

# Encoding Higher Inductive Types Without Boilerplate

## A Study in Agda Metaprogramming

Paventhan Vivekanandan  
School of Informatics, Computing and Engineering  
Indiana University  
Bloomington, Indiana, USA  
pvivekan@indiana.edu

David Thrane Christiansen  
Galois, Inc.  
Portland, Oregon, USA  
dtt@galois.com

### Abstract

Higher inductive types are inductive types that include non-trivial higher-dimensional structure, represented as identifications that are not reflexivity. While work proceeds on type theories with a computational interpretation of univalence and higher inductive types, it is convenient to encode these structures in more traditional type theories with mature implementations. However, these encodings involve a great deal of error-prone additional syntax. We present a library that uses Agda’s metaprogramming facilities to automate this process, allowing higher inductive types to be specified with minimal additional syntax.

**Keywords** Higher inductive type, Elaboration, Elimination rules, Computation rules

### 1 Introduction

Type theory unites programming and mathematics in a delightful synthesis, in which we can write programs and proofs in the same language. Work on higher-dimensional type theory has revealed a beautiful higher-dimensional structure, lurking just beyond reach. In particular, higher inductive types provide a natural encoding of many otherwise-difficult mathematical concepts, and univalence lets us work in our type theory the way we do on paper: up to isomorphism. Homotopy type theory, however, is not yet done. We do not yet have a mature theory or a mature implementation.

While work proceeds on prototype implementations of higher-dimensional type theories [4, 10], much work remains before they will be as convenient for experimentation with new ideas as Coq, Agda, or Idris is today. In the meantime, it is useful to be able to experiment with ideas from higher-dimensional type theory in our existing systems. If one is willing to put up with some boilerplate code, it is possible to encode higher inductive types and univalence using a mixture of postulated identities and traditional datatypes. We use a technique developed by Licata [17].

Boilerplate postulates, however, are not just inconvenient, they are also an opportunity to make mistakes. Luckily, this boilerplate code can be mechanically generated using Agda’s

recent support for *elaborator reflection* [7], a paradigm for metaprogramming in an implementation of type theory. An elaborator is the part of the implementation that translates a convenient language designed for humans into a much simpler, more explicit, verbose language designed to be easy for a machine to process. Elaborator reflection directly exposes the primitive components of the elaborator to metaprograms written in the language being elaborated, allowing them to put these components to new uses.

Using Agda’s elaborator reflection, we automatically generate the support code for many useful higher inductive types, specifically those that include additional paths between constructors, but not paths between paths. We automate the production of the datatype definitions, postulated paths, induction principles, and their computational behavior. Angiuli et al.’s encoding of patch theory as a higher inductive type [5] requires approximately 1500 lines of code when represented using Licata’s technique. Using our library, the encoding can be expressed in just 70 lines.

This paper makes the following contributions:

- We describe the design and implementation of a metaprogram that automates an encoding of higher inductive types with one path dimension using Agda’s new metaprogramming system.
- We demonstrate applications of this metaprogram to examples from the literature, including both standard textbook examples of higher inductive types as well as larger systems, including both patch theory and specifying cryptographic schemes.
- This metaprogram serves as an example of the additional power available in Agda’s elaborator reflection relative to earlier metaprogramming APIs.

### 2 Background

#### 2.1 Agda Reflection

Agda’s reflection library enables compile-time metaprogramming. This reflection library directly exposes parts of the implementation of Agda’s type checker and elaborator for use by metaprograms, in a manner that is similar to Idris’s elaborator reflection [6, 7] and Lean’s tactic metaprogramming [14]. The type checker’s implementation is exposed as effects in a monad called TC.

```

macro
  mc1 : Term → Term → TC ⊤
  mc1 exp hole =
    do exp' ← quoteTC exp
    unify hole exp'

sampleTerm : Term
sampleTerm = mc1 (λ (n : Nat) → n)

```

**Figure 1.** A macro that quotes its argument

Agda exposes a representation of its syntax to metaprograms, including datatypes for expressions (called *Term*) and definitions (called *Definition*). The primitives exposed in TC include declaring new metavariables, unifying two *Terms*, declaring new definitions, adding new postulates, computing the normal form or weak head normal form of a *Term*, inspecting the current context, and constructing fresh names. This section describes the primitives that are used in our code generation library; more information on the reflection library can be found in the Agda documentation [1].

TC computations can be invoked in three ways: by macros, which work in expression positions, using the `unquoteDecl` operator in a declaration position, which can bring new names into scope, and using the `unquoteDef` operator in a declaration position, which can automate constructions using names that are already in scope. This preserves the principle in Agda's design that the system never invents a name.

An Agda *macro* is a function of type  $t_1 \rightarrow t_2 \rightarrow \dots \rightarrow \text{Term} \rightarrow \text{TC } \top$  that is defined inside a macro block. Macros are special: their last argument is automatically supplied by the type checker, and consists of a *Term* that represents the metavariable to be solved by the macro. If the remaining arguments are quoted names or *Terms*, then the type checker will automatically quote the arguments at the macro's use site. At some point, the macro is expected to unify the provided metavariable with some other term, thus solving it.

Figure 1 demonstrates a macro that quotes its argument. The first step is to quote the quoted expression argument again, using `quoteTC`, yielding a quotation of a quotation. This doubly-quoted expression is passed, using Agda's new support for Haskell-style `do`-notation, into a function that unifies it with the hole. Because unification removes one layer of quotation, `unify` inserts the original quoted term into the hole. The value of `sampleTerm` is

```
lam visible (abs "n" (var 0 []))
```

The constructor `lam` represents a lambda, and its body is formed by the abstraction constructor `abs` that represents a scope in which a new name "n" is bound. The body of the abstraction is a reference back to the abstracted name using de Bruijn index 0.

The `unquoteTC` primitive removes one level of quotation. Figure 2 demonstrates the use of `unquoteTC`. The macro

```

macro
  mc2 : Term → Term → TC ⊤
  mc2 exp hole =
    do exp' ← unquoteTC exp
    unify hole exp'

sampleSyntax : Nat → Nat
sampleSyntax =
  mc2 (lam visible (abs "n" (var 0 [])))

```

**Figure 2.** A macro that unquotes its argument

`mc2` expects a quotation of a quotation, and substitutes its unquotation for the current metavariable.

The `unquoteDecl` and `unquoteDef` primitives, which run TC computations in a declaration context, will typically introduce new declarations by side effect. A function of a given type is declared using `declareDef`, and it can be given a definition using `defineFun`. Similarly, a postulate of a given type is defined using `declarePostulate`. Figure 3 shows an Agda implementation of addition on natural numbers, while figure 4 demonstrates an equivalent metaprogram that adds the same definition to the context.

```

plus : Nat → Nat → Nat
plus zero b = b
plus (suc n) b = suc (plus n b)

```

**Figure 3.** Addition on natural numbers

In Figure 4, `declareDef` declares the type of `plus`. The constructor `pi` represents dependent function types, but a pattern synonym is used to make it shorter. Similarly, `def` constructs references to defined names, and the pattern synonym ``Nat` abbreviates references to the defined name `Nat`, and `vArg` represents the desired visibility and relevance settings of the arguments. Once declared, `plus` is defined

```

pattern vArg x = arg (arg-info visible relevant) x
pattern _`⇒_ a b = pi (vArg a) (abs "_" b)
pattern `Nat = def (quote Nat) []

unquoteDecl plus =
  do declareDef (vArg plus) (`Nat `⇒ `Nat `⇒ `Nat)
  defineFun plus
    (clause (vArg (con (quote zero) [])) ::
      vArg (var "y") ::
      [])
    (var 0 []) ::
    clause (vArg (con (quote suc)
      (vArg (var "x") :: [])) ::
      vArg (var "y") ::
      [])
    (con (quote suc)
      (vArg (def plus
        (vArg (var 1 []) ::
          vArg (var 0 []) :: [])) ::
        [])) ::
    [])

```

**Figure 4.** Addition, defined by metaprogramming

using `defineFun`, which takes a name and a list of clauses, defining the function by side effect. Each clause consists of a pattern and a right-hand side. Patterns have their own datatype, while right-hand sides are `Terms`. The name `con` is overloaded: in patterns, it denotes a pattern that matches a particular constructor, while in `Terms`, it denotes a reference to a constructor.

## 2.2 Higher Inductive Types

Homotopy type theory [25] is a research program that aims to develop univalent, higher-dimensional type theories. A type theory is *univalent* when equivalences between types are considered equivalent to identifications between types; it is *higher-dimensional* when we allow non-trivial identifications that every structure in the theory must nevertheless respect. Identifications between elements of a type are considered to be at the lowest dimension, while identifications between identifications at dimension  $n$  are at dimension  $n + 1$ . Voevodsky added univalence to type theories as an axiom, asserting new identifications without providing a means to compute with them. While more recent work arranges the computational mechanisms of the type theory such that univalence can be derived, as is done in cubical type theories [4, 10], we are concerned with modeling concepts from homotopy type theory in existing, mature implementations of type theory, so we follow Univalent Foundations Program [25] in modeling paths using Martin-Löf's identity type. Higher-dimensional structure can arise from univalence, but it can also be introduced by defining new type formers that introduce not only introduction and elimination principles, but also new non-trivial identifications.

In homotopy type theories, one tends to think of types not as collections of distinct elements, but rather through the metaphor of topological spaces. The individual elements of the type correspond with points in the topological space, and identifications correspond to paths in this space.

While there is not yet a general schematic characterization of a broad class of higher inductive types along the lines of Dybjer's inductive families [13], it is convenient to syntactically represent the higher inductive types that we know are acceptable as if we had such a syntax. Thus, we sometimes specify a higher inductive type similarly to a traditional inductive type by providing its constructors (*i.e.* its points); we additionally specify the higher-dimensional structure by providing additional constructors for paths. For example, figure 5 describes `Circle`, which is a higher inductive type with one point constructor base and one non-trivial path constructor `loop`.

```
data Circle : Set where
  base : Circle
  loop : base ≡ base
```

**Figure 5.** A specification of a higher inductive type

Agda [19] is a functional programming language with full dependent types and dependent pattern matching. Agda's type theory has gained a number of new features over the years, among them the ability to restrict pattern matching to that subset that does not imply Streicher's Axiom K [9], which is inconsistent with univalence. The convenience of programming in Agda, combined with the ability to avoid axiom K, makes it a good laboratory for experimenting with the idioms and techniques of univalent programming while more practical implementations of univalent type theories are under development.

In Agda, we don't have built-in primitives to support the definition of higher inductive type such as `Circle`. One approach is to use Agda's rewrite rules [8] mechanism to define higher inductive types. In this approach, we define the dependent and non-dependent eliminators of a higher inductive type as parameterized modules inside which we declare the computation rules for points as rewrite rules using `{-# REWRITE , ... #-}` pragma. However, Agda's reflection library do not have interfaces to support introducing new pragmas and defining new modules. Another approach to define higher inductive types is to use Licata's method [17]. According to this method, a higher inductive type is defined using type abstraction inside a module. The module consists of a boiler-plate code segment which defines the higher inductive type using a private base type. Inside the module, the recursion and the induction principles acts on the constructors of the private base type. The abstract type is then exported allowing the reduction rules for point constructors to hold definitionally. For example, `Circle` is defined using Licata's method as follows.

Inside the module `Circle`, the type `S` is defined using a private datatype `S*`. The constructor `base` is defined using `base*` and the path `loop` is given as a postulated identification. The recursion and induction principles are defined by pattern matching on the constructor `base*` of the type `S*`, and thus compute as expected. The clients of `Circle` will not have access to the constructor `base*` of the private type `S*`, as it is not visible outside the module, which prevents them from writing functions that distinguish between multiple constructors of a higher inductive type that may be identified by additional path constructors. The client's *only* access to the constructor is through the provided elimination rules. The following code gives the non-dependent eliminator (sometimes called the *recursion rule*) `recS`.

`recS` ignores the path argument and simply computes to the appropriate answer for the point constructor. The computational behavior for the path constructor `loop` is postulated using reduction rule `βrecS`. The operator `apPerhaps` we should move the discussion of `ap` earlier, so that we don't need the digression is frequently referred to as `cong`, because it expresses that propositional equality is a congruence. However, when viewed through a homotopy

```

module Circle where
  private
    data S* : Set where
      base* : S*

  S : Set
  S = S*

  base : S
  base = base*

  postulate
    loop : base  $\equiv$  base*

  recS : {C : Set}  $\rightarrow$ 
    (cbase : C)  $\rightarrow$ 
    (cloop : cbase  $\equiv$  cbase)  $\rightarrow$ 
    S  $\rightarrow$  C
  recS cbase cloop base* = cbase

  postulate
     $\beta$ recS : {C : Set}  $\rightarrow$ 
      (cbase : C)  $\rightarrow$ 
      (cloop : cbase  $\equiv$  cbase)  $\rightarrow$ 
      ap (recS cbase cloop) loop  $\equiv$  cloop

  indS : {C : S  $\rightarrow$  Set}  $\rightarrow$ 
    (cbase : C base)  $\rightarrow$ 
    (cloop : transport C loop cbase  $\equiv$  cbase)  $\rightarrow$ 
    (circle : S)  $\rightarrow$  C circle
  indS cbase cloop base* = cbase

  postulate
     $\beta$ indS : {C : S  $\rightarrow$  Set}  $\rightarrow$ 
      (cbase : C base)  $\rightarrow$ 
      (cloop : transport C loop cbase  $\equiv$  cbase)  $\rightarrow$ 
      apd (indS {C} cbase cloop) loop  $\equiv$  cloop

```

**Figure 6.** A HIT encoded using Licata's method

type theory lens, it is often called *ap*, as it describes the action of a function on paths. In a higher inductive type, *ap* should compute new paths from old ones.

```

ap : {A B : Set} {x y : A}
  (f : A  $\rightarrow$  B) (p : x  $\equiv$  y)  $\rightarrow$  f x  $\equiv$  f y

```

In addition to describing the constructors of the points and paths of *S*, figure 6 additionally demonstrates the dependent eliminator (that is, the induction rule) *indS* and its computational meaning. The dependent eliminator relies on another operation on identifications, called *transport*, that coerces an inhabitant of a family of types at a particular index into an inhabitant at another index. Outside of homotopy type theory, *transport* is typically called *subst* or *replace*, because it also expresses that substituting equal elements for equal elements is acceptable.

```

transport : {A : Set} {x y : A}  $\rightarrow$ 
  (P : A  $\rightarrow$  Set)  $\rightarrow$  (p : x  $\equiv$  y)  $\rightarrow$  P x  $\rightarrow$  P y

```

In the postulated computation rule for *indS*, the function *apd* is the dependent version of *ap*: it expresses the action of dependent functions on paths.

```

apd : {A : Set} {B : A  $\rightarrow$  Set} {x y : A}  $\rightarrow$ 
  (f : (a : A)  $\rightarrow$  B a)  $\rightarrow$  (p : x  $\equiv$  y)  $\rightarrow$ 
  transport B p (f x)  $\equiv$  f y

```

The next section introduces the necessary automation features by describing the automatic generation of eliminators for a variant on Dybjer's inductive families. Section 4 then generalizes this feature to automate the production of eliminators for higher inductive types using Licata's technique. Section 5 revisits Angiuli et al.'s encoding of Darcs's patch theory [5] and demonstrates that the higher inductive types employed in that paper can be generated succinctly using our library.

### 3 Code Generation for Inductive Types

An inductive type *D* is a type that is freely generated by a finite collection of constructors. The constructors of *D* accept zero or more arguments, and result in an *D*. The constructors can also take an element of type *D* itself as an argument, but only *strictly positively*: any occurrences of the type constructor *D* in the type of an argument to a constructor of *D* must not be to the left of any arrows. Type constructors can have a number of *parameters*, which may not vary between the constructors, as well as *indices*, which may vary.

In Agda, constructors are given a function type. In Agda's reflection library, the constructor *data-type* of the datatype Definition stores the constructors of an inductive type as a list of Names. The type of a constructor can be retrieved by giving its Name as an input to the *getType* primitive. In this section, we discuss how to use the list of constructors and their types to generate code for the elimination rules of an inductive type.

#### 3.1 Non-dependent Eliminators

In Agda, we define an inductive type using *data* keyword. A definition of an inductive datatype declares its type and specifies its constructors. While Agda supports a variety of ways to define new datatypes, we will restrict our attention to the subset that correspond closely to Dybjer's inductive families. In general, the definition of an inductive datatype *D* with constructors  $c_1 \dots c_n$  has the following form:

```

data D (a1 : A1) ... (an : An) : (i1 : I1)  $\rightarrow$  ...  $\rightarrow$  (im : Im)  $\rightarrow$  Set where
  c1 :  $\Delta_1 \rightarrow D$  a1 ... an e11 ... e1m
      :
  cr :  $\Delta_n \rightarrow D$  a1 ... an er1 ... erm

```

where the index instantiations  $e_{k1} \dots e_{km}$  are expressions in the scope induced by the telescope  $\Delta_k$ . Every expression

```

data Vec (A : Set) : Nat → Set where
  [] : Vec A zero
  _::_ : {n : Nat} →
    (x : A) → (xs : Vec A n) →
    Vec A (suc n)

```

Figure 7. Length-indexed lists

in the definition must also be well-typed according to the provided declarations.

As an example, the datatype `Vec` represents lists of a known length. It is defined in figure 7. There is one parameter, namely  $(A : \text{Set})$ , and one index, namely  $\text{Nat}$ . The second constructor, `_::_`, has a recursive instance of `Vec` as an argument.

While inductive datatypes are essentially characterized by their constructors, it must also be possible to eliminate their inhabitants, exposing the information in the constructors. This section describes an Agda metaprogram that generates a non-dependent recursion principle for an inductive type; section 3.2 generalizes this technique to fully dependent induction principles.

For `Vec`, the recursion principle says that, in order to eliminate a `Vec A n`, one must provide a result for the empty `Vec` and a means for transforming the head and tail of a non-empty `Vec` combined with the result of recursion onto a tail into the desired answer for the entire `Vec`. Concretely, the type of the recursor `recVec` is:

```

recVec : (A : Set) →
  {n : Nat} → Vec A n →
  (C : Set) →
  (base : C) →
  (step : {n : Nat} →
    (x : A) →
    (xs : Vec A n) → C →
    C) → C

```

The recursor `recVec` maps the constructor `[]`, which takes zero arguments, to `base`. It maps  $(x :: xs)$  to  $(\text{step } x \text{ xs } (\text{recVec } xs \text{ C } \text{base } \text{step}))$ . Because `step` is applied to a recursive call to the recursor, it takes one more argument than the constructor `_::_`.

Based on the schematic presentation of inductive types  $D$  earlier in this section, we can define a schematic representation for their non-dependent eliminators  $D_{\text{rec}}$ .

```

Drec : (a1 : A1) → ... → (an : An) →
  (i1 : I1) → ... → (im : Im) →
  (tgt : D a1 ... an i1 ... in) →
  (C : Set) →
  (f1 : Δ'1 → C) → ... → (fr : Δ'r → C) →
  C

```

```

pattern _[_v]⇒_ a s b = pi (vArg a) (abs s b)
pattern _[_h]⇒_ a s b = pi (hArg a) (abs s b)

```

```

(agda-sort (lit 0) [ "A" h]⇒)      -- A
(def (quote Nat) [] [ "n" h]⇒)    -- n
(var 1 [] [ "x" v]⇒)              -- x
(def (quote Vec)                   -- xs : Vec A n
  (vArg (var 2 []) ::
    vArg (var 1 []) :: [])
  [ "xs" v]⇒)
(def (quote Vec)                   -- Vec A (suc n)
  (vArg (var 3 []) ::
    vArg (con (quote suc)
      (vArg (var 2 []) :: []))
  :: []))

```

Figure 8. Abstract syntax tree for the type of `_::_`

The type of  $f_i$ , which is the method for fulfilling the desired type  $C$  when eliminating the constructor  $c_i$ , is determined by the type of  $c_i$ . The telescope  $\Delta'_i$  is the same as  $\Delta_i$  for non-recursive constructor arguments. However,  $\Delta'_i$  binds additional variables when there are recursive occurrences of  $D$  in the arguments. If  $\Delta_i$  has an argument  $(y : B)$ , where  $B$  is not an application of  $D$  or a function returning such an application,  $\Delta'_i$  binds  $(y : B)$  directly. If  $B$  is an application of  $D$ , then an additional binding  $(y' : C)$  is inserted following  $y$ . Finally, if  $B$  is a function type  $\Psi \rightarrow D$ , the additional binding is  $(y' : \Psi \rightarrow C)$ .

To construct the type of `recVec`, we need to build the types of `base` and `step`. These are derived from the corresponding types of `base` and `_::_`, which can be discovered using reflection primitives. Since `[]` requires no arguments, its corresponding method is  $(\text{base} : C)$ . The constructor `pi` of type `Term` encodes the abstract syntax tree (AST) representation of `_::_` (figure 8). We can retrieve and traverse the AST of `_::_`, and add new type information into it to build a new type representing `step`.

During the traversal of abstract syntax tree of the type of `_::_`, when the type `Vec` occurs directly as an argument, the result type  $C$  is added next to it. For example, in figure 8, a new function is built from the argument  $(xs : \text{Vec } A \text{ n})$  by modifying it to  $(\text{Vec } A \text{ n}) \rightarrow C$  (figure 9). Arguments other than `Vec` require no modifications. Therefore,  $(x : A)$  is copied into the new type without any changes. Finally, the codomain `Vec A (suc n)` of `_::_`'s type is replaced with  $C$ , resulting in an AST for the type of `step`.

In general, when automating the production of  $D_{\text{rec}}$ , all the information that is needed to produce the type signature is available in the TC monad by looking up  $D$ 's definition. The constructor data-type contains the number of parameters occurring in a defined type. It also encodes the constructors of the type as a list of Names. Metaprograms can retrieve the index count by finding the difference between the number of parameters and the length of the constructor list. The constructors of  $D$  refer to the parameter and the index using de Bruijn indices.

```

(agda-sort (lit 0) [ "A" h]⇒          -- A
  (def (quote Nat) [] [ "n" h]⇒      -- n
    (def (quote Vec) (vArg (var 1 []) :: -- Vec A n
      vArg (var 0 []) :: [] [ "_" v]⇒
    (agda-sort (lit 0) [ "C" v]⇒      -- C
      (var 0 [] [ "_" v]⇒            -- base
        ((def (quote Nat) [] [ "n" h]⇒ -- step
          (var 5 [] [ "x" v]⇒         -- x
            (def (quote Vec)
              (vArg (var 6 []) ::
                vArg (var 1 []) ::
                []))
              [ "xs" v]⇒              -- xs
              (var 4 [] [ "_" v]⇒     -- → C
                var 5 []])))          -- C
          [ "_" v]⇒ var 2 []])))      -- C

```

Figure 9. Abstract syntax tree of recVec's type

The method step for the constructor  $_{::}$  in Vec, refers to the parameter and the index using de Bruijn indices. During the construction of the type of step, the recursor generator updates the de Bruijn indices accordingly. Note that not all indices occur as arguments to a constructor: in Vec, the constructor  $[]$  instantiates the index with a constant.

Once the AST for step's type has been found, it is possible to build the type of recVec in figure 9. To quantify over the return type ( $C : \text{Set}$ ), the Term constructor agda-sort refers to Set).

The general schema for the computation rules corresponding to  $D_{rec}$  and constructors  $c_1, \dots, c_n$  follows:

$$\begin{aligned}
 D_{rec} a_1 \dots a_n i_1 \dots i_m (c_1 \Delta_1) C f_1 \dots f_r &= \text{RHS}(f_1, \Delta'_1) \\
 \vdots \\
 D_{rec} a_1 \dots a_n i_1 \dots i_m (c_r \Delta_r) C f_1 \dots f_r &= \text{RHS}(f_r, \Delta'_r)
 \end{aligned}$$

Here,  $\overline{\Delta_j}$  is the sequence of variables bound in  $\Delta_j$ . RHS constructs the application of the method  $f_j$  to the arguments of  $c_j$ , such that  $C$  is satisfied. It is defined by recursion on  $\Delta_j$ .  $\text{RHS}(f_j, \cdot)$  is  $f_j$ , because all arguments have been accounted for.  $\text{RHS}(f_j, (y : B)\Delta_k)$  is  $\text{RHS}(f_j y, \Delta_k)$  when  $B$  does not mention  $D$ .  $\text{RHS}(f_j, (y : D)(y' : C)\Delta_k)$  is  $\text{RHS}(f_j y (D_{rec} \dots y \dots), \Delta_k)$ , where the recursive use of  $D_{rec}$  is applied to the recursive constructor argument as well as the appropriate indices, and the parameters, result type, and methods remain constant. Higher-order recursive arguments are a generalization of first-order arguments. Finally,

$$\text{RHS}(f_j, (y : \Psi \rightarrow D)(y' : \Psi \rightarrow C)\Delta_k)$$

is

$$\text{RHS}\left(f_j y \left(\lambda \overline{\Psi}. D_{rec} \dots (y \overline{\Psi}) \dots\right), \Delta_k\right)$$

where the recursive use of  $D_{rec}$  is as before.

After declaring recVec's type using declareDef, it is time to define its computational meaning. The computation rule

```

(cclause
  (vArg (con (quote _::_)          -- _::_
    (vArg (var "x") ::            -- x
      vArg (var "xs") :: []))     -- xs
    vArg (var "C") ::             -- C
    vArg (var "base") ::          -- base
    vArg (var "step") :: []))     -- step
  (var 0
    (vArg (var 4 []) ::           -- x
      vArg (var 3 []) ::         -- xs
      vArg
        (def recVec              -- recursion
          (vArg (var 3 [])        -- xs
            vArg (var 2 []) ::    -- C
            vArg (var 1 []) ::    -- base
            vArg (var 0 []) :: []]))) -- step

```

Figure 10. Clause definition for the computation rule of  $_{::}$ 

```

data W (A : Set) (B : A → Set) : Set where
  sup : (a : A) → (B a → W A B) → W A B

```

Figure 11. W-Type

representing the action of function recVec on  $[]$  and  $_{::}$  using clause (figure 10). The first argument to clause encodes variables corresponding to the above type, and it also includes the abstract representation of the constructors  $[]$  and  $_{::}$  on which the pattern matching should occur. The second argument to clause, which is of type Term, refers to the variables in the first argument using de Bruijn indices, and it encodes the output of recVec when the pattern matches. The constructor var of Pattern is used to introduce new variables in the clause definition. The type Pattern also has another constructor con that represents patterns that match specific constructors. The type Term has similar constructors var and con, that respectively represent variable references and constructor invocations. The computation rules for recVec are

```

recVec []          C base step =
  base
recVec (x :: xs) C base step =
  step x xs (f xs C base step)

```

Figure 10 presents the AST for the clause that matches  $_{::}$ . The de Bruijn index reference increments right-to-left, starting from the last argument. Definitions by pattern matching are added to the global context using the defineFun primitive.

Figure 11 presents an Agda encoding of Martin-Löf's well-orderings, the so-called *W-type*. W is interesting because its constructor exhibits a *higher-order* recursive instance of the type being defined. Following the recipe yields the recursor in figure 12, in which the recursive call occurs under a function representing arbitrary choices of tag.

The type of step is built by traversing the AST of sup's type. The first argument to sup, which is a constant type A,

```

recW : ∀ {A B}
  (tgt : W A B) →
  (C : Set) →
  (step : (x : A) →
    (f : B x → W A B) →
    (f' : B x → C) →
    C) →
  C
recW (sup x f) C step =
  step x f (λ b → recW (f b) C step)

```

Figure 12. The non-dependent eliminator for W

```

generateRec : Arg Name → Name →
  (indexList : List Nat) → TC T
generateRec (arg i f) t indLs =
  do indLs' ← getIndex t indLs
  cns ← getConstructors t
  lcons ← getLength cns
  cls ← getClause lcons zero t f indLs cns
  RTy ← getRtype t
  funType ← getRtype t indLs' zero RTy
  declareDef (arg i f) funType
  defineFun f cls

```

Figure 13. Implementation for generateRec

is copied directly into step's type. The second argument is a  $(B \times \rightarrow W A B)$ , which is a function whose codomain is a recursive instance of W. The resulting arguments must account for the recursion and are thus  $(B \times \rightarrow W A B)$  and  $(B \times \rightarrow C)$ . Finally, the codomain  $W A B$  of sup is replaced by C.

In the above computation rule, the third argument to step is a function that works for any choice of tag b. The arguments to lam are referenced using de Bruijn indices inside the lambda body. Thus, the de Bruijn indices for referring variables outside the lambda body must be updated accordingly.

Figure 13 demonstrates the implementation of generateRec, which constructs recursors. generateRec uses getClause and getRtype to build the computation and elimination rules respectively. It takes three arguments: the name of the function to be defined (represented by an element of type Arg Name), the quoted Name of the datatype, and a list containing the index count of the individual constructors. generateRec can be used to automate the generation of recursion rules for inductive types having the general schema given at the beginning of this section. The recursion rule generated by generateRec is brought into scope using unquoteDecl as follows.

```

unquoteDecl f = generateRec (vArg f)
  (quote Vec)
  (0 :: 1 :: [])

```

The third argument to generateRec is a list consisting of the index count for the constructors. It is required to pass

```

indW : {A : Set} → {B : A → Set} →
  (tgt : W A B) →
  (mot : W A B → Set) →
  (step : (x : A) →
    (f : B x → W A B) →
    (f' : (b : B x) → mot (f b)) →
    mot (sup x f)) →
  mot tgt
indW (sup x f) mot step =
  step x f (λ b → indW (f b) mot step)

```

Figure 14. The induction principle for W

the index count for each constructor explicitly as the Agda reflection library does not have built-in primitives to retrieve the index value.

### 3.2 Dependent Eliminators

The dependent eliminator for a datatype, also known as the *induction principle*, is used to eliminate elements of a datatype when the type resulting from the elimination mentions the very element being eliminated. The type of the induction principle for D is:

$$\begin{aligned}
 D_{ind} : & (a_1 : A_1) \rightarrow \dots \rightarrow (a_n : A_n) \rightarrow \\
 & (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
 & (tgt : D a_1 \dots a_n i_1 \dots i_m) \rightarrow \\
 & (C : (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
 & \quad D a_1 \dots a_n i_1 \dots i_n \rightarrow \\
 & \quad \text{Set}) \rightarrow \\
 & (f_l : \Delta'_l \rightarrow C e_{l1} \dots e_{lp} (c_l \overline{\Delta_l})) \rightarrow \\
 & (f_r : \Delta'_r \rightarrow C e_{r1} \dots e_{rp} (c_r \overline{\Delta_r})) \rightarrow \\
 & C i_1 \dots i_n tgt
 \end{aligned}$$

Unlike the non-dependent recursion principle  $D_{rec}$ , the result type is now computed from the target and its indices. Because it expresses the reason that the target must be eliminated, the function C is often referred to as the *motive*. Similarly to  $D_{rec}$ , the type of each method  $f_i$  is derived from the type of the constructor  $c_i$ —the method argument telescope  $\Delta'_k$  is similar, except the arguments that represent the result of recursion now apply the motive C to appropriate arguments. If  $\Delta_i$  has an argument  $(y : B)$ , where B is not an application of D or a function returning such an application,  $\Delta'_i$  still binds  $(y : B)$  directly. If B is an application of D to parameters  $a \dots$  and indices  $e \dots$ , then an additional binding  $(y' : C e \dots y)$  is inserted following y. Finally, if B is a function type  $\Psi \rightarrow D a \dots e \dots$ , the additional binding is  $(y' : \Psi \rightarrow C e \dots (y \overline{\Psi}))$ .

The computation rules for the induction principle are the same as for the recursion principle. Following these rules, the induction principle for W can be seen in figure 14.

```

(agda-sort (lit 0) [ "A" h]⇒          -- A
((var 0 [] [ "_" v]⇒                -- B : A → Set
  agda-sort (lit 0)) [ "B" h]⇒
  (def (quote W)
    (vArg (var 1 []) ::              -- tgt : W A B
      vArg (var 0 []) :: []) [ "tgt" v]⇒
    ((def (quote W)
      (vArg (var 2 []) ::            -- mot : W A B → Set
        vArg (var 1 []) :: []) [ "_" v]⇒
        agda-sort (lit 0)) [ "mot" v]⇒
        ((var 3 [] [ "x" v]⇒        -- x : A
          ((var 3
            (vArg (var 0 []) :: [])   -- f : B x → W A B
            [ "_" v]⇒
            def (quote W)
              (vArg (var 5 []) ::
                vArg (var 4 []) :: []) [ "f" v]⇒
              ((var 4 (vArg (var 1 []) :: -- f' : B x → mot (f x)
                [])
                [ "f'" v]⇒
                var 3
                (vArg (var 1 (vArg (var 0 []) ::
                  [])) ::
                  []))
                [ "z" v]⇒
                var 3                  -- mot (sup x f)
                (vArg
                  (con (quote sup)
                    (vArg (var 2 []) ::
                      vArg (var 1 []) :: [])
                    :: [])))))
                [ "_" v]⇒ var 1 (vArg (var 2 []) :: -- mot tgt
                  [])))))

```

**Figure 15.** Abstract syntax tree for the dependent eliminator of  $W$

Automating the production of the dependent eliminator is an extension of the procedure for automating the production of the non-dependent eliminator.

We can construct the AST of  $d$  using the static type information obtained from  $\text{sup}$ . To construct  $\text{indW}$ , during the traversal of the AST of  $\text{sup}$ 's type, the argument  $(a : A)$  is copied without any changes, just as it is in the case of the non-dependent eliminator. The next argument, however, is a function that returns a  $W$ . An additional argument, representing the induction hypothesis, is needed. The induction hypothesis  $(f' : (b : B x) \rightarrow \text{mot } (f b))$  takes the same arguments as  $f$ , but it returns the motive instantiated at the application of  $f$ . The final return type is found by applying the motive to the target and its indices (figure 15).

We can construct the type of the induction principle  $f$  using  $d$ . The type  $C$  in the mapping  $f$  depends on the element of the input type  $W A B$ . Operationally, the induction principle computes just like the recursion principle. It is constructed using clause definitions following the same approach. The generation of induction principles is carried out using  $\text{generateInd}$ , in figure 16.

$\text{generateInd}$  uses  $\text{getClauseDep}$  to generate the clause definitions representing the computation rules. The abstract representation of the type is provided by  $\text{getRtypeInd}$ . A version of the induction principle called  $\text{indW}'$  generated

```

generateInd : Arg Name → Name →
  (indexList : List Nat) → TC T
generateInd (arg i f) t indLs =
  do id' ← getIndex t indLs
  cns ← getConstructors t
  lcns ← getLength cns
  cls ← getClauseDep lcns zero t f id' cns
  RTy ← getType t
  funType ← getRtypeInd t zero id' RTy
  declareDef (arg i f) funType
  defineFun f cls

```

**Figure 16.** Implementation for  $\text{generateInd}$

```

data ArgPath {ℓ1} : Set (lsuc ℓ1) where
  argPath : Set ℓ1 → ArgPath

```

**Figure 17.** Definition of  $\text{ArgPath}$

by  $\text{generateInd}$  is brought into scope by  $\text{unquoteDecl}$  as follows:

```

unquoteDecl indW' = generateInd (vArg f)
  (quote W) []

```

## 4 Code Generation for Higher Inductive Types

In Agda, there are no built-in primitives to support the definition of higher inductive types. However, we can still define a higher inductive type with a base type using Licata's [17] method, as described in section 2.1. In this section, we discuss the automation of code generation for the boiler-plate code segments defining the higher inductive types.

### 4.1 Defining Higher Inductive Types

Our library defines a higher inductive type  $G$  as a top-level definition using a base type  $D$ , similar to the module `Circle` in section 2.1. The first step is to generate the private base type. After this, public definitions are created to provide access to the constructors, and the paths are postulated. Finally, the elimination principles are generated, and their actions on paths are postulated.

The first step is to copy the type former  $D$ 's type signature, re-using it for the declaration of  $G$ . Next, the library copies the type of each constructor  $c_i$  to a definition  $g_i$ , except each reference to  $D$  is replaced by a reference to  $G$ .

Unlike the type former and the point constructors, which can be copied from the base type, the paths in  $G$  must be provided explicitly. The code generator takes the path constructor types as input and postulates them as identifications using the reflection primitive `declarePostulate`. The data type  $\text{ArgPath}$  in figure 17 represents paths as their types.

HITs are defined by invoking the `data-hit` metaprogram with a syntax reminiscent of ordinary data declarations and the putative HIT syntax from figure 5.



```

data-hit :  $\forall \{\ell_1\}$ 
  (baseType : Name)  $\rightarrow$  (indType : Name)  $\rightarrow$ 
  (pointHolder : Name)  $\rightarrow$  (lcons : List Name)  $\rightarrow$ 
  (pathHolder : Name)  $\rightarrow$  (lpaths : List Name)  $\rightarrow$ 
  (lpathTypes : (List (ArgPath  $\{\ell_1\}$ )))  $\rightarrow$  TC  $\top$ 
data-hit base ind h1 lcons h2 lpaths pTy =
  do defineHindType base ind
     cns  $\leftarrow$  getConstructors base
     defineHitCons base ind cns lcons
     pTy'  $\leftarrow$  getPathTypes base ind cns lcons pTy
     defineHitPathCons lpaths pTy'
     definePointHolder h1 lcons
     definePathHolder h2 lpaths

```

**Figure 18.** The implementation of data-hit

```

data-hit (quote D) G
  Gpoints ( $g_1 :: \dots :: g_r :: []$ )
  Gpaths ( $p_1 :: \dots :: p_s :: []$ )
  (argPath
    ( $\{a_1 : A_1\} \rightarrow \dots \rightarrow \{a_n : A_n\} \rightarrow$ 
       $\Delta_1 \rightarrow$ 
       $(c_i e \dots) \equiv (c_j e \dots) ::$ 
       $\vdots$ 
    )
  argPath
    ( $\{x_1 : P_1\} \rightarrow \dots \rightarrow \{x_n : P_n\} \rightarrow$ 
       $\Delta_s \rightarrow$ 
       $(c_i e \dots) \equiv (c_j e \dots) :: []$ 
    )

```

data-hit defines top-level definitions Gpoints for point constructors and Gpaths for path constructors as part of the definition of the encoded higher inductive type  $G$ . This is because Agda does not natively support these definitions, so there is no way to retrieve the constructors of the  $G$  using the ordinary reflection mechanisms, so an additional registry is needed. This parallels the use of compile-time bindings in Racket [15] to store structure type metadata.

The elements of the argPath list represent the type of the path constructors  $p_1 \dots p_q$ . The constructor  $g_i$  is not in scope when used in the path type passed to argPath, so the base types' constructors are used instead. Figure 18 contains the implementation of data-hit.

In the implementation of data-hit in figure 18, the helper defineHindType defines the higher inductive type as a top-level definition using the base type, and defineHitCons defines specifies the point constructors of the higher inductive type as references to the base type's constructors. defineHitPathCons builds the path constructors of the higher

inductive type using the argPath list. For example, Circle from section 2.1 is defined thusly:

```

unquoteDecl S Spoints base Spaths loop =
  data-hit (quote S*) S
    Spoints (base :: []) -- point constructors
    Spaths (loop :: []) -- path constructors
    (argPath (base*  $\equiv$  base*) :: [])
    -- base replaces base*

```

The identity type input ( $base* \equiv base*$ ) to argPath represents the type of the path loop. The constructor base comes into scope only during the execution of unquoteDecl, and so cannot be used in the identity type reference in argPath. The references to base\* in the type of the path loop are replaced by base once it is defined.

## 4.2 Non-dependent Elimimators for HITs

The the recursion principle of a higher inductive type  $G$  maps the points and paths of  $G$  to points and paths in an output type  $C$ . We extend the general schema of the recursion principle given in section 3.1 by adding methods for path constructors as follows.

$$\begin{aligned}
 G_{rec} : & (a_1 : A_1) \rightarrow \dots \rightarrow (a_n : A_n) \rightarrow \\
 & (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
 & (tgt : G \ a_1 \dots a_n \ i_1 \dots i_n) \rightarrow \\
 & (C : \text{Set}) \rightarrow \\
 & (f_1 : \Delta'_1 \rightarrow C) \dots (f_r : \Delta'_r \rightarrow C) \rightarrow \\
 & (k_1 : \Delta'_1 \rightarrow (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 & \vdots \\
 & (k_q : \Delta'_q \rightarrow (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 & C
 \end{aligned}$$

The schematic definition of  $G_{rec}$  supports only we have given only one-dimensional paths. Our metaprogram currently supports only one-dimensional paths, but we are planning to improve the tool to support higher-dimensional paths in the future.

The type of the method  $f_i$  for the point constructor  $g_i$  in  $G_{rec}$  is built the same way as for the normal inductive type  $D$ , as described in section 3.1. The code generator builds the type of  $k_i$ , method for path constructor  $p_i$  in  $G_{rec}$ , by traversing the AST of  $p_i$ . The arguments of  $k_i$  are handled the same way as for the point constructor's method  $f_i$ . During the traversal, the code generator uses the base type recursor  $D_{rec}$  to map the point constructors  $g_i$  of  $G$  in the codomain of  $p_i$  to  $f_i$ . Determining the computation rules corresponding to points  $g_i$  is similar to the computation rules corresponding to constructors  $c_i$  of the inductive type  $D$ , except that there are additional methods to handle paths. Paths compute new paths; the computation rules that govern the interaction of recursors and paths  $p_i$  are named and postulated. They identify the action of the recursor on the path with the

```

(def (quote S) [] [ "_" v]⇒ -- S
  (agda-sort (lit 0) [ "C" v]⇒ -- C : Set
    (var 0 [] [ "cbase" v]⇒ -- cbase
      (def (quote _≡_) -- cloop
        (vArg (var 0 []) ::
          vArg (var 0 []) :: []))
      [ "cloop" v]⇒ var 2 []))))

```

**Figure 19.** Abstract syntax tree for `recS`'s type

corresponding method.

$$\begin{aligned}
 \beta G_{rec} : (a_1 : A_1) \rightarrow \dots \rightarrow (a_n : A_n) \rightarrow \\
 (C : \text{Set}) \rightarrow \\
 (f_1 : \Delta'_1 \rightarrow C) \dots (f_r : \Delta'_r \rightarrow C) \rightarrow \\
 (k_1 : \Delta'_1 \rightarrow (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 \vdots \\
 (k_q : \Delta'_q \rightarrow (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 \text{ap } (\lambda x. G_{rec} \ x \ C \ f_1 \dots f_r \ k_1 \dots k_q) \\
 (p_i \dots) \equiv (k_i \dots)
 \end{aligned}$$

As an example, if code for the circle HIT from section 2.1 has been generated, and the type is called `S`, then the recursor needs a method for base and one for loop. The method for base should be an inhabitant of `C`. If it is called `cbase`, then the method for loop should be a path `cbase ≡ cbase`. The types of the path methods depend on the values of the point methods. The code generator builds the type of loop's method by traversing the AST of loop's type, replacing references to point constructors with the result of applying the base type's recursor to the point methods.

The recursion rule `recS`, corresponding to figure 19, follows this pattern.

```

recS : S →
  (C : Set) →
  (cbase : C) →
  (cloop : cbase ≡ cbase) →
  C

```

The code generator builds the computation rule for the point constructor base using the same approach as described in section 3.1, as if it were for the base type. Additionally, it includes variables in the `clause` definition for the path constructor loop. The code generator postulates the following computation rule  $\beta_{recS}$  for the path constructor loop:

$$\begin{aligned}
 \beta_{recS} : (C : \text{Set}) \rightarrow (cbase : C) \rightarrow \\
 (cloop : cbase \equiv cbase) \rightarrow \\
 \text{ap } (\lambda x \rightarrow \text{recS } x \ C \ \text{cbase } \text{cloop}) \ \text{loop} \\
 \equiv \text{cloop}
 \end{aligned}$$

The application of function `recS` to the path loop substitutes the point base for the argument `x`, and it evaluates to the path `cloop` in the output type `C`. The definition of `generateRecHit` can be found in figure 20.

```

generateRecHit :
  Arg Name → List (Arg Name) →
  (baseType : Name) → (indexList : List Nat) →
  (baseRec : Name) → (indType : Name) →
  (points : List Name) →
  (paths : List Name) → TC ⊤
generateRecHit (arg i f) argD b il br i p1 p2 =
  do lcons ← getConstructors b
  lpoints ← getLength p1
  lpaths ← getLength p2
  clauses ← getPathClause lpoints lpaths br
  RTy ← getType baseType
  fTy ← getRtypePath b i br il p2 zero RTy
  declareDef (arg i f) fTy
  defineFun f clauses
  generateβRecHit argD b il br i f p1 p2

```

**Figure 20.** Implementation for `generateRecHit`

`generateRecHit` takes the base types' recursion rule as input and uses that to eliminate the points during the construction of the path methods in the recursor  $G_{rec}$ . The second argument `argD` is a list of terms representing the computation rules for the path constructors. The `generateβRecHit` interface takes `argD` as input and builds the computation rules for the path constructors. Other inputs to `generateRecHit` are the point and path collections declared during the definition of the HIT.

### 4.3 Dependent Elimimators for HITs

The dependent eliminator for a higher inductive type  $G$  is a dependent function that maps an element  $g$  of  $G$  to an output type  $Cg$ . The general schema for the induction principle of  $G$  is given as follows.

$$\begin{aligned}
 G_{ind} : (a_1 : A_1) \rightarrow \dots \rightarrow (a_n : A_n) \rightarrow \\
 (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
 (tgt : G \ a_1 \dots a_n \ i_1 \dots i_n) \rightarrow \\
 (C : (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
 G \ a_1 \dots a_n \ i_1 \dots i_n \rightarrow \\
 \text{Set}) \rightarrow \\
 (f_1 : \Delta'_1 \rightarrow C \ j_{11} \dots j_{1p} \ (c_1 \ \overline{\Delta_1})) \rightarrow \\
 \vdots \\
 (f_r : \Delta'_r \rightarrow C \ j_{r1} \dots j_{rp} \ (c_r \ \overline{\Delta_r})) \rightarrow \\
 (k_1 : \Delta'_1 \rightarrow \text{transport } C \ p_1 \ (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 \vdots \\
 (k_q : \Delta'_q \rightarrow \text{transport } C \ p_q \ (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
 C \ i_1 \dots i_n \ \text{tgt}
 \end{aligned}$$

Similar to  $G_{rec}$ , the type of  $f_i$  is built the same way as for the normal inductive type  $D$ . The code generator builds the

type of the method for path constructor  $p_i$ , called  $k_i$ , in  $G_{ind}$ , by traversing the AST of  $p_i$ . During the traversal, the code generator uses the base eliminator  $D_{ind}$  to map the point constructors  $g_i$  of  $G$  in the codomain of  $p_i$  to  $f_i$ . In the first argument to the identity type in the codomain of  $k_i$ , the code generator adds an application of transport to the motive  $C$  and the path  $p_i$ . The arguments of  $k_i$  are handled the same way as for  $f_i$ . The computation rules corresponding to paths  $p_i$  are postulated as follows:

$$\begin{aligned}
& \beta G_i : (a_1 : A_1) \rightarrow \dots \rightarrow (a_n : A_n) \rightarrow \\
& \quad (C : (i_1 : I_1) \rightarrow \dots \rightarrow (i_m : I_m) \rightarrow \\
& \quad \quad G a_1 \dots a_n i_1 \dots i_n \rightarrow \\
& \quad \quad \text{Set}) \rightarrow \\
& \quad (f_i : \Delta'_1 \rightarrow C j_{i1} \dots j_{ip} (c_1 \overline{\Delta_1})) \rightarrow \\
& \quad (f_r : \Delta'_r \rightarrow C j_{r1} \dots j_{rp} (c_r \overline{\Delta_r})) \rightarrow \\
& \quad (k_l : \Delta'_1 \rightarrow \text{transport } C p_l (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
& \quad (k_r : \Delta'_r \rightarrow \text{transport } C p_r (f_i \dots) \equiv (f_j \dots)) \rightarrow \\
& \quad \text{apd } (\lambda x . G_{ind} x C f_1 \dots f_r k_1 \dots k_r) \\
& \quad (p_i \dots) \equiv \\
& \quad (k_i \dots)
\end{aligned}$$

For the type  $S$  with point constructor base and path constructor loop, to define a mapping  $\text{indS} : (x : S) \rightarrow C x$ , we need  $\text{cbase} : C \text{ base}$  and  $\text{cloop} : \text{transport } C \text{ loop cbase} \equiv \text{cbase}$ , where  $\text{cloop}$  is a heterogeneous path transported over  $\text{loop}$ . The code generator builds the type of  $\text{cloop}$  by adding relevant type information to the type of  $\text{loop}$ . The type of the method for the path constructor  $\text{cloop}$  is derived by inserting a call to  $\text{transport}$  with arguments  $C$ ,  $\text{loop}$  and  $\text{cbase}$ . The code generator applies the eliminator of base type  $S^*$  to map  $\text{base}$  to  $\text{cbase}$  during the construction of the codomain of  $\text{cloop}$ . The following declaration gives the type of  $\text{indS}$ .

$$\begin{aligned}
& \text{indS} : (\text{circle} : S) \rightarrow \\
& \quad (C : S \rightarrow \text{Set}) \rightarrow \\
& \quad (\text{cbase} : C \text{ base}) \rightarrow \\
& \quad (\text{cloop} : \text{transport } C \text{ loop cbase} \equiv \text{cbase}) \rightarrow \\
& \quad C \text{ circle}
\end{aligned}$$

The computation rule for base, which defines the action of  $\text{indS}$  on  $\text{base}$ , is built using the same approach as for the non-dependent eliminator  $\text{recS}$ . The postulated computation rule  $\beta \text{indS}$  for the path  $\text{loop}$  uses  $\text{apd}$  which gives the action of dependent function  $\text{indS}$  on the path  $\text{loop}$ .

$$\begin{aligned}
& \beta \text{indS} : (C : S \rightarrow \text{Set}) \rightarrow \\
& \quad (\text{cbase} : C \text{ base}) \rightarrow \\
& \quad (\text{cloop} : \text{transport } C \text{ loop cbase} \equiv \text{cbase}) \rightarrow \\
& \quad \text{apd } (\lambda x \rightarrow \text{indS } x C \text{ cbase } \text{cloop}) \text{ loop} \equiv \text{cloop}
\end{aligned}$$

Figure 21 gives the implementation of  $\text{generateIndHit}$ . The helper  $\text{generate}\beta\text{IndHit}$  builds the computation rules for the path constructors.

```

generateIndHit : Arg Name → List (Arg Name) →
  (baseType : Name) → (indLs : List Nat) →
  (baseElm : Name) → (indType : Name) →
  (points : List Name) →
  (paths : List Name) → TC T
generateIndHit (arg i f) argD b il br i p1 p2 =
  do il' ← getIndex b il
  lcons ← getConstructors b
  lp1 ← getLength p1
  lp2 ← getLength p2
  clauses ← getPathClauseDep lp1 lp2
           b br il' lcons
  RTy ← getType b
  fTy ← getRTypPathDep b i br
        p1 p2 zero il' RTy
  declareDef (arg i f) fTy
  defineFun f clauses
  generateβIndHit argD b il br i f p1 p2

```

**Figure 21.** Implementation for  $\text{generateIndHit}$

## 5 Applications

Homotopy type theory has thus far primarily been applied to the encoding of mathematics, rather than to programming. Nevertheless, there are a few applications of homotopy type theory to programming. Applications such as homotopical patch theory [5] discuss a model of the core of the of the Darcs [23] version control system using patch theory [12, 18] encoded as a HIT. Containers in homotopy type theory [2, 3] implement data structures such as multisets and cycles. Automating the HIT boilerplate code allows more programmers to begin experimenting with programming with HITs.

### 5.1 Patch Theory Revisited

A patch is a syntactic representation of a function that modifies a repository when applied. For example, a patch  $(s_1 \leftrightarrow s_2 @ l)$ , which replaces string  $s_1$  with  $s_2$  at line  $l$ , when applied to a repository containing with string  $s_1$  at line  $l$  results in a repository with string  $s_2$  at line  $l$ . In homotopical patch theory [5], the patches are modeled as paths in a higher inductive type. Because patches are paths, they automatically satisfy groupoid laws, such as that the composition of patches is associative, that all patches have an inverse, and that inverse patches compose to the identity. Domain-specific laws related to the patches, such as that two swaps at independent lines commute, are implemented as higher dimensional paths. The computational content of the patches is extracted by mapping them to bijections in the universe with the help of univalence. Due to the functoriality of mappings in type theory, the functions preserve the path structures and thus the desired axioms.

We reimplemented Angiuli et al.'s patch theory in Agda. We implemented basic patches such as the insertion of a string as line  $l_1$  in a file and deletion of a line  $l_2$  from a file.

The functions implementing insertion and deletion in the universe are not bijective. So, to map the paths representing the patches insert and delete into the universe, we used Angiuli et al.'s patch history approach. According to this approach, we developed a separate higher inductive type History which serves as the types of patches. In addition to basic patches, we also implemented patches involving encryption or decryption with cryptosystems like RSA [22] and Paillier [20].

Having implemented patch theory, we then reimplemented History and the encrypted repository type cryptR it using our code generator. We also automated the code generation for the elimination and the computation rules for the higher inductive types History and cryptR. In addition to easing the implementation difficulties of higher inductive types, the code generator greatly reduced the code size. The type definitions shrank from around 1500 to around 70 lines, resulting in a 60% decrease in the overall number of lines of code in the development.

## 5.2 Cryptographic Protocols

Vivekanandan [28] models certain cryptographic protocols using homotopy type theory, introducing a new approach to formally specifying cryptographic schemes using types. The work discusses modeling cryptDB [21] using a framework similar to Angiuli et al.'s patch theory. CryptDB employs layered encryption techniques and homomorphic encryption. We can implement cryptDB by modeling the database queries as paths in a higher inductive type and mapping the paths to the universe using singleton types [5]. The code generator can be applied to generate code for the higher inductive type representing cryptDB and its corresponding elimination and computation rules. By using the code generator, we can decrease the length and increase the readability of the definitions, hopefully making it more accessible to the broad cryptographic community.

## 6 Related Work

Kokke and Swierstra [16] implemented a library for proof search using Agda's old reflection primitives, from before it had elaborator reflection. They describe a Prolog interpreter in the style of Stutterheim et al. [24]. It employs a hint database and a customizable depth-first traversal, with lemmas to assist in the proof search.

van der Walt and Swierstra [27] and van der Walt [26] discuss automating specific categories of proofs using proof by reflection. A key component of this proof technique is a means for converting an expression into a quoted representation. They automate this process, giving a user-defined datatype. van der Walt gives an overview of Agda's old metaprogramming tools.

Ongoing work on cubical type theories [4, 10, 11] provides a computational interpretation of univalence and HITs. We strenuously hope that these systems quickly reach maturity,

rendering our code generator obsolete. In the meantime, however, these systems are not yet as mature as Agda.

## 7 Conclusion and Future Work

We presented a code generator that generates the encodings of higher inductive types, developed using Agda's new support for Idris-style elaborator reflection. In particular, the tool generates formation, introduction, path, and elimination rules for 1-dimensional higher inductive types. This syntax is greatly simplified with respect to writing the encoding by hand. We demonstrated an extensive reduction in code size by employing our tool. Next, we intend to extend the tool to support higher-dimensional paths in the definition of HITs, bringing its benefits to a wider class of problems.

## References

- [1] 2017. *Agda's Documentation*. <http://agda.readthedocs.io/en/latest/language/reflection.html>.
- [2] Michael Abbott, Thorsten Altenkirch, and Neil Ghani. 2005. Containers: constructing strictly positive types. *Theoretic Computer Science*.
- [3] Thorsten Altenkirch. 2014. Containers in homotopy type theory. (January 2014). Talk at Mathematical Structures of Computation, Lyon.
- [4] Carlo Angiuli, Robert Harper, and Todd Wilson. 2017. Computational Higher-dimensional Type Theory. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL 2017)*. ACM, New York, NY, USA, 680–693.
- [5] Carlo Angiuli, Edward Morehouse, Daniel R. Licata, and Robert Harper. 2014. Homotopical Patch Theory. In *Proceedings of the 19th ACM SIGPLAN International Conference on Functional Programming (ICFP '14)*. ACM.
- [6] David Christiansen. 2016. *Practical Reflection and Metaprogramming for Dependent Types*. Ph.D. Dissertation. IT University of Copenhagen.
- [7] David Christiansen and Edwin Brady. 2016. Elaborator Reflection: Extending Idris in Idris. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming (ICFP '16)*. ACM, Nara, Japan.
- [8] Jesper Cockx and Andreas Abel. 2016. Sprinkles of Extensionality for Your Vanilla Type Theory. In *22nd International Conference on Types for Proofs and Programs (TYPES 2016)*.
- [9] Jesper Cockx, Dominique Devriese, and Frank Piessens. 2014. Pattern Matching Without K. In *Proceedings of the 19th ACM SIGPLAN International Conference on Functional Programming (ICFP '14)*. ACM, New York, NY, USA, 257–268.
- [10] Cyril Cohen, Thierry Coquand, Simon Huber, and Anders Mörtberg. 2015. Cubical Type Theory: a constructive interpretation of the univalence axiom. In *21st International Conference on Types for Proofs and Programs (21st International Conference on Types for Proofs and Programs)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Tallinn, Estonia, 262. <https://hal.inria.fr/hal-01378906>
- [11] Thierry Coquand, Simon Huber, and Anders Mörtberg. 2018. On Higher Inductive Types in Cubical Type Theory. (2018). arXiv:arXiv:1802.01170
- [12] Jason Dagit. 2009. Type-correct changes—a safe approach to version control implementation. (2009). MS Thesis.
- [13] Peter Dybjer. 1994. Inductive families. *Formal Aspects of Computing* 6, 4 (01 Jul 1994), 440–465. DOI: <http://dx.doi.org/10.1007/BF01211308>
- [14] Gabriel Ebner, Sebastian Ullrich, Jared Roesch, Jeremy Avigad, and Leonardo de Moura. 2017. A Metaprogramming Framework for Formal Verification. *Proceedings of the ACM on Programming Languages* 1, ICFP, Article 34 (August 2017), 29 pages.

- [15] Matthew Flatt, Ryan Culpepper, David Darais, and Robert Bruce Findler. 2012. Macros that Work Together: Compile-time bindings, partial expansion, and definition contexts. *Journal of Functional Programming* 22, 2 (2012), 181–216.
- [16] Pepijn Kokke and Wouter Swierstra. 2015. Auto in Agda. In *Lecture Notes in Computer Science*, vol 9129. Springer, Cham. In: Hinze R., Voigtländer J. (eds) Mathematics of Program Construction (MPC).
- [17] Daniel R. Licata. 2011. Running Circles Around (In) Your Proof Assistant; or, Quotients that Compute. (April 2011). <https://homotopytypetheory.org/2011/04/23/running-circles-around-in-your-proof-assistant>.
- [18] Samuel Mimram and Cinzia Di Giusto. 2013. A categorical theory of patches. *Electronic Notes in Theoretic Computer Science*, 298:283–307.
- [19] Ulf Norell. 2007. *Towards a practical programming language based on dependent type theory*. Ph.D. Dissertation. Department of Computer Science and Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden.
- [20] Pascal Paillier. 1999. Public-key cryptosystems based on composite degree residuosity classes. In: Proceedings of the 18th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT), Prague, Czech Republic.
- [21] Raluca Ada Popa, Catherine M.S. Redfield, and Hari Balakrishnan Nickolai Zeldovich. 2011. CryptDB: Protecting confidentiality with encrypted query processing. In: Proceedings of the 23rd ACM Symposium on Operating Systems Principles (SOSP), Cascais, Portugal.
- [22] Adi Shamir Ron Rivest and Leonard Adleman. 1978. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21 (2), pp. 120-126.
- [23] David Roundy. 2005. Darcs: Distributed version management in haskell. In *ACM SIGPLAN Workshop on Haskell*. ACM.
- [24] Jurriën Stutterheim, Wouter Swierstra, and Doaitse Swierstra. 2013. Forty hours of declarative programming: Teaching Prolog at the Junior College Utrecht. In *Electronic Proceedings in Theoretical Computer Science*, volume 106, pages 50–62, 2013. In: Proceedings First International Workshop on Trends in Functional Programming in Education, University of St. Andrews, Scotland, UK, June 2012.
- [25] The Univalent Foundations Program. 2013. *Homotopy Type Theory: Univalent Foundations of Mathematics*. <https://homotopytypetheory.org/book>, Institute for Advanced Study.
- [26] Paul van der Walt. 2012. Reflection in Agda. (2012). Master's thesis, Department of Computer Science, Utrecht University, Utrecht, Netherlands (2012).
- [27] Paul van der Walt and Wouter Swierstra. 2013. Engineering proof by reflection in Agda. In *Lecture Notes in Computer Science*, pages 157–173. Springer Berlin Heidelberg, 2013. In: In Ralf Hinze, (ed) Implementation and Application of Functional Languages.
- [28] Paventhan Vivekanandan. 2018. A Homotopical Approach to Cryptography. (July 2018). To be presented at Workshop on Foundations of Computer Security (FCS 2018), University of Oxford, UK.