... 33

A Classification Summary .......................................................... 35
Bibliography


# Computer Science Reports

**Department of Mathematics and Computer Science**  
**Technische Universiteit Eindhoven**

If you want to receive reports, send an email to: m.m.j.l.philips@tue.nl (we cannot guarantee the availability of the requested reports)

**In this series appeared:**

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>97/06</td>
<td>M. Franssen</td>
<td>Tools for the Construction of Correct Programs: an Overview, p. 33.</td>
</tr>
<tr>
<td>97/07</td>
<td>J.C.M. Baeten and J.A. Bergstra</td>
<td>Bounded Stacks, Bags and Queues, p. 15.</td>
</tr>
<tr>
<td>97/08</td>
<td>P. Hoogendijk and R.C. Backhouse</td>
<td>When do datatypes commute, p. 35.</td>
</tr>
<tr>
<td>97/13</td>
<td>W.M.P. van der Aalst</td>
<td>Exploring the Process Dimension of Workflow Management, p. 56.</td>
</tr>
<tr>
<td>97/14</td>
<td>J.F. Groote, F. Monin and J. Springintvelde</td>
<td>A computer checked algebraic verification of a distributed summation algorithm, p. 28.</td>
</tr>
<tr>
<td>97/15</td>
<td>M. Franssen</td>
<td>$\lambda p$: A Pure Type System for First Order Logic with Automated Theorem Proving, p. 35.</td>
</tr>
<tr>
<td>97/16</td>
<td>W.M.P. van der Aalst</td>
<td>On the verification of Inter-organizational workflows, p. 23</td>
</tr>
<tr>
<td>97/17</td>
<td>M. Vaceari and R.C. Backhouse</td>
<td>Calculating a Round-Robin Scheduler, p. 23.</td>
</tr>
<tr>
<td>98/01</td>
<td>W. Van der Aalst</td>
<td>Formalization and Verification of Event-driven Process Chains, p. 26</td>
</tr>
<tr>
<td>98/02</td>
<td>M. Voorhoeve</td>
<td>State / Event Net Equivalence, p. 25</td>
</tr>
<tr>
<td>98/03</td>
<td>J.C.M. Baeten and J.A. Bergstra</td>
<td>Deadlock Behaviour in Split and ST Bisimulation Semantics, p. 15.</td>
</tr>
<tr>
<td>98/04</td>
<td>R.C. Backhouse</td>
<td>Pair Algebras and Galois Connections, p. 14</td>
</tr>
<tr>
<td>98/05</td>
<td>D. Danis</td>
<td>Flat Fragments of CTL and CTL*: Separating the Expressive and Distinguishing Powers. P. 22.</td>
</tr>
</tbody>
</table>
| 98/07 | Proceedings of the workshop on Workflow Management: Net-based Concepts, Models, Techniques and Tools (WFM'98)  
| 98/08 | Informal proceedings of the Workshop on User Interfaces for Theorem Provers.  
K.M. van Hee and H.A. Reijers
G. van Hee and H.A. Reijers
An analytical method for assessing business processes, p. 29.

T. Basten and J. Hooman
Process Algebra in PVS

J. Zwanenburg
The Proof-assistant Yarrow, p. 15

Ninth ACM Conference on Hypertext and Hypermedia
Hypertext '98
Pittsburgh, USA, June 20-24, 1998
Proceedings of the second workshop on Adaptive Hypertext and Hypermedia.
Edited by P. Brusilovsky and P. De Bra, p. 95.

J.F. Groote, F. Morin and J. v.d. Pol
Checking verifications of protocols and distributed systems by computer.
Extended version of a tutorial at CONCUR'98, p. 27.

V. Bos and J.J.T. Kleijn
Structured Operational Semantics of \( \chi \), p. 27

H.M.W. Verbeek, T. Basten and W.M.P. van der Aalst
Diagnosing Workflow Processes using Woflan, p. 44

R.C. Backhouse and P. Hoogendijk
Final Dialgebra: From Categories to Allegories, p. 81

M. Franssen, R.C. Veltkamp and W. Wessels
Efficient Evaluation of Triangular B-splines, p. 13

T. Basten and W. v.d. Aalst
Inheritance of Workflows: An Approach to tackling problems related to change, p. 66

P. Brusilovsky and P. De Bra

D. Bosnacki, S. Mauw, and T. Willemse

J. v.d. Pol, J. Hooman and E. de Jong
Requirements Specification and Analysis of Command and Control Systems

T.A.C. Willemse
The Analysis of a Conveyor Belt System, a case study in Hybrid Systems and timed \( \mu \) CRL, p. 44.

J.C.M. Baeten and C.A. Middelburg
Process Algebra with Timing: Real Time and Discrete Time, p. 50.

S. Andova

K.M. van Hee, R.A. van der Toorn, J. van der Woude and P.A.C. Verkoulen
A Framework for Component-based Software Architectures, p. 19

A. Engels and S. Mauw
Why men (and octopuses) cannot juggle a four ball cascade, p. 10

J.F. Groote, W.H. Hesselink, S. Mauw, R. Vermeulen
An algorithm for the asynchronous Write-All problem based on process collision*, p. 11.

G.J. Houben, P. Lemmens

T. Basten, W.M.P. v.d. Aalst
Inheritance of Behavior, p.83

J.C.M. Baeten and T. Basten
Partial-Order Process Algebra (and its Relation to Petri Nets), p. 79

J.C.M. Baeten and C.A. Middelburg
Real Time Process Algebra with Time-dependent Conditions, p.33.

Proceedings Conferentie Informatiewetenschap 1999
Centrum voor Wiskunde en Informatica
12 november 1999, p.98
edited by P. de Bra and L. Hardman

J.C.M. Baeten and J.A. Bergstra
Mode Transfer in process Algebra, p. 14

J.C.M. Baeten
Process Algebra with Explicit Termination, p. 17.

S. Mauw and M.A. Reniers
A process algebra for interworkings, p. 63.

R. Bloo, J. Hooman and E. de Jong
Semantic Aspects of an Architecture for Distributed Embedded Systems*, p. 47.
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>00/05</td>
<td>Algebraic Process Verification, p. 65.</td>
<td>J.F. Groote and M.A. Reniers</td>
</tr>
<tr>
<td>00/06</td>
<td>The Parallel Composition of Uniform Processes with Data, p. 19</td>
<td>J.F. Groote and J. v. Wamel</td>
</tr>
<tr>
<td>00/07</td>
<td>Variable Binding Operators in Transition System Specifications, p. 27.</td>
<td>C.A. Middelburg</td>
</tr>
<tr>
<td>00/08</td>
<td>Grammars Compared: A study on determining a suitable grammar for parsing and generating natural language sentences in order to facilitate the translation of natural language and MSC use cases, p. 33.</td>
<td>I.D. van den Ende</td>
</tr>
<tr>
<td>00/09</td>
<td>A Formal Development of Distributed Summation, p. 35</td>
<td>R.R. Hoogerwoord</td>
</tr>
<tr>
<td>00/11</td>
<td>Enhancing Partial-Order Reduction via Process Clustering, p. 14</td>
<td>T. Basten and D. Bošnački</td>
</tr>
<tr>
<td>00/12</td>
<td>Message Sequence Charts in the Software Engineering Process, p. 26</td>
<td>S. Mauw, M.A. Reniers and T.A.C. Willemse</td>
</tr>
<tr>
<td>00/13</td>
<td>Termination in Timed Process Algebra, p. 36</td>
<td>J.C.M. Baeten, M.A. Reniers</td>
</tr>
<tr>
<td>00/14</td>
<td>Impossible Futures and Determinism, p. 14</td>
<td>M. Voorhoeve, S. Mauw</td>
</tr>
<tr>
<td>00/15</td>
<td>An Interactive Viewer for Mathematical Content based on Type Theory, p. 24.</td>
<td>M. Oostdijk</td>
</tr>
<tr>
<td>00/16</td>
<td>Characterizing $\lambda$-terms with equal reduction behavior, p. 12</td>
<td>F. Kamareddine, R. Bloo, R. Nederpelt</td>
</tr>
<tr>
<td>00/17</td>
<td>Belief Revision with Explicit Justifications: an Exploration in Type Theory, p. 30.</td>
<td>T. Borghuis, R. Nederpelt</td>
</tr>
<tr>
<td>00/18</td>
<td>Parameters in Pure Type Systems, p. 41</td>
<td>T. Laan, R. Bloo, F. Kamareddine, R. Nederpelt</td>
</tr>
<tr>
<td>00/19</td>
<td>Specifying Internet applications with DiCons, p. 9</td>
<td>J. Baeten, H. van Beek, S. Mauw</td>
</tr>
<tr>
<td>01/01</td>
<td>A Rewriting Approach to Binary Decision Diagrams, p. 27</td>
<td>H. Zaniema and J. v.d. Pol</td>
</tr>
<tr>
<td>01/02</td>
<td>Interpretations of Automata, p. 41</td>
<td>T.A.C. Willemse</td>
</tr>
</tbody>
</table>