

GROUPOID INFINITY
KYIV, UKRAINE

Intermediate Language with Dependent Types and Strong Normalization for Erlang/OTP applications.

Technical Article

Maxim Sokhatsky, Synrc Research Center

Kyiv 2016—2017

Contents

1 Introduction

1.1 History

LISP. Untyped lambda calculus was discovered as an inner language of the space at origin (Curry, Church, 1932). This language was manifested as LISP (McCarthy, 1958) that was built upon: cons, nil, eq, atom, car, cdr, lambda, apply and id. It was parts of inductive types lately known as inductive type constructors. Still untyped lambda calculus is used as an extraction target for many provers (Idris, F*), and also manifests in different domain languages (JavaScript, Erlang).

ML/LCF. Further teardown of inner space language was ML language, founded merely on algebraic datatypes and algebra on higher terms rather than categorical semantic. Lately it was fixed with categorical methods in CPL (Hagino, 1987) and Charity (Cockett, 1992). Milner, assisted by Morris and Newey designed Meta Language for the purpose of builing LCF in early 70-s. LCF was a predecessor family of automated math provers: HOL88, HOL90, HOL98 and HOL/Isabelle which is now built using Poly/ML.

Fully Automated Provers. In that period during 80-90s other automated math systems were appeared: AUTOMATH (de Bruijn, 1967), Mizar (Trybulec, 1989), PVS (Owre, Rushby, Shankar, 1995), ACL2 (Boyer, Kaufmann, Moore, 1996) and Otter (McCune, 1996).

MLTT. Contemporary provers (built upon consistent Martin-Löf Type Theory, 1972) like Agda, Coq, Lean, F*, Idris are based on Barendregt and Coquand' Calculus of Constructions with different flavours of inifinity universe hierarchies and Calculus of Inductive Constructions for modeling polynomial functors of well-founded trees. Some of them are automated and some are trying to be and general purpose programming languages with proving facilities, like Idris, Coq (coq.io), Agda (M-Alonso).

Cubical Type Theory The further extensions of MLTT theory is a Homotopy Types needed for reasoning about abstract shapes and multidimentional path types. The Cubical approach was chosen in favour of Globular and Simplical due to computability of Voevodsky's univalence axiom.

1.2 Background

From the partical point of view there are exists two approaches. One involve the usage of two languages: meta language for your models and separate language for prover, like in **HOL**, **Andromeda**. The second approach propose embedding models into single language. The only available prover that fits this criteria today is **cubicaltt**. The more

2 Motivation and Vision

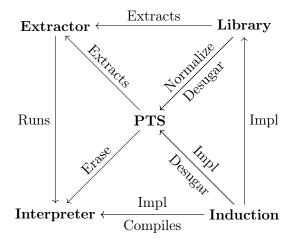
From PTS to HTS. We want to have flexible detachable layers on top of PTS core. Then Sigma for proving. Then well-founded trees or polynomial functors as known as data and record. Then higher path types, interval arithmetic, glue and comp for HIT. Each layers is driven by differenth math, the common in only the method – category theory.

Extensible Language Design. Encoding of inductive types is based on categorical semantic of compilation to PTS. All other syntax constructions are inductive definitions, plugged into the stream parser. AST of the PTS language is also defined in terms of inductive constructions and thus allowed in the macros. The language of polynomial functors (data and record) and core language of the process calculus (spawn, receive and send) are just macrosystem over Om language, its syntax extensions.

Changable Encodings. In pure CoC we have only arrows, so all inductive type encodings would be Church-encoding variations. Most extended nowadays is Church-Boehm-Berrarducci encoding, which dedicated to inductive types. Another well known are Scott (lazyness), Parigot (lazyness and constant-time iterators) and CPS (continuations) encodings. However most of them require variations of Fixpoint types.

Proved Categorical Semantic. There was modeled a math model (using higher-order categorical logic) of encoding, which calculates (co)limits in a cathegory of (co)algebras built with given set of (de)constructors. We call such encoding in honour of Lambek lemma that leeds us to the equality of (co)initial object and (co)limit in the categories of (co)algebras. Such encoding works with dependent types and its consistency is proved in Lean model.

General Architecture. This article covers only the central part of of the proving system, the **OM** intermediate language with strong normalization properties and extraction to Erlang/OTP bytecode.



3 Pure Type System as Intermediate Language

The Om language is a dependently typed lambda calculus, an extension of Barendregt' and Coquand Calculus of Constructions with predicative hierarchy of indexed universes. There is no fixpoint axiom needed for the definition of infinity term dependance.

All terms respect ranking Axioms inside sequence of universes Sorts and complexity of the dependent term is equal maximum complexity of term and its dependency Rules. The type system is completely described by the following PTS notation (due to Barendregt):

```
\begin{cases} Sorts = Type.\{i\}, \ i: Nat \\ Axioms = Type.\{i\}: Type.\{inc\ i\} \\ Rules = Type.\{i\} \leadsto Type.\{j\}: Type.\{max\ i\ j\} \end{cases}
```

An intermediate Om language is based on Henk [?] languages described first by Erik Meijer and Simon Peyton Jones in 1997. Leter on in 2015 Morte impementation of Henk design appeared in Haskell, using Boem-Berrarducci encoding of non-recursive lambda terms. It is based only on one type constructor Π , its special case λ and theirs eliminators: apply and curry, infinity number of universes, and one computation rule called β -reduction. The design of Om language resemble Henk and Morte both design and implementation. This language indended to be small, concise, easy provable and able to produce verifiable peace of code that can be distributed over the networks, compiled at target with safe trusted linkage.

3.1 BNF

Om syntax is compatible with λC Coquand's Calculus of Constructions presented in Morte and Henk languages. However it has extension in a part of specifying universe index as a **Nat** number.

Equivalent tree encoding for parsed terms is following:

```
Inductive OM := Star: \operatorname{nat} \to \operatorname{OM} | Var: \operatorname{name} \to \operatorname{OM} | App: \operatorname{OM} \to \operatorname{OM} \to \operatorname{OM} | Lambda: \operatorname{name} \to \operatorname{OM} \to \operatorname{OM} \to \operatorname{OM} | Arrow: \operatorname{OM} \to \operatorname{OM} \to \operatorname{OM} | Pi: \operatorname{name} \to \operatorname{OM} \to \operatorname{OM} \to \operatorname{OM}.
```

3.2 Universes

The OM language is a higher-order dependently typed lambda calculus, an extension of Coquand's Calculus of Constructions with the predicative/impredicative hierarchy of indexed universes. This extension is motivated avoiding paradoxes in dependent theory. Also there is no fixpoint axiom needed for the definition of infinity term dependance.

3.3 Predicative Universes

All terms obey the A ranking inside the sequence of S universes, and the complexity R of the dependent term is equal to a maximum of the term's complexity and its dependency. The universes system is completely described by the following PTS notation (due to Barendregt):

```
S (n : nat) = U n A_1 (n m : nat) = U n : U m where m > n - cumulative R_1 (m n : nat) = U m \rightarrow U n : U (max m n) - predicative
```

Note that predicative universes are incompatible with Church lambda term encoding. You can switch predicative vs impredicative universes by typecheker parameter.

$$\frac{i: Nat, j: Nat, i < j}{Type_i: Type_j} \tag{A_1}$$

$$\frac{i: Nat, j: Nat}{Type_i \to Type_j: Type_{max(i,j)}} \tag{R_1}$$

3.4 Impredicative Universes

Propositional contractible bottom space is the only available extension to predicative hierarchy that not leads to inconsistency. However there is another option to have infinite impredicative hierarchy.

$$A_2$$
 (n : nat) = U n : U (n + 1) - non-cumulative R_2 (m n : nat) = U m \rightarrow U n : U n - impredicative

$$\frac{i:Nat}{Type_i:Type_{i+1}} \tag{A_2}$$

$$\frac{i: Nat, \quad j: Nat}{Type_i \to Type_j: Type_j} \tag{R2}$$

3.5 Single Axiom Language

This language is called one axiom language (or pure) as eliminator and introduction adjoint functors inferred from type formation rule. The only computation rule of Pi type is called beta-reduction.

```
\forall \ (\mathbf{x}: \ \mathbf{A}) \to \mathbf{B} \ \mathbf{x} : \ \mathsf{Type} \\ \lambda \ (\mathbf{x}: \ \mathbf{A}) \to \mathbf{b} : \mathbf{B} \ \mathbf{x} \\ \mathbf{f} \ a : \ \mathbf{B} \ [\mathbf{a}/\mathbf{x}] \\ (\ \lambda \ (\mathbf{x}: \ \mathbf{A}) \to \mathbf{b}) \ a = \mathbf{b} [\mathbf{a}/\mathbf{x}] : \ \mathbf{B} [\mathbf{a}/\mathbf{x}] \\ \frac{x : A \vdash B : Type}{\Pi \ (x : A) \to B : Type} \\ \frac{x : A \vdash b : B}{\lambda \ (x : A) \to b : \Pi \ (x : A) \to B} \\ \frac{f : (\Pi \ (x : A) \to B) \quad a : A}{f \ a : B \ [a/x]} \\ \frac{x : A \vdash b : B \quad a : A}{(\lambda \ (x : A) \to b) \ a = b \ [a/x] : B \ [a/x]} \\ (\beta \text{-computation})
```

This language could be embedded in itself and used as Logical Framework for the Pi type:

3.6 Hierarchy

H returns the target Universe of B term dependendance on A. There are two dependance rules known as predicative and impredicative which returns max universe or universe of last term respectively.

```
dep A B impredicative → B
dep A B predicative → max A B
h Arg Out → dep Arg Out om:hierarchy(impredicative)
```

3.7 Universes

Star returns the number of Universe. If argument is not Universe it returns {error, }.

3.8 Functions

Func returns true if the argument is functional space. Otherwise returns $\{error, \}$.

```
func ((:forall,),(I,0)) → true
func T → (:error,(:forall,T))
```

3.9 Variables

Var returns true if the var N is defined in dictionary B. Otherwise returns $\{error, \}$.

3.10 Shift

Shift renames var N in N

3.11 Substitution

```
sub Term Name Value

ightarrow sub Term Name Value 0
                           (I,0)) N V L \rightarrow (:arrow,
sub (:arrow,
                                                                    sub I N V L, sub 0 N V L);
sub ((:forall,(N,0)),(I,0)) N V L \rightarrow ((:forall,(N,0)),sub I N V L,sub 0 N(sh V N 0)L+1)
sub ((:forall,(F,X)),(I,0)) N V L \rightarrow ((:forall,(F,X)),sub I N V L,sub 0 N(sh V F 0)L)
 \hbox{sub } ((:lambda\,,(N\,,0))\,,(I\,,0)) \ \ N \ \ V \ \ \rightarrow \ \ ((:lambda\,,(N\,,0))\,,\\ \hbox{sub } \ \ I \ \ N \ \ V \ \ L\,,\\ \hbox{sub } \ \ 0 \ \ N(sh \ \ V \ \ N \ \ 0)L+1) 
sub ((:lambda,(F,X)),(I,0)) N V L \rightarrow ((:lambda,(F,X)),sub I N V L,sub 0 N(sh V F 0)L)
                           (F,A)) N V L \rightarrow (:app,sub F N V L,sub A N V L)
sub (:app,
                           (N,L)) N V L \rightarrow V
sub (:var,
sub (:var,
                           (N,I)) N V L when I>L \rightarrow (:var,(N,I-1))
sub T
                                    \_ \_ \to T.
```

3.12 Normalization

```
norm : none

ightarrow :none
norm : any

ightarrow :any
                                     \rightarrow case norm F of
norm (:app,(F,A))
                                         ((:lambda,(N,0)),(I,0)) \rightarrow norm (subst 0 N A)
                                                                    NF \rightarrow (:app,(NF,norm A)) end
norm (:remote,N)
                                     \rightarrow cache (norm N [])
                            (I,0)) \rightarrow ((:forall,("_",0)),(norm I,norm 0))
norm (:arrow,
norm ((:forall,(N,0)),(I,0)) \rightarrow ((:forall,(N,0)), (norm I,norm 0))
 \text{norm } ((:lambda,(N,0)),(I,0)) \rightarrow ((:lambda,(N,0)), \quad (norm I,norm 0)) 
                                     \rightarrow T
norm T
```

3.13 Definitional Equality

```
eq ((:forall,("_",0)), X) (:arrow,Y)
                                                \rightarrow eq X Y
eq (:app,(F1,A1))
                              (:app,(F2,A2)) \rightarrow let true = eq F1 F2 in eq A1 A2
eq (:star,N)
                              (:star,N)
                                                \rightarrow true
eq (:var,(N,I))
                              (: var,(N,I))
                                                \rightarrow true
eq (:remote,N)
                              (:remote,N)
                                                \rightarrow true
eq ((:farall,(N1,0)),(I1,01))
   ((:forall,(N2,0)),(I2,02)) \rightarrow
   let true = eq I1 I2 in eq 01 (subst (shift 02 N1 0) N2 (:var,(N1,0)) 0)
eq ((:lambda,(N1,0)),(I1,01))
   ((:lambda,(N2,0)),(I2,02)) \rightarrow
   let true = eq I1 I2 in eq O1 (subst (shift O2 N1 0) N2 (:var,(N1,0)) 0)
eq (A,B)
                                   \rightarrow (:error,(:eq,A,B))
```

3.14 Type Checker

```
type (:star,N)
                                     \_ \rightarrow (:star,N+1)
type (:var,(N,I))
                                     {	t D} \, 	o \, {	t let} \, {	t true} \, = \, {	t var} \, {	t N} \, {	t D} \, {	t in} \, {	t keyget} \, {	t N} \, {	t D} \, {	t I}
type (:remote,N)
                                     D \rightarrow cache type N D
type (:arrow,(I,0))
                                     D → (:star,h(star(type I D)),star(type O D))
type ((:forall,(N,0)),(I,0)) D \rightarrow (:star,h(star(type\ I\ D)),star(type\ O\ [(N,norm\ I)|D]))
type ((:lambda,(N,0)),(I,0)) D \rightarrow let star (type I D),
                                            NI = norm I in ((:forall,(N,0)),(NI,type(0,[(N,NI)|D])))
                                     D \rightarrow let T = type(F,D),
type (:app,(F,A))
                                            true = func T,
                                            ((:forall,(N,0)),(I,0)) = T,
                                            Q = type A D,
                                            true = eq I Q in norm (subst 0 N A)
```

3.15 Target Erlang VM and LLVM platforms

This work expects to compile to limited target platforms. For now Erlang, Haskell and LLVM are awaiting. Erlang version is expected to be useful both on LING and BEAM Erlang virtual machines. This language allows you to define trusted operations in System F and extract this routines to Erlang/OTP platform and plug as trusted resourses. As example we also provide infinite coinductive process creation and inductive shell that linked to Erlang/OTP IO functions directly.

3.16 Recursive Example

3.17 Corecursive Example

4 Language with Inductive Types

Exe is a general purpose functional language with functors, lambdas on types, recursive algebraic types, higher order functions, corecursion, free monad for effects encoding. It compiles to a small MLTT core of dependent type system with inductive types and equality.

4.1 BNF

4.2 Compiler Passes

The underlying OM typechecker and compiler is a target language for EXE general purpose language.

 $\begin{array}{lll} EXPAND & IND-Macroexpansion \\ NORMAL & PTS-Term\ normalization\ and\ typechecking \\ ERASE & PTS-Delete\ information\ about\ types \\ COMPACT & PTS-Term\ Compactification \\ EXTRACT & PTS-Extract\ Erlang\ Code \\ \end{array}$

4.3 AST

The model in Cubical and Coq of the Exe language is available at infinity¹ repository of groupoid organization.

```
data tele (A: U)
                   = emp | tel (n: name) (b: A) (t: tele A)
data branch (A: U) =
                           br (n: name) (args: list name) (term: A)
data label (A: U) =
                           lab (n: name) (t: tele A)
data ind
   = star
                                  (n: nat)
   | var
            (n: name)
                                  (i: nat)
   | app
                      (f a: ind)
   | lambda (x: name) (d c: ind)
            (x: name) (d c: ind)
   | pi
            (n: name) (a b: ind)
   | sigma
                      (d c: ind)
   arrow
   | pair
                      (a b: ind)
   | fst
                      (p:
                             ind)
                            ind)
   snd
                      (p:
   | id
                      (a b: ind)
   | idpair
                      (a b: ind)
   | idelim
                      (a b c d e: ind)
                                                list (label ind))
   | data_
            (n: name) (t: tele ind) (labels:
   case
            (n: name) (t: ind)
                                     (branches: list (branch ind))
            (n: name)
                                     (args:
                                                list ind)
   | ctor
```

¹https://github.com/groupoid/infinity/tree/master/base

4.4 Inductive Types

There are two types of recursion: one is least fixed point (as $F_A X = 1 + A \times X$ or $F_A X = A + X \times X$), in other words the recursion with a base (terminated with a bounded value), lists and trees are examples of such recursive structures (so we call induction recursive sums); and the second is greatest fixed point or recursion without a base (as $F_A X = A \times X$) — such kind of recursion on infinite lists (codata, streams, coinductive types) we can call recursive products.

4.5 Polynomial Functors

Least fixed point trees are called well-founded trees and encode polynomial functors.

Natural Numbers: $\mu X \rightarrow 1 + X$

List A: $\mu X \to 1 + A \times X$

Lambda calculus: $\mu X \rightarrow 1 + X \times X + X$

Stream: $\nu X \to A \times X$

Potentialy Infinite List A: $\nu X \rightarrow 1 + A \times X$

Finite Tree: $\mu X \rightarrow \mu Y \rightarrow 1 + X \times Y = \mu X = List X$

As we know there are several ways to appear for variable in recursive algebraic type. Least fixpoint are known as an recursive expressions that have a base of recursion Both recursive and corecursive datatypes could be encoded using Boem-Berarducci encoding as an non-recursive definitions of folds that include in indentity signature all the constructor components of (co)inductive type.

4.6 Lists

The data type of lists over a given set A can be represented as the initial algebra $(\mu L_A, in)$ of the functor $L_A(X) = 1 + (A \times X)$. Denote $\mu L_A = List(A)$. The constructor functions $nil: 1 \to List(A)$ and $cons: A \times List(A) \to List(A)$ are defined by $nil = in \circ inl$ and $cons = in \circ inr$, so in = [nil, cons]. Given any two functions $c: 1 \to C$ and $h: A \times C \to C$, the catamorphism $f = ([c, h]): List(A) \to C$ is the unique solution of the equation system:

$$\begin{cases} f \circ nil = c \\ f \circ cons = h \circ (id \times f) \end{cases}$$

where f = foldr(c, h). Having this the initial algebra is presented with functor $\mu(1 + A \times X)$ and morphisms sum $[1 \to List(A), A \times List(A) \to List(A)]$ as catamorphism. Using this encoding the base library of List will have following form:

```
\begin{cases} foldr = ([f \circ nil, h]), f \circ cons = h \circ (id \times f) \\ len = ([zero, \lambda \ a \ n \rightarrow succ \ n]) \\ (++) = \lambda \ xs \ ys \rightarrow ([\lambda(x) \rightarrow ys, cons])(xs) \\ map = \lambda \ f \rightarrow ([nil, cons \circ (f \times id)]) \end{cases}
\begin{aligned} & \text{data list: (A: *)} \rightarrow * := \\ & \text{(nil: list A)} \\ & \text{(cons: A \rightarrow list A \rightarrow list A)} \end{cases} \\ & \{ list = \lambda \ ctor \rightarrow \lambda \ cons \rightarrow \lambda \ nil \rightarrow ctor \\ cons = \lambda \ x \rightarrow \lambda \ xs \rightarrow \lambda \ list \rightarrow \lambda \ cons \rightarrow \lambda \ nil \rightarrow cons \ x \ (xs \ list \ cons \ nil) \\ & nil = \lambda \ list \rightarrow \lambda \ cons \rightarrow \lambda \ nil \rightarrow nil \end{cases} \\ & \text{record lists: (A B: *) :=} \\ & \text{(len: list A \rightarrow list A \rightarrow list A)} \\ & \text{(map: (A \rightarrow B) \rightarrow (list A \rightarrow list B))} \\ & \text{(filter: (A \rightarrow bool) \rightarrow (list A \rightarrow list A))} \end{cases} \\ & \begin{cases} len = foldr \ (\lambda \ x \ n \rightarrow succ \ n) \ 0 \\ (++) = \lambda \ ys \rightarrow foldr \ cons \ ys \\ map = \lambda \ f \rightarrow foldr \ (\lambda x \ xs \rightarrow cons \ (f \ x) \ xs) \ nil \\ filter = \lambda \ p \rightarrow foldr \ (\lambda x \ xs \rightarrow if \ p \ x \ then \ cons \ x \ xs \ else \ xs) \ nil \\ foldl = \lambda \ f \ v \ xs = foldr \ (\lambda \ xg \rightarrow (\lambda \rightarrow g \ (f \ a \ x))) \ id \ xs \ v \end{cases}
```

4.7 Normal Forms

Lists/Map

4.8 Prelude Base Library

```
data Nat: Type :=
         (\textit{Zero: Unit} \ \rightarrow \ \textit{Nat})
         (Succ: Nat \rightarrow Nat)
  data List (A: Type) : Type :=
         (Nil: Unit \rightarrow List A)
         (Cons: A \rightarrow List A \rightarrow List A)
record list: Type :=
         (len: List A \rightarrow integer)
         ((++): List A \rightarrow List A \rightarrow List A)
         (map: (A,B: Type) (A \rightarrow B) \rightarrow (List A \rightarrow List B))
         (filter: (A \rightarrow bool) \rightarrow (List A \rightarrow List A))
record String: List Nat := List.Nil
  data IO: Type :=
         (getLine: (String \rightarrow IO) \rightarrow IO)
         (putLint: String 	o IO)
         (pure: () \rightarrow I0)
record IO: Type :=
         (data: String)
         ([>>=]: ...)
record Morte: Type :=
         (recursive: IO.replicateM Nat.Five
                         (IO.[>>=] IO.data Unit IO.getLine IO.putLine))
```

References

Category Theory

- [1] S.MacLane Categories for the Working Mathematician 1972
- [2] W.Lawvere Conceptual Mathematics 1997
- [3] P.Curien Category theory: a programming language-oriented introduction 2008

Pure Type Systems

- [4] P.Martin-Löf Intuitionistic Type Theory 1984
- [5] T.Coquand The Calculus of Constructions. 1988
- [6] E.Meijer Henk: a typed intermediate language 1997
- [7] H.Barendregt Lambda Calculus With Types 2010

Inductive Type Systems

- [8] F.Pfenning Inductively defined types in the Calculus of Constructions 1989
- [9] P.Wadler Recursive types for free 1990
- [10] N.Gambino Wellfounded Trees and Dependent Polynomial Functors 1995
- [11] P.Dybjer Inductive Famalies 1997
- [12] B.Jacobs (Co)Algebras) and (Co)Induction 1997
- [13] V. Vene Categorical programming with (co)inductive types 2000
- [14] H.Geuvers Dependent (Co)Inductive Types are Fibrational Dialgebras 2015

Homotopy Type Systems

- [15] T.Streicher A groupoid model refutes uniqueness of identity proofs 1994
- [16] T.Streicher The Groupoid Interpretation of Type Theory 1996
- [17] B.Jacobs Categorical Logic and Type Theory 1999
- [18] S.Awodey Homotopy Type Theory and Univalent Foundations 2013
- [19] S.Huber A Cubical Type Theory 2015
- [20] A.Joyal What is an elementary higher topos 2014
- [21] A.Mortberg Cubical Type Theory: a constructive univalence axiom 2017