Barrelfish Project  
ETH Zurich  

Filet-o-Fish  
When French Cuisine Meets Swiss Fishes  

Barrelfish Technical Note 024  

Pierre-Evariste Dagand  

02.06.2017  

Systems Group  
Department of Computer Science  
ETH Zurich  
CAB F.79, Universitätstrasse 6, Zurich 8092, Switzerland  
http://www.barrelfish.org/
# Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>02.06.2017</td>
<td>TR</td>
<td>Converted to Technical Note from Pierre’s LHS</td>
</tr>
</tbody>
</table>
# Contents

I The Filet-o-Fish Language 7

1 Filet-o-Fish Syntax 9
   1.1 Filet-o-Fish pure expressions ................................................. 9
       1.1.1 Types .................................................................................. 9
       1.1.2 Pure Expressions .................................................................. 12
   1.2 Filet-o-Fish standard constructs ..................................................... 15
       1.2.1 Functor instance .................................................................... 17

2 Filet-o-Fish Semantics 19
   2.1 Functional core interpreter .............................................................. 19
   2.2 Building the FoF interpreter and compiler ......................................... 23
       2.2.1 Gluing the Interpreter ............................................................. 23
       2.2.2 Gluing the Compiler .............................................................. 24
   2.3 Plumbing Machinery ........................................................................ 25
       2.3.1 The Semantics Monad ............................................................. 26
       2.3.2 Folding the Free Monad ........................................................ 26
       2.3.3 Sequencing in the Free Monad ............................................... 26

3 Filet-o-Fish Operators 27
   3.1 Arrays ........................................................................................... 27
       3.1.1 Smart Constructors .............................................................. 27
       3.1.2 Run Instantiation .................................................................. 28
       3.1.3 Compile Instantiation ............................................................ 28
   3.2 Conditionals ..................................................................................... 29
       3.2.1 Smart Constructors .............................................................. 29
       3.2.2 Compile Instantiation ............................................................ 30
       3.2.3 Run Instantiation .................................................................. 31
   3.3 Enumeration ..................................................................................... 33
       3.3.1 Smart Constructors .............................................................. 33
       3.3.2 Compile Instantiation ............................................................ 33
       3.3.3 Run Instantiation .................................................................. 34
   3.4 Function Definition ........................................................................... 34
       3.4.1 Smart Constructors .............................................................. 34
       3.4.2 Compile Instantiation ............................................................ 35
       3.4.3 Run Instantiation .................................................................. 35
   3.5 Reference Cells .................................................................................. 36
       3.5.1 Smart Constructors .............................................................. 36
       3.5.2 Compile Instantiation ............................................................ 37
       3.5.3 Run Instantiation .................................................................. 37
   3.6 Strings .............................................................................................. 38
       3.6.1 Smart Constructors .............................................................. 38
       3.6.2 Compile Instantiation ............................................................ 38
       3.6.3 Run Instantiation .................................................................. 38
   3.7 Structures Definition ........................................................................ 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7.1</td>
<td>Smart Constructors</td>
<td>39</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Compile Instantiation</td>
<td>40</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Run Instantiation</td>
<td>40</td>
</tr>
<tr>
<td>3.8</td>
<td>Type Definition</td>
<td>41</td>
</tr>
<tr>
<td>3.8.1</td>
<td>Smart Constructors</td>
<td>41</td>
</tr>
<tr>
<td>3.8.2</td>
<td>Compile Instantiation</td>
<td>42</td>
</tr>
<tr>
<td>3.8.3</td>
<td>Run Instantiation</td>
<td>42</td>
</tr>
<tr>
<td>3.9</td>
<td>Unions Definition</td>
<td>42</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Smart Constructors</td>
<td>42</td>
</tr>
<tr>
<td>3.9.2</td>
<td>Compile Instantiation</td>
<td>43</td>
</tr>
<tr>
<td>3.9.3</td>
<td>Run Instantiation</td>
<td>44</td>
</tr>
<tr>
<td>4.1</td>
<td>Printf</td>
<td>45</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Smart Constructors</td>
<td>45</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Compile Instantiation</td>
<td>45</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Run Instantiation</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Assert</td>
<td>46</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Smart Constructors</td>
<td>46</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Compile Instantiation</td>
<td>46</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Run Instantiation</td>
<td>46</td>
</tr>
<tr>
<td>5.1</td>
<td>Has Descendants</td>
<td>47</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Smart Constructors</td>
<td>47</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Compile Instantiation</td>
<td>47</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Run Instantiation</td>
<td>48</td>
</tr>
<tr>
<td>5.2</td>
<td>Mem To Phys</td>
<td>48</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Smart Constructors</td>
<td>48</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Compile Instantiation</td>
<td>48</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Run Instantiation</td>
<td>48</td>
</tr>
<tr>
<td>II</td>
<td>The Filet-o-Fish Compiler</td>
<td>49</td>
</tr>
<tr>
<td>6.1</td>
<td>The FoF Intermediate Language</td>
<td>51</td>
</tr>
<tr>
<td>6.2</td>
<td>Translating FoFCode to IL.FoF</td>
<td>52</td>
</tr>
<tr>
<td>6.2.1</td>
<td>The compiler</td>
<td>52</td>
</tr>
<tr>
<td>6.2.2</td>
<td>The machinery</td>
<td>53</td>
</tr>
<tr>
<td>6.3</td>
<td>Evaluator</td>
<td>55</td>
</tr>
<tr>
<td>7.1</td>
<td>The Paka Intermediate Language</td>
<td>56</td>
</tr>
<tr>
<td>7.2</td>
<td>Paka building blocks</td>
<td>56</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Low-level machinery</td>
<td>58</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Building PakaCode</td>
<td>58</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Building PakaIntra</td>
<td>59</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Building ILPaka</td>
<td>60</td>
</tr>
<tr>
<td>7.3</td>
<td>Translating IL.FoF to IL.Paka</td>
<td>61</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Compiling Function definition</td>
<td>61</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Compiling Constant</td>
<td>62</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Compiling Closing statements</td>
<td>62</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Compiling control-flow operators</td>
<td>62</td>
</tr>
<tr>
<td>7.3.5</td>
<td>Compiling statements</td>
<td>64</td>
</tr>
<tr>
<td>7.4</td>
<td>Translating IL.Paka to C</td>
<td>72</td>
</tr>
</tbody>
</table>
Introduction

The Filet-o-Fish contains a battered fish patty made mostly from pollock and/or hoki.

Wikipedia

Filet-o-Fish, abbreviated FoF hereafter, is a tool for the working language designer. Developed in the context of Barrellfish[5], FoF aims at easing the development of Domain-Specific Languages (DSL) as well as enhancing their safety. As a side effect of FoF’s design, it also becomes easier for the user of a DSL to understand “what is going on”.

To achieve this goal, Filet-o-Fish defines a set of combinators. A combinator is a Haskell function manipulating some Haskell data-types. In this case, our combinators manipulate an abstraction of the C language constructs, such as integers, floats, structures, arrays, etc. Altogether, this set of combinators defines an embedded language in Haskell. To avoid the confusion with the DSLs we are willing to implement, we term this embedded language the meta-language.

You seems confused now. Listen. The Hamlet compiler is implemented with Filet-o-Fish. Hamlet is a Domain-Specific Language. In Hamlet’s compiler, we use FoF to get the job done, ie. to get the actual C code out of our capability system description. Hence, the Hamlet compiler is partly developed in the FoF meta-language. Understood?

However, Filet-o-Fish is much more than a language to get the job done: being able to compile the meta-language to C is just one side-effect of our work. By writing a DSL compiler with FoF, you actually define the semantics of the DSL. Whereas the syntax defines the set of legal expressions of a language, the semantics assign a meaning to the terms of the language. Note that the C language does not have any formal semantics. And, no, this is not normal. This is Evil.

For a DSL, the benefit of having a formal semantics is twofold. First, the semantics of your DSL is the most precise and accurate description of the behavior of your domain-specific constructs. An informal, in-English specification of the DSL might fail to capture some specific points. The formal semantics is an ultimate documentation, which doesn’t lie. Second, defining a formal semantics is a necessary step before any compiler correctness proof, be it mechanized or on paper. Therefore, thanks to FoF, you get a formal, mechanized semantics of your DSL. And this is for free.

Finally, this document is the literate Haskell code of Filet-o-Fish: the code described in the following pages is the one that is compiled by the Haskell compiler. Therefore, this is the most accurate, up-to-date documentation of FoF’s internals.

So much marketing, let us look at the code.
Part I

The Filet-o-Fish Language
The Filet-O-Fish Language

Filet-o-Fish is organized in a modular way. This is reflected by the definition of the syntax of the language in Chapter 1. Indeed, the language is organized around the purely functional core of C, as described in Section 1.1. This core is extended by several constructs that are the operationally rich building blocks of the language, as described in Section 1.2.

The functional semantics of this language is then implemented in Chapter 2. Following the modular definition of the language, we first implement an interpreter for the core language (Section 2.1). In Section 2.2, we gather the per-construct interpreter under one general function. In Section 2.3, we build the machinery to automatically compute an interpreter and a compiler for the whole language.

Further, in Chapter 3, we implement the interpreter and Filet-o-Fish interpretation of the constructs. Similarly, Chapter 4 and Chapter 5, we define foreign functions mirroring the C library and the barreelfish library. These chapters are bound to be extended as long as foreign functions are needed. This is a natural process made easy by the modular design of the syntax and semantics of Filet-o-Fish.
Chapter 1

Filet-o-Fish Syntax

1.1 Filet-o-Fish pure expressions

The core of Filet-o-Fish is organized around the purely functional core of C. It consists of C types as well as C expressions.

1.1.1 Types

Data-type Definitions

The TypeExpr data-type encompasses the following types:

- Void,
- Integers, of various signedness and size,
- Float,
- Named structures and unions,
- Named pointers, i.e., a pointer recurring in a structure or union,
- Arrays, and
- Pointers

Note that a value of type TInt or TFloat is a constant, like 2, 3/7, or sizeof(struct foo). In FoF meta-language, a C variable is not a value — but a construct. So, the type of the variable x defined by int32_t x = 4 is not TInt Signed TInt32.

```
data TypeExpr = TVoid
   | TInt Signedness Size
   | TFloat
   | TChar
```
Functions  A function is represented by an Haskell function, taking a list of arguments and computing the body of the function. In the jargon, this is called an higher-order abstract syntax. So, the function definition is represented by the following type:

```haskell
data Function = Fun ([PureExpr] → FoFCode PureExpr)
```

Because `TypeExpr` is showable, `Function` has to be showable too. While we could define a more complete `Show` instance for `Function`, we will not do so here and simply return an opaque name.

```haskell
instance Show Function where
  show _ = "<fun>"
```

Concerning equality, this becomes more tricky. We would have to define what “equality” means and if that definition is decidable. Here, we consider syntactic equality and although we could decide whether two functions are syntactically equal or not, we will not do so for the moment. We simply consider functions as always distinct.

```haskell
instance Eq Function where
  _ ≡ _ = False
```

Composed data-types  Composed data-types have several allocation policies: they might be declared dynamically, using malloc, or statically, on the stack. This is reflected by the following definitions. We chose to use different definitions for each kind of data-type because they are likely to evolve in future versions and diverge from this common scheme.

```haskell
data AllocStruct = StaticStruct |
  DynamicStruct
  deriving (Eq, Show)
data AllocUnion = StaticUnion |
  DynamicUnion
  deriving (Eq, Show)
data AllocArray = StaticArray Int |
  DynamicArray
  deriving (Eq, Show)
```

Both Structures and Unions rely on the `TFieldList` synonym. Basically, the type of a Structure corresponds to its name as well as the list of its field names and respective types.

```haskell
type TFieldList = [(String, TypeExpr)]
```
Integers  Signedness and size of integers is defined as usual. An integer is either signed or unsigned and its size may vary from 8 to 64 bits. Interestingly, we derive Ord on these data-types: Ord provides us with a comparison function on the signedness and size. In practice, we can check that a cast is a correct downcasting by enforcing that the sign and size we cast to is bigger than the original sign and size.

```haskell
data Signedness = Unsigned
  | Signed
deriving (Eq, Ord, Show)
data Size = TInt8
  | TInt16
  | TInt32
  | TInt64
deriving (Eq, Ord, Show)
```

Pointers  As we understand that the suspense is unbearable, we are going to reveal you the type of $x$ defined above. Actually, the type of $x$ is TPointer (TInt Signed TInt32) Avail. A pointer? Indeed, a variable does actually points to a location in memory. This choice allows us to capture the notion of variables and pointers in a single abstraction, called a reference cell.

A reference cell can be in one of the following states: either Available or Read. This distinction makes sense during the compilation process, it can ignored otherwise.

```haskell
data Mode = Avail
  | Read
deriving (Eq, Show)
```

Smart Constructors

In some circumstances, it is necessary to explicitly write the type of an expression. However, explicitly combining the previously defined types can be quite cumbersome. For example, we can naturally define the base types as follow:

```haskell
voidT :: TypeExpr
voidT = TVoid
uint8T, uint16T, uint32T, uint64T :: TypeExpr
uint8T = TInt Unsigned TInt8
uint16T = TInt Unsigned TInt16
uint32T = TInt Unsigned TInt32
uint64T = TInt Unsigned TInt64
int8T, int16T, int32T, int64T :: TypeExpr
int8T = TInt Signed TInt8
int16T = TInt Signed TInt16
int32T = TInt Signed TInt32
int64T = TInt Signed TInt64
floatT :: TypeExpr
floatT = TFloat
charT :: TypeExpr
charT = TChar
uintptrT :: TypeExpr
uintptrT = TCompPointer "void"
```
And, similarly, we can build up composed types by applying them on smaller types:

\[
\begin{align*}
\text{arrayDT} :: & \quad \text{TypeExpr} \rightarrow \text{TypeExpr} \\
\text{arrayDT} \ typ = & \quad \text{ToArray DynamicArray} \ typ \\
\text{arrayST} :: & \quad \text{Int} \rightarrow \text{TypeExpr} \rightarrow \text{TypeExpr} \\
\text{arrayST} \ size \ typ = & \quad \text{ToArray (StaticArray} \ size) \ typ \\
\text{ptrT} :: & \quad \text{TypeExpr} \rightarrow \text{TypeExpr} \\
\text{ptrT} \ typ = & \quad \text{TPointer} \ typ \ \text{Avail} \\
\text{structDT}, & \quad \text{unionDT}, \\
\text{structST}, & \quad \text{unionST} :: \quad \text{String} \rightarrow \text{TFieldList} \rightarrow \text{TypeExpr} \\
\text{structDT} \ name \ fields = & \quad \text{TStruct DynamicStruct} \ name \ fields \\
\text{unionDT} \ name \ fields = & \quad \text{TUnion DynamicUnion} \ name \ fields \\
\text{structST} \ name \ fields = & \quad \text{TStruct StaticStruct} \ name \ fields \\
\text{unionST} \ name \ fields = & \quad \text{TUnion StaticUnion} \ name \ fields \\
\text{enumT} :: & \quad \text{String} \rightarrow \{(\text{String}, \text{Int})\} \rightarrow \text{TypeExpr} \\
\text{enumT} \ name \ fields = & \quad \text{TEnum} \ name \ fields \\
\text{typedef} :: & \quad \text{TypeExpr} \rightarrow \text{String} \rightarrow \text{TypeExpr} \\
\text{typedef} \ typ \ name = & \quad \text{TTypedef} \ typ \ name
\end{align*}
\]

Finally, the named pointer – which is actually a fix-point – takes as input the name of the structure or union it refers to.

\[
\begin{align*}
\text{cptrT} :: & \quad \text{String} \rightarrow \text{TypeExpr} \\
\text{cptrT} \ id = & \quad \text{TCompPointer} \ id
\end{align*}
\]

1.1.2 Pure Expressions

In a first step, we are going to define the expressions composing FoF meta-language. As for types, this consists in a data-type, \text{PureExpr}, capturing the syntax of expressions. Then, we also define some smart constructors.

Data-type Definitions

An expression is one of the following object:

- \text{void}, the only object populating the type \text{Void},
- an integer, of specific signedness and size,
- a float,
- a reference to an object in memory,
- a unary operation, applied to an object,
- a binary operation, applied on two objects,
- the \text{sizeof} operator, applied to a type,
- a conditional expression, testing an object against 0, returning one of two objects, and
- a \text{cast} operator, casting an object to a given type

\[
\text{data PureExpr} = \text{Void} \\
\quad \mid \text{CLInteger Signedness Size Integer} \\
\quad \mid \text{CLFloat Float} \\
\quad \mid \text{CLChar Char}
\]
Variable names  A reference is identified by a name. A Generated name has been forged by FoF. A Provided name has been defined by the compiler designer. An Inherited name results from an operation performed on another variable. We carefully track the origin of names for compilation purpose: for example, if a variable name has been Generated, we should try to eliminate it, to make the compiled code more readable.

```
data VarName = Generated String
  | Provided String
  | Inherited Int VarName
  deriving (Show, Eq)
```

A reference is also decorated by its origin. This field is used by the compiler to identify the scope of variables. Therefore, the compiler can enforce some safety checks, such as verifying that the address of a local variable is not assigned to a global one, for example. Sadly, this information is not always precisely maintained nor correctly used in the current implementation. More care and more checks should be added in the future, to ensure the correctness of the generated code.

```
data Origin = Local
  | Global
  | Param
  | Dynamic
  deriving (Eq, Show)
```

Unary operations  The unary operations are either the arithmetic minus operation, or the logic complement operation, or the logic negation operation.

```
data UnaryOp = Minus | Complement | Negation
  deriving (Eq, Show)
```

Binary operations  The binary operations are either arithmetic operators (+, −, ×, ÷, and %), Boolean operators (<<, >>, &, bitwise-or, and ^), or comparison operators (<, <=, >, >=, ==, and !=).

```
data BinaryOp = Plus | Sub | Mul | Div | Mod
  | Shl | Shr | AndBit | OrBit | XorBit
  | Le | Leq | Ge
  | Geq | Eq | Neq
  deriving (Eq, Show)
```

Smart Constructors

As usual, we define some constructors for the C programmer to feel at home with FoF. Let us start with the constants first:
Then come the unary operators:

```
minus, comp, neg :: PureExpr → PureExpr
minus = Unary Minus
comp = Unary Complement
neg = Unary Negation
```

And the binary operators. Note that they are defined *infix*. Therefore, it becomes possible to write the following code:

```
exampleInfix :: PureExpr
exampleInfix = (uint8 1) .< .((uint8 2) .+ .(uint8 4))
```

Although not specified yet, we could have set up the left/right associativity and precedence rules of these operators. This would reduce the parenthesizing overhead. It is just a matter of doing it.

```
( .+.), (.-.), (.*.), (/./), (.%.),
( .<< .), ( .>> .), ( .& .), ( .| .), ( .^ .),
( . < .), (. <= .), (. > .),
( . >= .), (. == .), (. /= .) :: PureExpr → PureExpr → PureExpr
```

( .+. ) = Binary Plus
( .- . ) = Binary Sub
( .*. ) = Binary Mul
( ./ ) = Binary Div
( .%/.) = Binary Mod
( .<< . ) = Binary Shl
( .>> . ) = Binary Shr
( .& . ) = Binary AndBit
( .| . ) = Binary OrBit
( .^ . ) = Binary XorBit
( .< . ) = Binary Le
( .<= . ) = Binary Leq
( .> . ) = Binary Ge

FiletOFish - 14
Barrellfish TN-024
Finally, sizeof, conditionals, and cast have their straightforward alter-ego in FoF:

\[
\begin{align*}
\text{sizeof} & \colon \text{TypeExpr} \rightarrow \text{PureExpr} \\
\text{sizeof } t & = \text{Sizeof } t \\
\text{test} & \colon \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \\
\text{test } c \text{ if } t \text{ iff } & = \text{Test } c \text{ if } t \text{ iff} \\
\text{cast} & \colon \text{TypeExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \\
\text{cast } t \text{ e } & = \text{Cast } t \text{ e}
\end{align*}
\]

When compiling foreign function calls, one might need to turn a (Haskell) string into a FoF quote object. This is achieved by the following combinator. One must avoid using this operation as much as possible: this quotation has no semantic meaning, therefore one should use it only when we are really sure we are not interested in the quoted semantic anymore.

\[
\begin{align*}
\text{quote} & \colon \text{String} \rightarrow \text{PureExpr} \\
\text{quote } s & = \text{Quote } s
\end{align*}
\]

### 1.2 Filet-o-Fish standard constructs

The FoF language is defined by the syntax tree below. It gathers every constructs defined in the Constructs directory as well as foreign functions defined in the Libc and Libbarrelfish directories.

```haskell
data FoFConst a

Foreign-call to libc Assert:

\[ = \text{Assert } \text{PureExpr } a \]

Foreign-call to libc Printf:

\[ | \text{Printf } \text{String} \left[ \text{PureExpr} \right] \text{ a} \]

Foreign-call to libarrelfish has_.descendants:

\[ | \text{HasDescendants } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Foreign-call to libarrelfish mem_.to_.phys:

\[ | \text{MemToPhys } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Foreign-call to Hamlet get_.address:

\[ | \text{GetAddress } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Support for Union:

\[ | \text{NewUnion } (\text{Maybe } \text{String}) \text{ AllocUnion } \text{String} \left[ \left( \text{String}, \text{TypeExpr} \right) \right] \left( \text{String}, \text{Data} \right) \left( \text{Loc } \rightarrow \text{ a} \right) \]

\[ | \text{ReadUnion } \text{Loc } \text{String } (\text{Data } \rightarrow \text{ a}) \]

\[ | \text{WriteUnion } \text{Loc } \text{String } \text{Data } \text{ a} \]

Support for TYPedef:

\[
(\geq) = \text{Binary Geq} \\
(\eq) = \text{Binary Eq} \\
(\neq) = \text{Binary Neq}
\]

1.2 Filet-o-Fish standard constructs

The FoF language is defined by the syntax tree below. It gathers every constructs defined in the Constructs directory as well as foreign functions defined in the Libc and Libbarrelfish directories.

```haskell
data FoFConst a

Foreign-call to libc Assert:

\[ = \text{Assert } \text{PureExpr } a \]

Foreign-call to libc Printf:

\[ | \text{Printf } \text{String} \left[ \text{PureExpr} \right] \text{ a} \]

Foreign-call to libarrelfish has_.descendants:

\[ | \text{HasDescendants } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Foreign-call to libarrelfish mem_.to_.phys:

\[ | \text{MemToPhys } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Foreign-call to Hamlet get_.address:

\[ | \text{GetAddress } (\text{Maybe } \text{String}) \text{ PureExpr } (\text{PureExpr } \rightarrow \text{ a}) \]

Support for Union:

\[ | \text{NewUnion } (\text{Maybe } \text{String}) \text{ AllocUnion } \text{String} \left[ \left( \text{String}, \text{TypeExpr} \right) \right] \left( \text{String}, \text{Data} \right) \left( \text{Loc } \rightarrow \text{ a} \right) \]

\[ | \text{ReadUnion } \text{Loc } \text{String } (\text{Data } \rightarrow \text{ a}) \]

\[ | \text{WriteUnion } \text{Loc } \text{String } \text{Data } \text{ a} \]

Support for TYPedef:
typedef TypeExpr a
typedef String TypeExpr a

Support for Structures:
| NewStruct (Maybe String) AllocStruct String [(String, (TypeExpr, Data))] (Loc → a)
| ReadStruct Loc String (Data → a)
| WriteStruct Loc String Data a

Support for Strings:
| NewString (Maybe String) String (Loc → a)

Support for Reference cells:
| NewRef (Maybe String) Data (Loc → a)
| ReadRef Loc (Data → a)
| WriteRef Loc Data a

Support for Functions:
| NewDef [FunAttr] String Function TypeExpr [(TypeExpr, Maybe String)]
  (PureExpr → a)
| CallDef (Maybe String) PureExpr [PureExpr]
  (PureExpr → a)
| Return PureExpr

Support for Enumerations:
| NewEnum (Maybe String) String Enumeration String (Loc → a)

Support for Conditionals:
| If (FoFCode PureExpr)
  (FoFCode PureExpr)
  (FoFCode PureExpr) a
| For (FoFCode PureExpr)
  (FoFCode PureExpr)
  (FoFCode PureExpr)
  (FoFCode PureExpr) a
| While (FoFCode PureExpr)
  (FoFCode PureExpr) a
| DoWhile (FoFCode PureExpr)
  (FoFCode PureExpr) a
| Switch PureExpr
  [((PureExpr, FoFCode PureExpr))]
  (FoFCode PureExpr) a
| Break
| Continue

Support for Arrays:
| NewArray (Maybe String) AllocArray [Data] (Loc → a)
| ReadArray Loc Index (Data → a)
| WriteArray Loc Index Data a

The following type synonyms have been used above as a documentation purpose. A Data represents a value used to initialize a data-structure. A Loc represents a reference. An Index is a value used to index an array.

code

| type Data = PureExpr
| type Loc = PureExpr
| type Index = PureExpr

desc
A function can be characterized by the following attributes, following their C semantics:

```haskell
data FunAttr = Static
| Inline
deriving (Eq)
```

```haskell
instance Show FunAttr where
  show Static = "static"
  show Inline = "inline"
```

### Enumeration

When defining an enumeration, we use the following type synonym to describe the list of pair name-value:

```haskell
type Enumeration = [(String, Int)]
```

### 1.2.1 Functor instance

A crucial specificity of FoFConst is that it defines a functor. This functor is defined as follow.

```haskell
instance Functor FoFConst where
  fmap f (Assert a b) = Assert a (f b)
  fmap f (Printf a b c) = Printf a b (f c)
  fmap f (HasDescendants a b c) = HasDescendants a b (f c)
  fmap f (MemToPhys a b c) = MemToPhys a b (f c)
  fmap f (GetAddress a b c) = GetAddress a b (f c)
  fmap f (NewUnion a b c d e) = NewUnion a b c d e (f c)
  fmap f (ReadUnion a b c d) = ReadUnion a b (f c)
  fmap f (WriteUnion a b c d) = WriteUnion a b c d (f c)
  fmap f (Typedef a c) = Typedef a (f c)
  fmap f (TypedefE a b c) = TypedefE a b (f c)
  fmap f (NewStruct a b c d e) = NewStruct a b c d e (f c)
  fmap f (ReadStruct a b c d) = ReadStruct a b c d (f c)
  fmap f (WriteStruct a b c d) = WriteStruct a b c d (f c)
  fmap f (NewString a b c) = NewString a b c (f c)
  fmap f (NewRef a b c) = NewRef a b c (f c)
  fmap f (ReadRef a b c) = ReadRef a b c (f c)
  fmap f (WriteRef a b c) = WriteRef a b c (f c)
  fmap g (NewDef a b c d e f) = NewDef a b c d e (g f)
  fmap f (CallDef a b c d) = CallDef a b c d (f c)
  fmap f (Return a) = Return a
  fmap f (If a b c d e) = If a b c d e (f c)
  fmap f (For a b c d e) = For a b c d e (f c)
  fmap f (While a b c d) = While a b c d (f c)
  fmap f (DoWhile a b d) = DoWhile a b c d (f c)
  fmap f (Switch a b c d) = Switch a b c d (f c)
  fmap f (Break) = Break
  fmap f (Continue) = Continue
  fmap f (NewArray a b c d e) = NewArray a b c d e (f c)
  fmap f (ReadArray a b c d) = ReadArray a b c d (f c)
  fmap f (WriteArray a b c d) = WriteArray a b c d (f c)
```
Thanks to this functor structure, it makes sense to embed \textit{FoFConst} in a \textit{Semantics}: the machinery we build in Chapter 2.3 will take care of transforming this functor into a free monad. Hence the following type synonym.

\begin{verbatim}
    type FoFCode a = Semantics FoFConst a
\end{verbatim}
Chapter 2

Filet-o-Fish Semantics

So, logically...
If...
she...
weighs...
the same as a duck...
she's made of wood.

Monty Python

2.1 Functional core interpreter

In this Section, we implement an expression evaluator. Given any (correct) expression, it will compute
the corresponding value. The implementation is decomposed in several steps. In Section 2.1, we eval-
uate top-level expressions. Doing so, we rely on case-specific evaluators. This includes unary operators
(Section 2.1), binary operators (Section 2.1), the sizeof operation (Section 2.1), the conditional operation
(Section 2.1), and the cast operation (Section 2.1).

Note that the following functions are partial: not all expressions can be successfully evaluated. Indeed,
some operations are simply meaningless. For example, computing the sum of a structure and a float
is illegal. Currently, we are simply ignore these errors and this might result in run-time errors of the
DSL compiler. Satisfactory solutions of this problem exist, though. For example, we could implement
a type-checker that would ensure the absence of run-time errors. Another approach would be improve
our error handling code.

Top-level Evaluation

The purpose of this section is implement the following function:

\[
\text{symbEval} :: \text{PureExpr} \rightarrow \text{PureExpr}
\]

That reduces a given expression to a value. Hence, for values, this is trivial:

\[
\text{symbEval Void = Void} \\
\text{symbEval x@(CLInteger _) = x} \\
\text{symbEval x@(CLFloat _) = x} \\
\text{symbEval x@(CLRef _) = x}
\]

Then, for inductive constructions, we rely on the specific functions implemented in the following sec-
tions.
Unary Operator Evaluation

For unary operators, we need to implement the following function:

\[
symbEvalUnary :: \text{UnaryOp} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr}
\]

Hence the following code:

\[
symbEvalUnary \text{ Minus } x = \\
\hspace{1em} \text{case } x \text{ of } \\
\hspace{2em} \text{CLInteger Signed size x } \rightarrow \text{CLInteger Signed size } (-x) \\
\hspace{2em} \text{CLFloat x } \rightarrow \text{CLFloat } (-x) \\
\hspace{2em} _ \rightarrow \text{error } "\text{symbEvalUnary: minus on wrong type}" \\
symbEvalUnary \text{ Complement } x = \\
\hspace{1em} \text{case } x \text{ of } \\
\hspace{2em} \text{CLInteger sg sz x } \rightarrow \text{CLInteger sg sz } (\text{complement } x) \\
\hspace{2em} _ \rightarrow \text{error } "\text{symbEvalUnary: complement on wrong type}" \\
symbEvalUnary \text{ Negation } x = \\
\hspace{1em} \text{case } x \text{ of } \\
\hspace{2em} \text{CLInteger sg sz 0 } \rightarrow \text{CLInteger sg sz 1} \\
\hspace{2em} \text{CLInteger sg sz } _ \rightarrow \text{CLInteger sg sz 0} \\
\hspace{2em} _ \rightarrow \text{error } "\text{symbEvalUnary: negation on wrong type}" \\
\]

Binary Operator Evaluation

For binary operators, here is our goal:

\[
symbEvalBinary :: \text{BinaryOp} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr}
\]

Achieved by the following, messy codes.

**Arithmetic Operations**

\[
symbEvalBinary \text{ Plus } (\text{CLInteger sg sz x}) (\text{CLInteger sg' sz' y}) \\
\| \text{sg} \equiv \text{sg'} \land \text{sz} \equiv \text{sz'} = \text{CLInteger sg sz } (x + y) \\
\| \text{otherwise } = \text{error } "\text{symbEvalBinary: Plus undefined}" \\
symbEvalBinary \text{ Plus } (\text{CLInteger } _ _ x) (\text{CLFloat y}) =
\]
More checks should be added here. For examples, we should ensure that the result of the subtraction of two unsigned numbers is still positive, or make it wrap.

Boolean Operations

Boolean Operations
```
symbEvalBinary AndBit _ _ = error "symbEvalBinary: And undefined"
symbEvalBinary OrBit (CLInteger sg si x) (CLInteger sg' si' y)
  | sg \equiv sg' \land si \equiv si' = CLInteger sg si (x B_.&. y)
  | otherwise = error "symbEvalBinary: Or undefined"
symbEvalBinary XorBit (CLInteger sg si x) (CLInteger sg' si' y)
  | sg \equiv sg' \land si \equiv si' = CLInteger sg si (x `xor` y)
  | otherwise = error "symbEvalBinary: Xor undefined"
symbEvalBinary XorBit _ _ = error "symbEvalBinary: Xor undefined"

Comparison Operations
symbEvalBinary op (CLInteger sg si x) (CLInteger sg' si' y)
  | sg \equiv sg' \land si \equiv si' = symbEvalComp op x y
  | otherwise = error ("symbEvalBinary: " ++ show op ++ " undefined")
symbEvalBinary op (CLInteger _ _ x) (CLFloat y) =
  symbEvalComp op (fromRational $ toRational x) y
symbEvalBinary op (CLFloat x) (CLInteger _ _ y) =
  symbEvalComp op x (fromRational $ toRational y)
symbEvalBinary op (CLFloat x) (CLFloat y) = symbEvalComp op x y

symbEvalComp :: (Ord a, Num a) \Rightarrow BinaryOp \rightarrow a \rightarrow a \rightarrow PureExpr
symbEvalComp op x y =
  let cmp = case op of
    Le \rightarrow (<)
    Leq \rightarrow (\leq)
    Ge \rightarrow (>)
    Geq \rightarrow (\geq)
    Eq \rightarrow (\equiv)
    Neq \rightarrow (\not\equiv) in
  if cmp x y then
    CLInteger Unsigned TInt64 1
  else CLInteger Unsigned TInt64 0

Sizeof Evaluation

Our sizeof operator follows the corresponding C operation:
symbEvalSizeof :: TypeExpr \rightarrow PureExpr
symbEvalSizeof TVoid = CLInteger Unsigned TInt64 1
symbEvalSizeof (TInt _ TInt8) = CLInteger Unsigned TInt64 1
symbEvalSizeof (TInt _ TInt16) = CLInteger Unsigned TInt64 2
symbEvalSizeof (TInt _ TInt32) = CLInteger Unsigned TInt64 4
symbEvalSizeof (TInt _ TInt64) = CLInteger Unsigned TInt64 8
symbEvalSizeof TFloat = CLInteger Unsigned TInt64 4
```
Conditionals Evaluation

The semantics of the conditional mimics a restricted version of the C standard: True corresponds to everything which is not a float or integer equal to zero. Hence, we evaluate the corresponding branch accordingly.

\[
\text{symbEvalTest} :: \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr}
\]

\[
\text{symbEvalTest} (\text{CLInteger } 0) \_ \_ y = \text{symbEval } y
\]

\[
\text{symbEvalTest} (\text{CLFloat } 0) \_ \_ y = \text{symbEval } y
\]

\[
\text{symbEvalTest} \_ x \_ = \text{symbEval } x
\]

Cast Evaluation

Here is our stripped-down version of cast. It will probably deserve some work in the future, as it is quite restrictive. Also, it should ensure that the type modification are reflected on the data: converting a signed, negative number to an unsigned form changes the value of this number. This is currently unsupported.

\[
\text{symbEvalCast} :: \text{TypeExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr} \rightarrow \text{PureExpr}
\]

\[
\text{symbEvalCast} (\text{TInt } \text{sg } \text{sz}) (\text{CLInteger } \text{sg}' \text{ sz}') x
\]

\[
| \text{sg}' < \text{sg} \land \text{sz}' < \text{sz} = \text{CLInteger } \text{sg sz } x
\]

\[
\text{otherwise } = \text{error} \ "\text{symbEvalCast: illegal integer cast}"
\]

\[
\text{symbEvalCast} \text{TFloat (CLInteger } \_ \_ x) =
\]

\[
\text{CLFloat (fromRational } \$ \text{toRational } x)
\]

\[
\text{symbEvalCast} \text{TFloat } vx@ (\text{CLFloat } x) = vx
\]

\[
\text{symbEvalCast} \_ \_ =
\]

\[
\text{error} \ "\text{symbEvalCast: Not yet implemented/undefined cast}"
\]

2.2 Building the FoF interpreter and compiler

In this section, we glue together the constructs of the FoF language, defined in the Constructs, Libc, and Libbarrelfish directories. This gluing builds a one-step interpreter for FoF, compileAlgebra (Section 2.2.1), and a one-step compiler, compileAlgebra (Section 2.2.2). We rely on the machinery defined in Section 2.3 to automatically build an interpreter and a compiler from these functions.

2.2.1 Gluing the Interpreter

The run-time is actually quite simple. It is described by a heap, in which we first store fresh identifiers, freshLoc, freshSLoc, and freshALoc. When we want to store a value in memory, we pick a fresh identifier and, respectively update the refMap, strMap, or arrayMap with a new map from the identifier to the value. Similarly, we can read and modify these mappings. Intuitively, the Heap is a representation of the machine’s memory.
These different maps have different purposes: \texttt{refMap} maps an identifier to a single value, \texttt{strMap} maps an identifier to a mapping from strings to values (modelling a structure or union), and \texttt{arrayMap} maps an identifier to a bounded array of values.

\[
\textbf{data} \texttt{Heap} = \texttt{Hp} \{ \\text{freshLoc} :: \texttt{Int}, \\
\text{refMap} :: [\langle \texttt{VarName}, \texttt{Data} \rangle], \\
\text{freshSLoc} :: \texttt{Int}, \\
\text{strMap} :: [\langle \texttt{VarName}, [\langle \texttt{String}, \texttt{Data} \rangle] \rangle], \\
\text{freshALoc} :: \texttt{Int}, \\
\text{arrayMap} :: [\langle \texttt{VarName}, [\texttt{Data}] \rangle] \}
\]

Then, the one-step interpreter takes a FoF term, a Heap, and returns a pair of value and resulting heap. This is simply implemented by matching the term and calling the corresponding construct-specific interpreter.

\[
\begin{align*}
\text{runAlgebra} \colon \text{FoFConst} \ (\texttt{Heap} \to (\texttt{PureExpr}, \texttt{Heap})) &\to \texttt{Heap} \to (\texttt{PureExpr}, \texttt{Heap}) \\
\text{runAlgebra} \ x@(@\texttt{NewArray - - - -}) &= \text{runArrays} \ x \\
\text{runAlgebra} \ x@(@\texttt{ReadArray - - - -}) &= \text{runArrays} \ x \\
\text{runAlgebra} \ x@(@\texttt{WriteArray - - - -}) &= \text{runArrays} \ x \\
\text{runAlgebra} \ x@(@\texttt{If - - - -}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{For - - - -}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{While - - - -}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{DoWhile - - - -}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{Switch - - - -}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{Break}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{Continue}) &= \text{runConditionals} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewEnum - - - -}) &= \text{runEnumerations} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewDef - - - -}) &= \text{runFunctions} \ x \\
\text{runAlgebra} \ x@(@\texttt{CallDef - - - -}) &= \text{runFunctions} \ x \\
\text{runAlgebra} \ x@(@\texttt{Return - - - -}) &= \text{runFunctions} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewRef - - - -}) &= \text{runReferences} \ x \\
\text{runAlgebra} \ x@(@\texttt{ReadRef - - - -}) &= \text{runReferences} \ x \\
\text{runAlgebra} \ x@(@\texttt{WriteRef - - - -}) &= \text{runReferences} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewString - - - -}) &= \text{runString} \ x \\
\text{runAlgebra} \ x@(@\texttt{Typedef - - - -}) &= \text{runTypedef} \ x \\
\text{runAlgebra} \ x@(@\texttt{TypedefE - - - -}) &= \text{runTypedef} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewStruct - - - -}) &= \text{runStructures} \ x \\
\text{runAlgebra} \ x@(@\texttt{ReadStruct - - - -}) &= \text{runStructures} \ x \\
\text{runAlgebra} \ x@(@\texttt{WriteStruct - - - -}) &= \text{runStructures} \ x \\
\text{runAlgebra} \ x@(@\texttt{NewUnion - - - -}) &= \text{runUnions} \ x \\
\text{runAlgebra} \ x@(@\texttt{ReadUnion - - - -}) &= \text{runUnions} \ x \\
\text{runAlgebra} \ x@(@\texttt{WriteUnion - - - -}) &= \text{runUnions} \ x \\
\text{runAlgebra} \ x@(@\texttt{Assert - - - -}) &= \text{runAssert} \ x \\
\text{runAlgebra} \ x@(@\texttt{Printf - - - -}) &= \text{runPrintf} \ x \\
\text{runAlgebra} \ x@(@\texttt{HasDescendants - - - -}) &= \text{runHasDescendants} \ x \\
\text{runAlgebra} \ x@(@\texttt{MemToPhys - - - -}) &= \text{runMemToPhys} \ x \\
\text{runAlgebra} \ x@(@\texttt{GetAddress - - - -}) &= \text{runGetAddress} \ x
\end{align*}
\]

\subsection{Gluing the Compiler}

Similarly, the one-step compiler is organized around the notion of Binding environment: this environment is carried over the compilation process. Hence, the Binding represents the compiler’s state:

\begin{itemize}
\item \texttt{freshVar} is a free identifier, used to generate unique variable names,
\item \texttt{def ...} maps the defined structure names with their type
\end{itemize}
\[
\text{data } \text{Binding} = \text{Binding } \{ \text{freshVar} :: \text{Int},
\text{defStructs} :: [(\text{String}, \text{TypeExpr})],
\text{defUnions} :: [(\text{String}, \text{TypeExpr})],
\text{defEnums} :: [(\text{String}, [\{(\text{String}, \text{Int})]\})] \}
\]

This binding is then modified by the one-step compiler, which takes a term, a binding, and return an FoF expression as well as an updated binding.

\[
\begin{align*}
\text{compileAlgebra} &: \text{FoFConst } (\text{Binding} \rightarrow (\text{ILFoF}, \text{Binding})) \rightarrow \\
& (\text{Binding} \rightarrow (\text{ILFoF}, \text{Binding}))
\end{align*}
\]

\[
\begin{align*}
\text{compileAlgebra } x@& (\text{NewArray } _- -_-) = \text{compileArrays } x \\
\text{compileAlgebra } x@& (\text{ReadArray } _- -_-) = \text{compileArrays } x \\
\text{compileAlgebra } x@& (\text{WriteArray } _- -_-) = \text{compileArrays } x \\
\text{compileAlgebra } x@& (\text{If } _- -_-) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{While } _- -_-) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{DoWhile } _- -_-) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{Switch } _- -_-) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{Break}) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{Continue}) = \text{compileConditionals } x \\
\text{compileAlgebra } x@& (\text{CallDef } _- -_-) = \text{compileFunctions } x \\
\text{compileAlgebra } x@& (\text{Return}) = \text{compileFunctions } x \\
\text{compileAlgebra } x@& (\text{NewEnum } _- -_-) = \text{compileEnumerations } x \\
\text{compileAlgebra } x@& (\text{NewRef } _- -_-) = \text{compileReferences } x \\
& (\text{ReadRef } _- -_-) = \text{compileReferences } x \\
& (\text{WriteRef } _- -_-) = \text{compileReferences } x \\
\text{compileAlgebra } x@& (\text{NewString } _- -_-) = \text{compileString } x \\
\text{compileAlgebra } x@& (\text{NewStruct } _- -_-) = \text{compileStructures } x \\
\text{compileAlgebra } x@& (\text{ReadStruct } _- -_-) = \text{compileStructures } x \\
\text{compileAlgebra } x@& (\text{WriteStruct } _- -_-) = \text{compileStructures } x \\
\text{compileAlgebra } x@& (\text{NewUnion } _- -_-) = \text{compileUnions } x \\
\text{compileAlgebra } x@& (\text{ReadUnion } _- -_-) = \text{compileUnions } x \\
\text{compileAlgebra } x@& (\text{WriteUnion } _- -_-) = \text{compileUnions } x \\
\text{compileAlgebra } x@& (\text{Assert}) = \text{compileAssert } x \\
\text{compileAlgebra } x@& (\text{Printf}) = \text{compilePrintf } x \\
\text{compileAlgebra } x@& (\text{HasDescendants}) = \text{compileHasDescendants } x \\
\text{compileAlgebra } x@& (\text{MemToPhys}) = \text{compileMemToPhys } x \\
\text{compileAlgebra } x@& (\text{GetAddress}) = \text{compileGetAddress } x
\end{align*}
\]

## 2.3 Plumbing Machinery

The material presented in this chapter relies on some hairy concepts from Category Theory. If you are curious about these things, Edward Kmett wrote a nice blog post [2] on the subject. The first version of FoF, and in particular this file, relied on Wouter Swierstra solution to the expression problem [6]. However, the burden of this approach on the type-system was unbearable for our users.

Our motivation is to build a monad in which one can naturally write sequential code, just as an imperative language. Each construct of the language is defined in \texttt{Constructs} by the \texttt{FoFConst} data-type. Purposely, this data-type implements a functor. The code below generically turn a functor \( f \) into a \texttt{Semantics }\( f \) monad. Hence, in \texttt{Constructs}, we apply this machinery to make a monad out of \texttt{FoFConst}. 

---

Barrelfish TN-024

FileOfFish - 25
2.3.1 The Semantics Monad

We build a monad $\text{Semantics } f$ out of a function $f$ thanks to the following data-type:

\[
\text{data } \text{Semantics } f \ a = \begin{cases} 
\text{Pure } a \\
\text{Impure (} f \ (\text{Semantics } f \ a) \\text{)}
\end{cases}
\]

First of all, we show that this defines a functor:

\[
\text{instance Functor } f \Rightarrow \text{Functor } (\text{Semantics } f) \quad \text{where}
\]

\[
fmap f (\text{Pure } x) = \text{Pure } (f \ x) \\
fmap f (\text{Impure } t) = \text{Impure } (fmap (fmap f) \ t)
\]

We need to (as of GHC 7.10) implement Applicative

\[
\text{instance (Functor } f \Rightarrow \text{Applicative } (\text{Semantics } f) \quad \text{where}
\]

\[
\text{pure } = \text{return} \\
(<*>) = \text{ap}
\]

Then, we obtain the monad:

\[
\text{instance Functor } f \Rightarrow \text{Monad } (\text{Semantics } f) \quad \text{where}
\]

\[
\text{return } = \text{Pure} \\
(\text{Pure } x) \gg f = f x \\
(\text{Impure } t) \gg f = \text{Impure } (fmap (\gg f) \ t)
\]

Terms are embedded into the monad thanks the following function:

\[
\text{inject } :: f \ (\text{Semantics } f \ a) \rightarrow \text{Semantics } f \ a \\
\text{inject } x = \text{Impure } x
\]

2.3.2 Folding the Free Monad

Finally, once we have built the monad, we will need to manipulate its content. For example, we will be willing to evaluate it, or to compile it, etc. All these operations can be implemented by folding over the monadic code, that is traversing the constructs in their definition order and computing an output of type $b$. Note that we have to distinguish $\text{Pure}$ terms, which are simply values, from $\text{Impure}$ ones, which are the embedded constructs.

\[
\text{foldSemantics } :: \text{Functor } f \Rightarrow (a \rightarrow b) \rightarrow (f \ b \rightarrow b) \rightarrow \text{Semantics } f \ a \rightarrow b
\]

\[
\text{foldSemantics } \text{pure } \text{imp } (\text{Pure } x) = \text{pure } x \\
\text{foldSemantics } \text{pure } \text{imp } (\text{Impure } t) = \text{imp } \$ \text{fmap } (\text{foldSemantics } \text{pure } \text{imp}) \ t
\]

2.3.3 Sequencing in the Free Monad

Provided a list of monadic code, we are able to turn them into a single monadic code returning a list of terms. This corresponds to the $\text{sequence}$ function in the IO monad:

\[
\text{sequenceSem } ms = \text{foldr } k (\text{return } []) \ ms \\
\quad \text{where } k \ m \ m' = \\
\quad \quad \text{do } \\
\quad \quad \quad x \leftarrow m \\
\quad \quad \quad xs \leftarrow m' \\
\quad \quad \quad \text{return } (x : xs)
\]
Chapter 3

Filet-o-Fish Operators

Listen.
Strange women lying in ponds distributing swords is no basis for a system of government. Supreme executive power derives from a mandate from the masses, not from some farcical aquatic ceremony.

Monty Python

3.1 Arrays

The Array construct, as well as the subsequent constructs, is organized as follow. First, we define some smart constructors, which are directly used by the DSL designer when implementing the compiler. Then, we implement the one-step interpreter and compiler to FoF.

Array offers an abstraction over C arrays, both statically defined or statically allocated. Hence, it offers the possibility to create, read from, and write into arrays.

3.1.1 Smart Constructors

We can create dynamic and static anonymous arrays using the following combinators:

\[
\text{newArray} :: [\text{Data}] \rightarrow \text{FoFCode Loc}
\]
\[
\text{newArray value} = \text{inject} (\text{NewArray} \ \text{Nothing} \ \text{DynamicArray} \ \text{value} \ \text{return})
\]

\[
\text{newStaticArray} :: [\text{Data}] \rightarrow \text{FoFCode Loc}
\]
\[
\text{newStaticArray value} = \text{inject} (\text{NewArray} \ \text{Nothing} \ (\text{StaticArray} \ \text{length} \ \text{value}) \ \text{value} \ \text{return})
\]

Similarly, they can be named:

\[
\text{newArrayN} :: \text{String} \rightarrow [\text{Data}] \rightarrow \text{FoFCode Loc}
\]
\[
\text{newArrayN name value} = \text{inject} (\text{NewArray} \ (\text{Just} \ \text{name}) \ \text{DynamicArray} \ \text{value} \ \text{return})
\]

\[
\text{newStaticArrayN} :: \text{String} \rightarrow [\text{Data}] \rightarrow \text{FoFCode Loc}
\]
\[
\text{newStaticArrayN name value} = \text{inject} (\text{NewArray} \ (\text{Just} \ \text{name}) \ (\text{StaticArray} \ \text{length} \ \text{value}) \ \text{value} \ \text{return})
\]

Then, we can read the content of an array:
As well as write some data in a cell:

\[ \text{writeArray} :: \text{Loc} \to \text{Index} \to \text{Data} \to \text{FoFCode} () \]
\[
\text{writeArray } l f d = \text{inject } (\text{WriteArray } l f d \text{ return }())
\]

3.1.2 Run Instantiation

The interpretation of an array operation is dispatched by the following code.

\[ \text{runArrays} :: \text{FoFConst} (\text{Heap} \to (a, \text{Heap})) \to (\text{Heap} \to (a, \text{Heap})) \]
\[
\text{runArrays} (\text{NewArray } a b c r) \text{ heap } = \text{uncurry } r \quad \text{runNewArray } b c \text{ heap } \\
\text{runArrays} (\text{ReadArray } a b r) \text{ heap } = \text{uncurry } r \quad \text{runReadArray } a b \text{ heap } \\
\text{runArrays} (\text{WriteArray } a b c r) \text{ heap } = r \quad \text{runWriteArray } a b c \text{ heap }
\]

Creating, reading, and writing to or from an array are trivially implemented by the following code:

\[ \text{runNewArray} :: \text{AllocArray} \to [\text{Data}] \to \text{Heap} \to (\text{Loc}, \text{Heap}) \]
\[
\text{runNewArray } alloc \text{ initData } \text{ heap } = \\
\text{let } \text{loc } = \text{freshALoc } \text{heap } \text{in} \\
\text{let } \text{sizeInt } = \text{length } \text{initData } \text{in} \\
\text{let } \text{name } = \text{makeVarName } \text{Dynamic } \text{loc } \text{in} \\
\text{let } \text{ref } = \text{CLRef } \text{Dynamic} (\text{ToArray } alloc \text{ typeof } \text{head } \text{initData}) \text{ name } \text{in} \\
\text{let } \text{heap1 } = \text{heap } \{ \text{freshALoc } = \text{loc } + 1, \\
\text{arrayMap } = (\text{name }, \text{initData }): (\text{arrayMap } \text{heap}) \} \text{ in} \\
(\text{ref }, \text{heap1})
\]

\[ \text{runReadArray} :: \text{Loc} \to \text{Index} \to \text{Heap} \to (\text{Data}, \text{Heap}) \]
\[
\text{runReadArray } (\text{CLRef } (\text{ToArray } \_)) \text{ loc index } \text{heap } = \\
\text{let } \text{array } = \text{fromJust } \text{loc 'lookup' (arrayMap } \text{heap}) \text{ in} \\
\text{let } (\text{CLInteger } \_ \text{indexInt}) = \text{symbEval } \text{index } \text{in} \\
\text{let } \text{val } = \text{array }!! (\text{fromInteger } \text{indexInt}) \text{ in} \\
(\text{val }, \text{heap})
\]

\[ \text{runWriteArray} :: \text{Loc} \to \text{Index} \to \text{Data} \to \text{Heap} \to \text{Heap} \]
\[
\text{runWriteArray } (\text{CLRef } (\text{ToArray } \_)) \text{ loc dat } \text{ heap } = \\
\text{let } \text{array } = \text{fromJust } \text{loc 'lookup' (arrayMap } \text{heap}) \text{ in} \\
\text{let } (\text{CLInteger } \_ \text{indexInt}) = \text{symbEval } \text{index } \text{in} \\
\text{let } (\text{arrayBegin }, \text{arrayEnd }) = \text{splitAt } (\text{fromInteger } \text{indexInt}) \text{ array } \text{in} \\
\text{let } \text{array1 } = \text{arrayBegin }++ (\text{dat } : \text{tail } \text{arrayEnd }) \text{ in} \\
\text{let } \text{heap1 } = \text{heap } \{ \text{arrayMap } = (\text{loc }, \text{array1 }): \text{arrayMap } \text{heap} \} \text{ in} \\
\text{heap1}
\]

3.1.3 Compile Instantiation

Similarly, the compilation of array operations consists in implementing the following function:

\[ \text{compileArrays} :: \text{FoFConst} (\text{Binding} \to (\text{ILFoF}, \text{Binding})) \to (\text{Binding} \to (\text{ILFoF}, \text{Binding})) \]
\[
\text{The translation from the } \text{FoFConst } \text{terms to } \text{FoF } \text{terms is almost automatic. The added value of this } \text{process consists in generating or deriving names for the references.}
\]
3.2 Conditionals

The Conditionals constructs consist of all control-flow operators defined in the C language, excepted the goto statement and fall-through switches.

3.2.1 Smart Constructors

We provide the DSL designer with all standard C control-flow operators. Hence, we define the following combinators: ifc, for, while, doWhile, break, and continue.

ifc :: FoFCode PureExpr →
    FoFCode PureExpr →
    FoFCode PureExpr →
    FoFCode PureExpr
ifc cond ifTrue ifFalse =
    inject (If cond ifTrue ifFalse (return Void))
for :: FoFCode PureExpr →
    FoFCode PureExpr →
    FoFCode PureExpr →
    FoFCode PureExpr →
    FoFCode PureExpr
for init cond incr loop =
  inject (For init cond incr loop (return Void))
while :: FoFCode PureExpr →
  FoFCode PureExpr →
  FoFCode PureExpr
while cond loop =
  inject (While cond loop (return Void))
doWhile :: FoFCode PureExpr →
  FoFCode PureExpr →
  FoFCode PureExpr
doWhile loop cond =
  inject (DoWhile loop cond (return Void))
break :: FoFCode PureExpr
break = inject Break
continue :: FoFCode PureExpr
continue = inject Continue

The \texttt{switch} statement is slightly different from the C one: every case is automatically terminated by a \texttt{break} statement. Hence, it is impossible to \texttt{fall through} a case.

\begin{verbatim}
switch :: PureExpr →
  [(PureExpr, FoFCode PureExpr)] →
  FoFCode PureExpr →
  FoFCode PureExpr
switch cond cases defaultCase =
  inject (Switch cond cases defaultCase (return Void))
\end{verbatim}

\subsection{Compile Instantiation}

The compilation step is mostly standard. Note that we often have to compile sub-blocks of code. Therefore, we need to carefully update the relevant binding states, so as to ensure the freshness of generated names while respecting the scope of locally defined variables.

\begin{verbatim}
compileConditionals (If condi ifTrue ifFalse r) binding =
  (FIf compCond compIfTrue compIfFalse cont,
   binding2)
  where (compCond, binding1) = compileSemtoFoF' condi binding
       (compIfTrue, binding1') = compileSemtoFoF' ifTrue binding1
       (compIfFalse, binding1'') = compileSemtoFoF' ifFalse
       (binding1' |-> binding1)
       (cont, binding2) = r (binding1'' |-> binding)

compileConditionals (While condW loop r) binding =
  (FWhile compCond compLoop cont,
   binding3)
  where (compCond, binding1) = compileSemtoFoF' condW binding
       (compLoop, binding2) = compileSemtoFoF' loop binding1
       (cont, binding3) = r (binding2 |-> binding)

compileConditionals (DoWhile loop condD r) binding =
  (FDAWhile compLoop compCond cont,
   binding3)
  where (compLoop, binding1) = compileSemtoFoF' loop binding
       (compCond, binding2) = compileSemtoFoF' condD
       (binding1 |-> binding)
\end{verbatim}
(cont, binding3) = r (binding2 |-> binding)

compileConditionals (For init test inc loop r) binding =
(FFor compInit compTest compInc compLoop cont, binding5)
  where (compInit, binding1) = compileSemtoFoF' init binding
        (compTest, binding2) = compileSemtoFoF' test binding1
        (compInc, binding3) = compileSemtoFoF' inc binding2
        (compLoop, binding4) = compileSemtoFoF' loop
          (binding1 |-> binding3)
        (cont, binding5) = r (binding4 |-> binding)

compileConditionals (Switch test cases defaultC r) binding =
(FSwitch test compCases compDefault cont, binding3)
  where compileCase (compCodes, binding) (i, code) =
    ((i, compCode) : compCodes,
     (binding1 |-> binding))
  where (compCode, binding1) = compileSemtoFoF' code binding
        (compCases, binding1) =
          foldl' compileCase ([], binding) cases
        (compDefault, binding2) =
          compileSemtoFoF' defaultC (binding1 |-> binding)
        (cont, binding3) = r (binding2 |-> binding)

compileConditionals Break binding =
(FClosing $ FBreak, binding)

compileConditionals Continue binding =
(FClosing $ FContinue, binding)

3.2.3 Run Instantiation

The implementation of the interpreter is straightforward. We start by dispatching calls to construct-specific functions:

runConditionals (If a b c r) heap =
  r $ runIf a b c heap
runConditionals (For a b c d r) heap =
  r $ runFor a b c d heap
runConditionals (While a b r) heap =
  r $ runWhile a b heap
runConditionals (DoWhile a b r) heap =
  r $ runDoWhile a b heap
runConditionals (Switch a b c r) heap =
  r $ runSwitch a b c heap
runConditionals Break heap =
  error "runAlgebra: Break not yet implemented"
runConditionals Continue heap =
  error "runAlgebra: Continue not yet implemented"

Then, we implement the semantics of each of these constructs:

runIf :: FoFCode PureExpr ->
FoFCode PureExpr ->
FoFCode PureExpr ->
Heap -> Heap
runIf test ifTrue ifFalse heap =

Barrelfish TN-024
let (vtest, heap1) = run test heap in
let CLInteger _ _ valVtest = symbEval vtest in
if (valVtest ≠ 0) then
    let (_, heap2) = run ifTrue heap1 in
else
    let (_, heap2) = run ifFalse heap1 in
heap2

runFor :: FofCode PureExpr →
    FofCode PureExpr →
    FofCode PureExpr →
    FofCode PureExpr →
    Heap → Heap
runFor init test incr loop heap =
    let (_, heap1) = run init heap in
loopWhile heap1
    where loopWhile heap =
        let (vtest, heap1) = run test heap in
        let CLInteger _ _ valVtest = symbEval vtest in
        if (valVtest ≠ 0) then
            let (_, heap2) = run loop heap1 in
        else
            heap1
        loopWhile heap3
runWhile :: FofCode PureExpr →
    FofCode PureExpr →
    Heap → Heap
runWhile test loop heap =
    let (vtest, heap1) = run test heap in
    let CLInteger _ _ valVtest = symbEval vtest in
    if (valVtest ≠ 0) then
        let (_, heap2) = run loop heap1 in
runWhile test loop heap2
else heap1
runDoWhile :: FofCode PureExpr →
    FofCode PureExpr →
    Heap → Heap
runDoWhile loop test heap =
    let (_, heap1) = run loop heap in
    let (vtest, heap2) = run test heap1 in
    let CLInteger _ _ valVtest = symbEval vtest in
    if (valVtest ≠ 0) then
        runDoWhile loop test heap2
else
    heap2
runSwitch :: PureExpr →
    [(PureExpr, FofCode PureExpr)] →
    FofCode PureExpr →
    Heap → Heap
runSwitch test cases defaultCase heap =
    let res = symbEval test in
    case res 'lookup' cases of
        Just stmt → let (_, heap1) = run stmt heap in
            heap1
        Nothing → let (_, heap1) = run defaultCase heap in
            heap1
3.3 Enumeration

The Enumeration construct mirrors the enum data-type of C. It allows us to name a finite number of natural constants and manipulate these names instead of numbers.

3.3.1 Smart Constructors

The newEnum combinator is used to create a member value belonging to one of the fields of nameEnum.

\[
\text{newEnum} :: \text{String} \rightarrow \text{Enumeration} \rightarrow \text{String} \rightarrow \text{FoFCode PureExpr} \\
\text{newEnum nameEnum fields value} = \\
\text{inject (NewEnum Nothing nameEnum fields value return)}
\]

Similarly, newEnumN creates a named member of an enumeration.

\[
\text{newEnumN} :: \text{String} \rightarrow \text{String} \rightarrow \text{Enumeration} \rightarrow \text{String} \rightarrow \text{FoFCode PureExpr} \\
\text{newEnumN name nameEnum fields value} = \\
\text{inject (NewEnum (Just name) name fields value return)}
\]

3.3.2 Compile Instantiation

A NewEnum is compiled as follow.

\[
\text{compileEnumerations} (\text{NewEnum name enumName vals value r}) \text{ binding} = \\
(\text{FStatement (FNewEnum publicName enumName vals value) cont, binding3)} \\
\text{ where (publicName, binding2) = case name of} \\
\text{ Just x \rightarrow (Provided x, binding)} \\
\text{ Nothing \rightarrow (makeVarName Local loc, binding1)} \\
\text{ where (loc, binding1) = getFreshVar binding} \\
\text{ ret = CLRef Global int64T (Provided value)} \\
\text{ (cont, binding3) = r ret binding2}
\]

Note that ret is actually the name of the enumerated value: it is treated as a constant and passed as such to the remaining code. A more standard implementation would have been to create a variable containing this constant value and pass the reference to the variable to the subsequent code. However, when switch-ing over an enumerated value, the case would match a variable instead of a constant, which is refused by the C compiler.

Clearly, a clean solution to this implementation must be found. However, the current solution, if not perfect, seems to be good enough.
### 3.3.3 Run Instantiation

Running a `newEnum` simply consists in getting the associated value.

```haskell
runEnumerations (NewEnum _ _ enum name r) heap =
  let ref = uint64 $ toInteger $ fromJust $ name `lookup` enum in
    r ref heap
```

### 3.4 Function Definition

This module abstracts the function definition and manipulation mechanisms found in C. This consists in a `def` constructor, to define functions, a `call` and `callN` functions to call functions, as well as a `returnc` combinator to return from a function call.

#### 3.4.1 Smart Constructors

When defining a function, we provide a list of attributes, its name, its body, its return type, and a list of arguments types:

```haskell
def :: [FunAttr] →
  String →
  ([PureExpr] → FoFCode PureExpr) →
  TypeExpr →
  [(TypeExpr, Maybe String)] →
  FoFCode PureExpr

def attr name fun returnT argsT =
  inject (NewDef attr name (Fun fun) returnT argsT return)
```

Then, it is possible to call into a function, provided a list of parameters. The result, if any, can be named by using the `callN` construct.

Currently, both the interpreter and the compiler are extremely optimistic about their inputs: in the future, we should add more safety checks. For example, we should check that we are calling the functions with the right arguments.

```haskell
call :: PureExpr → [PureExpr] → FoFCode PureExpr
call funRef params =
  inject (CallDef Nothing funRef params return)
callN :: String → PureExpr → [PureExpr] → FoFCode PureExpr
callN varName funRef params =
  inject (CallDef (Just varName) funRef params return)
```

Finally, it is possible to return from a function thanks to the usual `return`. This should not be confused with the monadic `return` of Haskell.

```haskell
returnc :: PureExpr → FoFCode PureExpr
returnc value = inject (Return value)
```
3.4.2 Compile Instantiation

Compiling functions is a little bit more tricky than usual. It requires generating or handling arguments, as well as handling the return value, if any. This corresponds to the following code.

```haskell
compileFunctions (NewDef attr nameF (Fun func) return args r)
  binding =
  (FNewDef attr nameF compBody return instanceArgs cont, binding2)
  where instanceArgs = instanciateArgs args
        (compBody, binding1) = compileSemtoFoF' (func instanceArgs) binding
        ref = CLRRef Global (TFun nameF (Fun func) return args) (Provided nameF)
        (cont, binding2) = r ref (binding1 |-> binding)
    instanciateArgs :: [(TypeExpr, Maybe String)] -> [PureExpr]
    instanciateArgs params = reverse $ foldl' instanciateArg [] $ zip [1..] params
    where instanciateArg l (idx, (typ, mName)) = (CLRRef Param typ name) : l
          where name = case mName of
                     Just x -> Provided x
                     Nothing -> makeVarName Param idx

compileFunctions (CallDef mName f@(CLR _ (TFun nameF func returnT argsT)) args r) binding =
  (FStatement (FCallDef name f args) cont, binding2)
  where (name, binding1) = case returnT of
         TVoid -> (Nothing, binding)
         _    -> case mName of
                 Just x -> (Just $ Provided x, binding)
                 Nothing ->
                 (Just $ makeVarName Local loc, binding')
                 where (loc, binding') = getFreshVar binding
          (cont, binding2) = case returnT of
         TVoid -> r Void binding1
         _    -> r (CLRRef Local returnT (fromJust name)) binding1
```

The translation of the return statement, on the other hand, is trivial.

```haskell
compileFunctions (Return e) binding =
  (FClosing $ FReturn e, binding)
```

3.4.3 Run Instantiation

As usual, we dispatch here:
runFunctions (NewDef _ _ f _ _ r) heap =  
  uncurry r $ runNewDef f heap 
runFunctions (CallDef _ a b r) heap =  
  uncurry r $ runCallDef a b heap 
runFunctions (Return a) heap =  
  runReturn a heap  -- OK??

And compute there:

runReturn :: PureExpr -> Heap -> (PureExpr, Heap) 
runReturn e heap = (e, heap) 
runNewDef :: Function -> Heap -> (PureExpr, Heap) 
runNewDef function heap =  
  (CLRef Global (TFun ⊥ function ⊥⊥ ⊥⊥ ⊥), heap) 
runCallDef :: PureExpr -> [PureExpr] -> Heap ->  
  (PureExpr, Heap) 
runCallDef (CLRef _ (TFun _ (Fun function _) _ _) _) args heap =  
  let (result, heap1) = run (function args) heap in  
    (result, heap1)

3.5 Reference Cells

The reference cell construct provides an abstraction to both variables and C pointers. It composed by three combinators to create, read from, and write to reference cells. It can be compared to OCaml references or Haskell IORef.

3.5.1 Smart Constructors

A reference cell is created in an initialized state. The variant newRefN allows the DSL designer to provide a name to the created variable.

newRef :: Data -> FoFCode Loc  
newRef d = inject (NewRef Nothing d return) 
newRefN :: String -> Data -> FoFCode Loc  
newRefN name d = inject (NewRef (Just name) d return)

Follow primitives to read from and write to these reference cells:

readRef :: Loc -> FoFCode Data  
readRef l = inject (ReadRef l return) 
writeRef :: Loc -> Data -> FoFCode PureExpr  
writeRef l d = inject (WriteRef l d (return Void))

The current implementation lacks lots of sanity checks:

- read and Write on CLRef,
- write from and to compatible types,
- do not write local pointers into param/global ones,
- ...

FiletOFish - 36 Barrellfish TN-024
3.5.2 Compile Instantiation

The compilation is tricky when it comes to computing the pointer type. I wouldn’t be surprised if some bugs were lying there. This concerns `newRef` and `readRef`, which effect on references is not trivial.

```
compileReferences (NewRef refName ref r) binding =
  (FStatement (FNewRef publicName ref) cont, binding2)
  where (publicName, binding1) = case refName of
    Just x -> (Provided x, binding)
    Nothing ->
      let (loc, binding1) = getFreshVar binding in
      (makeVarName Local loc, binding1)
  ret = CLRef Local (TPointer (typeOf ref) Avail) publicName
  (cont, binding2) = r ret binding1
```

```
compileReferences (ReadRef ref @(CLRef _ zloc) r) binding =
  (FStatement (FReadRef name ref) cont, binding2)
  where (loc, name, binding1) = heritVarName binding zloc
    ret = CLRef Local (unfoldPtrType ref) name
  (cont, binding2) = r ret binding1
```

writeRef is straightforward.

```
compileReferences (WriteRef ref d r) binding =
  (FStatement (FWriteRef ref d) cont, binding1)
  where (cont, binding1) = r binding
```

3.5.3 Run Instantiation

On the other hand, the implementation of the interpreter is much simpler. We start with the dispatcher:

```
runReferences (NewRef d r) heap = uncurry r $ runNewRef d heap
runReferences (ReadRef l r) heap = uncurry r $ runReadRef l heap
runReferences (WriteRef l v r) heap = r $ runWriteRef l v heap
```

And the per-construct interpreters follow:

```
runNewRef :: Data -> Heap -> (Loc, Heap)
runNewRef value heap =
  (CLRef Local typeOfVal name, heap2)
  where typeOfVal = typeOf value
    loc = freshLoc heap
    refs = refMap heap
    name = makeVarName Local loc
    heap1 = heap {freshLoc = loc + 1}
    heap2 = heap1 {refMap = (name, value) : refs}
runReadRef :: Loc -> Heap -> (Data, Heap)
runReadRef (CLRef _ location) heap =
  let refs = refMap heap in
  let val = fromJust $ location `lookup` refs in
  (val, heap)
runWriteRef :: Loc -> Data -> Heap -> Heap
```
\begin{verbatim}
runWriteRef (CLRef _ _ location) value heap =
  let refs = refMap heap in
  let refs1 = (location, value) : refs in
  heap { refMap = refs1 }
\end{verbatim}

3.6 Strings

The \textit{String} construct corresponds to static arrays of characters. However, they are implemented here as a special case as they are specially dealt with by the C compiler.

3.6.1 Smart Constructors

We only provide string creation combinators: accessing a string can be achieved thanks to \textit{Arrays} combinators. As usual, we provide two combinators: one to create an anonymous string, one to create a named string.

\begin{verbatim}
newString :: String \rightarrow FoFCode Loc
newString value = inject (NewString Nothing value return)
newStringN :: String \rightarrow String \rightarrow FoFCode Loc
newStringN name value = inject (NewString (Just name) value return)
\end{verbatim}

3.6.2 Compile Instantiation

The compilation is straightforward, on the model of static array declaration.

\begin{verbatim}
compileString (NewString name dat r) binding =
  let (publicName, binding1)
    = case name of
      Just x -> (Provided x, binding)
      Nothing ->
        let (loc, binding1) = getFreshVar binding in
        (makeVarName Global loc,
         binding1) in
  let ret = CLRef Global
            (TArray (StaticArray $ length dat) TChar)
            publicName in
  let (cont, binding2) = r ret binding1 in
  (FStatement (FNewString publicName dat) cont,
               binding2)
\end{verbatim}

3.6.3 Run Instantiation

Similarly, the interpreter is simple.

\begin{verbatim}
runString (NewString a b r) heap = uncurry r $ runNewString b heap
\end{verbatim}

\begin{verbatim}
runNewString :: String \rightarrow Heap \rightarrow (Loc, Heap)
runNewString string heap =
\end{verbatim}
let loc = freshALoc heap in
let size = length string in
let name = makeVarName Dynamic loc in
let ref = CLRef Dynamic (TArray (StaticArray size) TChar) name in
let heap1 = heap { freshALoc = loc + 1,
arrayMap = (name, map cchar string) : (arrayMap heap)} in
(ref, heap1)

3.7 Structures Definition

The Structure construct allows you to mirror the struct data-type of C. It is composed by a newStruct combinator, to instantiate an element of this type, a readStruct combinator, to read a field from a structure, and a writeStruct combinator, to write into a field.

3.7.1 Smart Constructors

As often with instantiation operators, we can chose between statically or dynamically allocating the value. Then, it is possible to chose between an anonymous or a named value. All these choices are provided by the following four combinators.

newStaticStruct :: String → ((TypeExpr, String, Data)] → FoFCode Loc
newStaticStruct name stt =
    inject (NewStruct Nothing StaticStruct name
           (map (λ(t,n,v) → (n,(t,v))) stt)
        return)

newStaticStructN :: String → String → ((TypeExpr, String, Data)] → FoFCode Loc
newStaticStructN nameStr name stt =
    inject (NewStruct (Just nameStr) StaticStruct name
           (map (λ(t,n,v) → (n,(t,v))) stt)
        return)

newStruct :: String → ((TypeExpr, String, Data)] → FoFCode Loc
newStruct name stt =
    inject (NewStruct Nothing DynamicStruct name
           (map (λ(t,n,v) → (n,(t,v))) stt)
        return)

newStructN :: String → String → ((TypeExpr, String, Data)] → FoFCode Loc
newStructN nameStr name stt =
    inject (NewStruct (Just nameStr) DynamicStruct name
           (map (λ(t,n,v) → (n,(t,v))) stt)
        return)
Follow the read and write combinators:

```haskell
readStruct :: Loc → String → FoFCode Data
readStruct l f = inject (ReadStruct l f return)
writeStruct :: Loc → String → Data → FoFCode ()
writeStruct l f d = inject (WriteStruct l f d (return ()))
```

### 3.7.2 Compile Instantiation

Apart from type handling, the compilation naturally follows the definition. As often, computing the `CLRef` is a magic voodoo, which is far from being provably correct.

```haskell
compileStructures (NewStruct refName allocStruct name fields r) binding =
    (FStatement newS cont, binding2)
  where (loc, binding1) = getFreshVar binding
    structName = case refName of
      Just x → Provided x
      Nothing → makeVarName Dynamic loc
    fieldsTypeStr = [(field, typ)
        | (field, (typ, _)) ← fields]
    typeStr = TStruct DynamicStruct name fieldsTypeStr
    ret = CLRef Dynamic typeStr structName
    (cont, binding2) = r ret binding1
    newS = FNewStruct structName allocStruct name fields

compileStructures (ReadStruct ref @(CLRef origin
typ @(TStruct alloc name fields)
  zloc)
field r) binding =
    (FStatement readS cont, binding2)
  where (loc, varName, binding1) = heritVarName binding zloc
    typeField = fromJust $ field 'lookup' fields
    ret = CLRef (allocToOrigin alloc) (readOf typeField) varName
    (cont, binding2) = r ret binding1
    readS = FReadStruct varName ref field
    allocToOrigin StaticStruct = Local
    allocToOrigin DynamicStruct = Dynamic

compileStructures (WriteStruct ref @(CLRef origin
typ @(TStruct alloc name fields)
  zloc)
field
value r) binding =
    (FStatement writeS cont, binding1)
  where (cont, binding1) = r binding
    writeS = FWriteStruct ref field value
```

### 3.7.3 Run Instantiation

The interpreter follows with a dispatcher:
runStructures (NewStruct a b c r) heap =
  uncurry r $ runNewStruct a b c heap
runStructures (ReadStruct a b r) heap =
  uncurry r $ runReadStruct a b heap
runStructures (WriteStruct a b c r) heap =
  r $ runWriteStruct a b c heap

And the per-construct implementation:

runNewStruct :: AllocStruct → String → [(String, (TypeExpr, Data))] →
  Heap → (Loc, Heap)
runNewStruct alloc name struct heap =
  let
    structT = map (λ(x1, (x2, _)) → (x1, x2)) struct in
    structD = map (λ(x1, (_, x2)) → (x1, x2)) struct in
    loc = freshLoc heap in
    structs = strMap heap in
    varName = makeVarName Local loc in
    heap1 = heap { freshLoc = loc + 1 } in
    heap2 = heap1 { strMap = (varName, structD) : structs } in
    (CLRef Local (TStruct alloc name structT) varName, heap2)

runReadStruct :: Loc → String → Heap → (Data, Heap)
runReadStruct (CLRef _ location) field heap =
  let
    structs = strMap heap in
    struct = fromJust $ location ‘lookup’ structs in
    val = fromJust $ field ‘lookup’ struct in
    (val, heap)

runWriteStruct :: Loc → String → Data → Heap → Heap
runWriteStruct (CLRef _ location) field value heap =
  let
    structs = strMap heap in
    struct = fromJust $ location ‘lookup’ structs in
    struct1 = (field, value) : struct in
    structs1 = (location, struct1) : structs in
    heap { strMap = structs1 }

3.8 Type Definition

The `Typedef` construct provides a similar service than the C `typedef`.

3.8.1 Smart Constructors

In particular, `Typedef` offers two combinators. The first one, `alias` allows you to locally define a type alias.

```
alias :: TypeExpr → FoFCCode PureExpr
alias typedef = inject (Typedef typedef (return void))
```

The other one, `aliasE` allows you to mention an aliasing declared in an external library, such as `<stdbool.h>` that declares a `bool` as an integer.

```
aliasE :: String → TypeExpr →
```
3.8.2 Compile Instantiation

The compilation to FoF is straightforward:

```plaintext
compileTypedef (Typedef (TTypedef typ aliasName) r) binding =
let (cont, binding1) = r binding in
(FStatement (FTypedef typ aliasName) cont, binding1)
```

```plaintext
compileTypedef (TypedefE inclDirective typeDef @(TTypedef typ aliasName) r) binding =
let (cont, binding1) = r binding in
(FStatement (FTypedefE inclDirective typeDef) cont, binding1)
```

3.8.3 Run Instantiation

These operations occurring at the type-level, the interpreter doesn’t pay any attention to them:

```plaintext
runTypedef (Typedef _ r) heap = r heap
runTypedef (TypedefE _ _ r) heap = r heap
```

3.9 Unions Definition

The Union constructs abstracts the union data-type of C.

3.9.1 Smart Constructors

Hence, creating an union is available in four flavors, statically or dynamically allocated, and anonymous or named.

```plaintext
newStaticUnion :: String →
[((TypeExpr, String)] →
String →
Data →
FoFCode Loc
newStaticUnion name fields field dat =
inject (NewUnion Nothing StaticUnion name
(map (λ(s1, s2) → (s2, s1)) fields)
(field, dat)
return)
newStaticUnionN :: String →
String →
[((TypeExpr, String)] →
String →
Data →
FoFCode Loc
```
newStaticUnionN nameU name fields field dat =
  inject (NewUnion (Just nameU) StaticUnion name
        (map (λ(s1, s2) → (s2, s1)) fields)
        (field, dat)
        return)

newUnion :: String →
  [(TypeExpr, String)] →
  String →
  Data →
  FoFCode Loc
newUnionN name fields field dat =
  inject (NewUnion Nothing DynamicUnion
        name
        (map (λ(s1, s2) → (s2, s1)) fields)
        (field, dat)
        return)

newUnionN :: String →
  [(TypeExpr, String)] →
  String →
  Data →
  FoFCode Loc
newUnionN nameU name fields field dat =
  inject (NewUnion (Just nameU) DynamicUnion
        name
        (map (λ(s1, s2) → (s2, s1)) fields)
        (field, dat)
        return)

Reading and writing follow the usual scheme:

readUnion :: Loc → String → FoFCode Data
readUnion l f = inject (ReadUnion l f return)

writeUnion :: Loc → String → Data → FoFCode ()
writeUnion l f d = inject (WriteUnion l f d (return ()))

3.9.2 Compile Instantiation

As usual the difficulty of the compilation stands in not messing up created and read types. Apart from
that, it is a simple translation.

compileUnions (NewUnion refName allocUnion nameU fields (initField, initData) r) binding =
  (FStatement newU cont,
   binding2)
  where typeUnion = TUnion DynamicUnion nameU fields
          (loc, binding1) = getFreshVar binding
          name = case refName of
            Nothing → makeVarName Dynamic loc
            Just x → Provided x
          ret = CLRef Dynamic typeUnion name
          (cont, binding2) = r ret binding1
          newU = FNewUnion name allocUnion nameU fields (initField, initData)
  compileUnions (ReadUnion ref @(CLRef _ typeU @(TUnion alloc

Barrefish TN-024

43
nameU fields) xloc
field r) binding =
(FStatement readU cont,
  binding2)
  where (loc, name, binding1) = heritVarName binding xloc
  typeField = fromJust $ field 'lookup' fields
  origin = allocToOrigin alloc
  ret = CLRef origin (readOf typeField) name
  (cont, binding2) = r ret binding1
  readU = FReadUnion name ref field
  allocToOrigin StaticUnion = Local
  allocToOrigin DynamicUnion = Dynamic
compileUnions (WriteUnion ref @((CLRef origin
  typ @((TUnion alloc _ fields)
  xloc)
  field
  value r) binding =
  (FStatement writeU cont,
    binding1)
    where (cont, binding1) = r binding
    writeU = FWriteUnion ref field value

3.9.3 Run Instantiation

This part has not been implemented yet. Hence, the interpreter will blow up in presence of unions. To
get an idea of the desired implementation, take a look at the reference cells interpreter. It should be
similarly easy.

runUnions (NewUnion _ a b c d r) heap = error "runUnions: not yet implemented"
runUnions (ReadUnion a b r) heap = error "runUnions: not yet implemented"
runUnions (WriteUnion a b c r) heap = error "runUnions: not yet implemented"
Chapter 4

Lib-C Operators

Mortician: Bring out your dead! [clang] …
Customer: Here’s one – nine pence.
Dead person: I’m not dead!
Mortician: What?
Customer: Nothing – here’s your nine pence.

Monty Python

4.1 Printf

The *Printf* constructs is a simple foreign function wrapper around the C library `printf`.

4.1.1 Smart Constructors

Provided with a format string and a list of parameters, the `printf` Pcombinator emulates `printf`.

```
printf :: String → [PureExpr] → FoFCode PureExpr
printf format params = inject (Printf format params (return Void))
```

4.1.2 Compile Instantiation

Compilation is a natural foreign function call. Note the quoting of `format`: we sacrify the semantics of the format string. We could possibly apply some tricks to recover it, or to get it in a "nice" format thanks to the `printf` combinator. However, for simplicity, we drop its semantics for now.

```
compilePrintf (Printf format params r) binding =
  let (cont, binding1) = r binding in
  (FStatement (FFFICall "printf" ((quote format) : params)) cont, binding1)
```
4.1.3 Run Instantiation

For the reason mentioned above, it is a pain to recover the semantics of the printf. Hence, we drop its side-effect when interpreting it.

\[
\text{runPrintf } (\text{Printf } abr) \text{ heap } = r \text{ heap}
\]

An esthetically satisfying solution would be to store this (and others) side-effecting operations in a stream, along with its arguments. Hence, we could compare side-effecting programs by their so-called trace. By ignoring the effect of printf here, we consider that side-effects have no semantic significance. This is kind of lie when interpreting an imperative language.

4.2 Assert

The construct Assert embeds the C assert function into FoF.

4.2.1 Smart Constructors

The use of assert is obvious, by its definition.

\[
\text{assert } :: \text{PureExpr } \rightarrow \text{FoFCode PureExpr}\\
\text{assert test } = \text{inject (Assert test (return Void))}
\]

4.2.2 Compile Instantiation

The compilation is a direct translation into a foreign function:

\[
\text{compileAssert } (\text{Assert } e r) \text{ binding } = \\
\text{let } (\text{cont, binding1}) = r \text{ binding in}\\
(\text{FStatement } (\text{FFFFICall "assert" [e] cont, binding1})
\]

4.2.3 Run Instantiation

As mentioned with Printf, we take here the easy option of ignoring the run-time behaviour of an assertion.

\[
\text{runAssert } (\text{Assert } a r) \text{ heap } = r \text{ heap}
\]

Being able to capture the semantics of that operation would be helpful when debugging a compiler. So, some efforts are worth being devoted here.
Chapter 5

Lib-barrefish Operators

Here may be found the last words of Joseph of Aramathea. He who is valiant and pure of spirit may find the Holy Grail in the Castle of Monty Python

5.1 Has Descendants

The construct HasDescendants embeds the libarrefish function has_descendants into FoF.

5.1.1 Smart Constructors

This function is provided in two flavors: an anonymous one, which stores its result in an anonymous variable, and a named one, which allows you to name the resulting variable.

\[
\begin{align*}
\text{has\_descendants} & \:: \text{PureExpr} \to \text{FoFCode PureExpr} \\
\text{has\_descendants cte} &= \text{inject (HasDescendants Nothing cte return)} \\
\text{has\_descendantsN} & \:: \text{String} \to \text{PureExpr} \to \text{FoFCode PureExpr} \\
\text{has\_descendantsN name cte} &= \text{inject (HasDescendants (Just name) cte return)}
\end{align*}
\]

5.1.2 Compile Instanciation

This function is translated into a foreign function definition, as usual:

\[
\begin{align*}
\text{compileHasDescendants (HasDescendants mName arg r) binding} &= \\
\text{let (loc, binding1) = getFreshVar binding in} \\
\text{let name = case mName of} \\
\text{Nothing \to makeVarName Local loc} \\
\text{Just x \to Provided x in} \\
\text{let ref = CLRef Local uint64T name in} \\
\text{(FStatement (FFFCall "has\_descendants" [ref, arg]) cont, binding2)}
\end{align*}
\]
5.1.3 Run Instantiation

As for libc functions, we have not yet implemented the semantics of that operation. A trace-based semantics would make sense, too.

\[
\text{runHasDescendants (HasDescendants \_ a r) heap = error "HasDescendants: eval not implemented"}
\]

5.2 Mem To Phys

This construct embeds the libarrelfish function \texttt{mem\_to\_phys} into FoF.

5.2.1 Smart Constructors

As for \texttt{HasDescendants}, both named and anonymous function are provided. They are direct wrappers around the \texttt{mem\_to\_phys} function.

\[
\text{mem\_to\_phys} :: \text{PureExpr} \rightarrow \text{FoFCode PureExpr} \\
\text{mem\_to\_phys} \text{cte} = \text{inject (MemToPhys Nothing cte return)} \\
\text{mem\_to\_physN} :: \text{String} \rightarrow \text{PureExpr} \rightarrow \text{FoFCode PureExpr} \\
\text{mem\_to\_physN} \text{name cte} = \text{inject (MemToPhys (Just name) cte return)}
\]

5.2.2 Compile Instantiation

Compiling is straightforward: just declare a foreign function.

\[
\text{compileMemToPhys (MemToPhys mName arg r) binding =} \\
\text{let (loc, binding1) = getFreshVar binding in} \\
\text{let name = case mName of} \\
\text{ Just x \rightarrow Provided x} \\
\text{ Nothing \rightarrow makeVarName Local loc in} \\
\text{let ref = CLRef Local uint64T name in} \\
\text{let (cont, binding2) = r ref binding1 in} \\
\text{(FStatement (FFFICall "mem\_to\_phys" [ref, arg]) cont, binding2)}
\]

5.2.3 Run Instantiation

However, the semantics remains to be defined.

\[
\text{runMemToPhys (MemToPhys \_ a r) heap = error "MemToPhys: eval not implemented"}
\]
Part II

The Filet-o-Fish Compiler
The Filet-O-Fish Compiler(s)

The Filet-o-Fish to C compiler is major component of Filet-o-Fish. Major in the sense that it is a big chunk of code, which correctness is critical. So, when playing with this part of the code, better be cautious. The high-level specification of the compiler is straightforward: given a Filet-o-Fish code, it should translate it into a semantically equivalent C code. Well, it is a compiler, after all.

However, from a usability point of view, it is vital to be able to understand what the generated code is doing: think of a debugging session that needs to go through some code generated by Filet-o-Fish. Hence, we have implemented some so-called optimizations that tidy up the generated code. In order to ease the implementation of these optimizations we rely on two standard compiler techniques: first, we define a bunch of intermediate languages (IL) to tackle a specific optimization issue, second we implement the optimizer as a data-flow analysis solver. The current state of affair is not as idyllic and the reader is referred to Chapter A to get an overview of my dreams.

Let us sketch the compilation process.

\[
\text{compile} :: \text{Semantics } \text{FoFConst } \text{PureExpr} \rightarrow \text{PakaCode}
\]

\[
\text{compile } \text{sem} = \\
\text{optimizePaka } \text{sem} \\
\text{compileFoFtoPaka } \text{sem} \\
\text{compileSemtoFoF } \text{sem}
\]

First of all, The compiler is provided a value of type \text{Semantics } \text{FoFConst } \text{PureExpr}, built by the operators of Chapter 3. While this structure has a nice functional definition, making it convenient for interpretation, it is bothersome to navigate on it. Therefore, the first pass of the compiler is to reify this data-structure, as explained in Chapter 6.

At the end of this compilation pass, the initial input has been translated into an (hopefully) equivalent one in the FoF intermediate language. In order to remove unnecessary variable assignments, a second pass of the compiler translate the FoF code into Paka code. In a nutshell, the Paka language only captures variable assignments, ignoring the computational parts of statements. Hence, seeking and simplifying redundant assignments is made easy: it corresponds to an optimization phase applied to the resulting Paka code.

Because different optimizations will focus on different aspects of the code, one could imagine several intermediate languages and refinements between them. FoF and Paka are just an example of what could be done. The name Paka comes from a retired hurricane: to pursue that tradition, you can look up the list of retired hurricane names [1]. There is fair amount of ILs to be implemented.
Chapter 6

The FoF Intermediate Language

- [...] For, since the tragic death of her father –
- He's not quite dead!
- Since the near fatal wounding of her father–
- He's getting better!
- For, since her own father... who, when he
seemed about to recover, suddenly felt the
icy hand of death upon him,...
- Oh, he's died!
- And I want his only daughter to look upon
me... as her own dad – in a very real, and
legally binding sense. And I feel sure that the
merger – uh, the union – between the
Princess and the brave, but dangerous, Sir
Launcelot of Camelot...

Monty Python

6.1 The FoF Intermediate Language

The FoF IL is nothing more than a direct translation of the Filet-o-Fish operators. In retrospect, calling it FoF might be confusing. Never forget that lives in the IL/ directory, so it is simply not the abbreviation for Filet-o-Fish, and that's it.

Having said that, it is also obvious that, essentially, FoF is Filet-o-Fish: it is a dumb translation of the Filet-o-Fish constructs into a data-type. Hence, an ILFoF term is the reification of the language constructs:

```data ILFoF
    = FConstant PureExpr
    | FStatement FStatement ILFoF
    | FClosing FClosing
    | FNewDef [FunAttr] String ILFoF TypeExpr [PureExpr] ILFoF
    | IF [ILFoF ILFoF ILFoF ILFoF ILFoF]
    | FFor ILFoF ILFoF ILFoF ILFoF ILFoF
    | FWhile ILFoF ILFoF ILFoF
    | FDoWhile ILFoF ILFoF ILFoF
    | FSwitch PureExpr [(PureExpr, ILFoF)] ILFoF ILFoF
```

Where an FStatement is one of the sequential statement of the Filet-o-Fish language, that is:
data FStatement =
  FNewUnion VarName AllocUnion String [(String, TypeExpr)] (String, Data)
| FReadUnion VarName Loc String
| FWriteUnion VarName Loc String Data
| FTypedef TypeExpr String
| FTypedefE String TypeExpr
| FNewStruct VarName AllocStruct String [(String, (TypeExpr, Data))] (String, Data)
| FReadStruct VarName Loc String Data
| FWriteStruct Loc String Data
| FNewString VarName String
| FNewRef VarName Data
| FReadRef VarName Loc
| FWriteRef Loc Data
| FNewEnum VarName String Enumeration String
| FNewArray VarName AllocArray [Data]
| FReadArray VarName Loc Index
| FWriteArray Loc Index Data
| FCallDef (Maybe VarName) PureExpr [PureExpr]
| FFFICall String [PureExpr]

And an FClosing is a standard C end of something statement:

data FClosing =
  FReturn PureExpr
| FBreak
| FContinue

6.2 Translating FoFCode to IL.FoF

6.2.1 The compiler

We already know how to translate individual statements of the FoFCode language, by using the one
step compiler compileAlgebra defined in ./Expressions.lhs and provided a Binding capturing the
state of the compiler. The game is then to chain up these compilation steps into a single one. Here,
foldSemantics nicely comes to the rescue and automatically build this compiler.

    compileSemtoFoF :: FoFCode PureExpr → Binding → (ILFoF, Binding)
    compileSemtoFoF = foldSemantics compilePure compileAlgebra

Where compilePure is used to compile pure expressions. Pure expressions are, by definition, constants
and returned as such. This is used when generating tests for conditional expressions: the computational
part is generated above the test handler and only the (pure) result is tested.

    compilePure :: PureExpr → Binding → (ILFoF, Binding)
    compilePure x binding = (FConstant x, binding)

For our convenience, we can define the following compileSemToFoF function that takes a closed FoFCode
and compiles it in the empty environment: that’s our compiler for self-contained expressions.

    compileSemtoFoF :: FoFCode PureExpr → ILFoF
    compileSemtoFoF term = fst $ compileSemtoFoF' term emptyBinding

where emptyBinding = Binding { freshVar = 1, defStructs = [], defUnions = [], def Enums = []}
6.2.2 The machinery

Manipulating the compiler environment

We very often need to generate fresh names, while keeping the freshness invariant of the compiler environment. The following function just does that:

\[\text{getFreshVar} :: \text{Binding} \to (\text{Int}, \text{Binding})\]
\[\text{getFreshVar} \ \text{binding} = (\text{loc}, \text{binding1})\]
\[\text{where}\  \text{loc} = \text{freshVar} \ \text{binding}\]
\[\text{binding1} = \text{binding} \{ \text{freshVar} = \text{loc} + 1 \}\]

Note that a clever implementation would be something of type:

\[\text{better\_getFreshVar} :: \text{Binding} \to (\text{Int} \to \text{Binding} \to t) \to t\]
\[\text{better\_getFreshVar} \ \text{binding} \ \text{f} = \perp\]

Which enforces the fact that the function \( f \) is provided a synchronized compiler state. This ensures that people don’t inadvertently mess up the compiler state. This remark holds for too many functions below, I’m a bit sad about that.

In order to ensure the freshness of names across bindings, we define the following \text{passFreshVar} function that builds a \text{stableBinding} whose fresh variables are ensured not to clash with the one generated using \text{upBinding}. Similarly, it carries the structures defined in \text{upBinding}.

\[\text{passFreshVar} :: \text{Binding} \to \text{Binding} \to \text{Binding}\]
\[\text{passFreshVar} \ \text{upBinding} \ \text{stableBinding} =\]
\[\text{stableBinding} \{ \text{freshVar} = \text{freshVar} \ \text{upBinding},\]
\[\text{defStructs} = \text{defStructs} \ \text{upBinding},\]
\[\text{defUnions} = \text{defUnions} \ \text{upBinding},\]
\[\text{defEnums} = \text{defEnums} \ \text{upBinding}\}\]
\[(|-|) = \text{passFreshVar}\]

From variable identifier and an origin, we can later make \text{VarName}. In a craze of Hungarian naming, the origin dictates the name of variables.

\[\text{makeVarName} :: \text{Origin} \to \text{Int} \to \text{VarName}\]
\[\text{makeVarName} \ \text{orig} \ \text{loc} = \text{Generated} \ \$ \ \text{makeVarName'} \ \text{orig} \ \text{loc}\]
\[\text{where}\ \text{makeVarName'} \ \text{Local} \ x = "\text{fof}_x" + \text{show} \ x\]
\[\text{makeVarName'} \ \text{Param} \ x = "\text{fof}_y" + \text{show} \ x\]
\[\text{makeVarName'} \ \text{Dynamic} \ x = "\text{fof}_d" + \text{show} \ x\]
\[\text{makeVarName'} \ \text{Global} \ x = "\text{fof}_g" + \text{show} \ x\]

The Hungarian fever can go further: when a variable is somehow related to another \text{VarName}, the \text{heritVarName} makes it explicit at the name level by deriving a fresh name from the previous one.

\[\text{heritVarName} :: \text{Binding} \to \text{VarName} \to (\text{Int}, \text{VarName}, \text{Binding})\]
\[\text{heritVarName} \ \text{binding} \ \text{name} = (\text{loc}, \text{Inherited} \ \text{loc} \ \text{name}, \ \text{binding1})\]
\[\text{where} \ (\text{loc}, \ \text{binding1}) = \text{getFreshVar} \ \text{binding}\]

From Expressions to Types

Let us be honest: the code which follows is tricky. Change something there and the generated code will be wrong, if it is not already. I’m looking at you \text{readOf} and \text{liftType}. They came to life during the implementation of References and its painful compiler. After a lot of work, I came to the conclusion (and
proof) that they are correct. The question is now: are they correct when mixed with complex data-types, such as structs and arrays. The practician seems to say "yes", the theoretician remains proofless.

The intrinsic difficulty is that a Reference abstracts both a C variable and a C pointer. However, in C, both concepts are quite distinct. Hence, the compiler needs to be clever to translate the unified notion of Reference in two semantically different objects. Hence that horrible machinery.

\[\text{typeOf: }\] Obviously, there exists a map going from each well-typed element of PureExpr to an element of TypeExpr. Hence, this map assigns a type to a given, well-typed expression. As for ill-typed expressions, we simply return an error message.

Computing the type of base values as well as of unary operations is straightforward:

\[
\begin{align*}
\text{typeOf} :: & \text{PureExpr} \rightarrow \text{TypeExpr} \\
\text{typeOf} (\text{Void}) &= T\text{Void} \\
\text{typeOf} (\text{CLInteger} \text{ sign size } \_ ) &= T\text{Int} \text{ sign size} \\
\text{typeOf} (\text{CLFloat} \_ ) &= T\text{Float} \\
\text{typeOf} (\text{CLRef} \_ \text{typ} \_ ) &= \text{typ} \\
\text{typeOf} (\text{Unary} \_ x ) &= \text{typeOf} x
\end{align*}
\]

A binary operation is well-typed if and only if both sub-terms are well-typed and of same type. The same goes for the branches of a conditional expression:

\[
\begin{align*}
\text{typeOf} (\text{Binary} \_ x y ) &= \\
& \text{if} (\text{typeOf}x \equiv \text{typeOf}y) \text{ then} \\
& \text{typeOf}x \\
& \text{else error} "\text{typeOf: Binop on distinct types.}" \\
& \text{where} \quad \text{typeOf}x = \text{typeOf} x \\
& \text{typeOf}y = \text{typeOf} y \\
\text{typeOf} (\text{Test} \_ t1 t2 ) &= \\
& \text{if} (\text{typeOf}t1 \equiv \text{typeOf}t2) \text{ then} \\
& \text{typeOf}t1 \\
& \text{else error} "\text{typeOf: Test exits on distinct types}"
\end{align*}
\]

By convention, the value returned by sizeof is an unsigned 64 bits integer:

\[
\text{typeOf} (\text{Sizeof} \_ ) = T\text{Int} \text{ Unsigned } T\text{Int64}
\]

Finally, the type of a casted expression is the assigned type. Note that we do not judge of the legality of this cast here. This aspect is handled by the dynamic semantics of FoF’s meta-language.

\[
\text{typeOf} (\text{Cast} t \_ ) = t
\]

\text{readOf and unfoldPtrType: } When we read the content of the reference cell, of type TPointer typeCell modeCell, the type of the object read is either:

- A constant of type typeCell, or
- A reference cell of type typeCell, in a Read mode

We can distinguish both cases thanks to typeCell. If typeCell is a TPointer itself (first case, below), this means that we are dealing with a reference cell. If typeCell is a base type (second case), this means that this is a constant.
readOf :: TypeExpr → TypeExpr
readOf (TPointer typ _) = TPointer typ Read
readOf x = x

unfoldPtrType :: PureExpr → TypeExpr
unfoldPtrType (CLRef (TPointer typ _)) = readOf typ

**liftType**: Although our Reference Cell representation abstracts away the distinction between variables and pointers, it has one drawback. A variable is assigned a `TPointer` type, whereas, in C, we will be working one `TPointer`-level below: our reference cell types corresponds to the same C type but one pointer dereference. Hence, we introduce the following lifting function:

\[
\text{liftType} :: \text{TypeExpr} \rightarrow \text{TypeExpr} \\
\text{liftType} (\text{TPointer } x) = x \\
\text{liftType } x = x
\]

deref: The `deref` is another operator dealing with the specify of reference cells. In the compiler, we translate the high-level reference cell operators by pointer manipulations and assignment. Therefore, when manipulating a reference cell, we will not interested in its actual content but its address. Hence the following function. Values will manipulated just as usual, by value.

\[
\text{deref} :: \text{PureExpr} \rightarrow \text{String} \\
\text{deref} (\text{CLRef } (\text{TPointer } \_ \_ \_)) = "&" \\
\text{deref } \_ = ""
\]

### 6.3 Evaluator

Just as for the compiler, described in the previous section, the implementation of the Filet-o-Fish interpreter is automatically derived from the one-step interpreters. Again, `foldSemantics` comes to the rescue and computes the interpreter:

\[
\text{run} :: \text{Semantics FoFConst PureExpr} \rightarrow \text{Heap} \rightarrow (\text{PureExpr, Heap}) \\
\text{run} = \text{foldSemantics } (\_ \_ \_ ) \text{ runAlgebra}
\]
Chapter 7

The Paka Intermediate Language

Listen, lad.
I’ve built this kingdom up from nothing.
When I started here, all there was was swamp. All the kings said I was daft to build a castle in a swamp, but I built it all the same, just to show ’em. It sank into the swamp.
So, I built a second one. That sank into the swamp.
So I built a third one. That burned down, fell over, then sank into the swamp.
But the fourth one stayed up. An’ that’s what your gonna get, lad – the strongest castle in these islands.

Monty Python

7.1 The Paka Intermediate Language

The purpose of Paka is to ease the task of tracking down unnecessary variable assignment in the to-be-generated C code. Therefore, its syntax is extremely close to C and focused on intra-procedural statements. This is reflected by the definition of PakaCode: the structure of the C file is almost here, with includes, type definitions and prototypes, function prototypes, and function definitions, in this order.

Note that they are all defined by a Map or associative list from String to something else. The String plays the role of an identifier which should be compiled only once in the C code. Typically, a type definition should appear only once, otherwise the C compiler will complain. Map is used when the definition order is not important, associative list is used when we want to keep it (when a declaration might be defined in term of another declaration defined earlier).

data PakaCode
  = PakaCode { includes :: Map.Map String Doc,
                types :: Map.Map String Doc,
                declarations :: [(String, Doc)],
                prototypes :: Map.Map String Doc,
                globalVars :: [(String, Doc)],
                functions :: Map.Map String (Doc, Doc, String, Doc, PakaIntra, ILPaka) }

emptyCode = PakaCode { includes = Map.empty,
Each function is defined by a `PakaIntra` record, which stands for *intra-procedural*. In there, we find local variable definitions and, potentially, a constant. This constant is used to carry the result of a side-effecting test: the side-effecting is compiled before the test-handler and the constant is tested instead.

```haskell
data PakaIntra
    = PakaIntra { localVars :: Map.Map String Doc,
                  expr :: (Maybe PureExpr) }
    deriving Show
de‘ndef emptyIntra = PakaIntra { localVars = Map.empty,
                                    expr = Nothing }
```

As part of the definition of functions, we find the body of the function. This is presented as an `ILPaka` data-type. This is a strip-down version of the `FoF IL`: we have kept most of the control-flow structures (at the exception of the `for` loop, translated into `while` loops) and statements. Because we are describing intra-procedural code, we have removed the function definition construct.

```haskell
data ILPaka
    = PVoid
    | PClosing PakaClosing
    | PStatement PakaStatement ILPaka
    | PIf ILPaka PureExpr ILPaka ILPaka ILPaka
    | PWhile ILPaka PureExpr ILPaka ILPaka
    | PDoWhile ILPaka ILPaka PureExpr ILPaka
    | PSwitch PureExpr [((PureExpr, ILPaka)] ILPaka ILPaka
```

However, the major specificity of Paka is its definition of a statement: a statement is either an assignment or an instruction. An assignment `PAssign x t ys` is a term `t` in which the variable `x` is assigned a value computed from the variables `ys`. On the other hand, an instruction `PInstruction t ys` is a side-effecting operation `t` making use of the variables `ys`.

In a nutshell, when chasing redundant assignments, we will track down raw assignment `PAssign x t [y]`, remove the assignment, and replace all use of `x` by `y`.

```haskell
data PakaStatement
    = PAssign PakaVarName Term [PakaVarName]
    | PInstruction Term [PakaVarName]
```

A `Term` is an almost valid C statement, with holes in it. The holes correspond to the variable names: provided with the list of variable names, it computes a C statement.

Hence, when we have settled the input and output variables of a `PAssign x t ys`, we obtain the corresponding C statement by applying `t x : xs`. Similarly, we get the C code from an instruction `PInstruction t ys` by computing `t ys`.

```haskell
type Term = [Doc] → Doc
```

However, things are not that simple. First, we need more information about the variable: are they raw C variables, or pointers, or dereferenced from somewhere else? This information is vital to avoid aliasing issues.

Similarly, when a variable `y` is used in some operationally non-trivial term `t`, we cannot simply replace `x by y`: we would have to compute some sort of `t y` to be correct. Although it would be doable, we do not support that at the moment and tag the variable name as `Complex`, meaning “non prone to simplification”.

---

Barrelfish TN-024

FileOfFish - 57
Finally, constants are a gold opportunity we don’t want to miss, hence we explicitly carry the value instead of variable name. Therefore, we are able to do some naive constant propagation for free.

```
data PakaVarName
    = Var String
    | Ptr PakaVarName
    | Deref PakaVarName
    | Complex PakaVarName
    | K PureExpr
    deriving (Show, Eq)
```

```
data PakaClosing
    = PReturn PureExpr
    | PBreak
    | PContinue
    deriving Show
```

### 7.2 Paka building blocks

I’m particularly proud of the Paka code generation architecture. To build a Paka term, we simply call some builders functions which are chained up together with the `#` operator. These builders take care of inserting the definitions in the right place in `PakaCode`, `PakaIntra`, or sequentially extend the `ILPaka` code. Thanks to that machinery, we don’t have to explicitly build these data-structures, we just call functions.

Hence, a builder is just putting a brick in the `PakaBuilding` wall:

```
type PakaBuilding = (ILPaka -> ILPaka, PakaCode, PakaIntra)
```

That is, operations taking some arguments and extending a `PakaBuilding` into a new one.

#### 7.2.1 Low-level machinery

To give a feeling of “sequential code”, the `#` operator is simply an inversed composition operation:

```
f # g = \lambda x \rightarrow g (f x)
```

Using `#`, we will compose our builders with a sequential feeling.

Because most, if not all, operations modify one element of the `PakaBuilding` triple, we define the following combinators:

```
first :: (a -> b) -> (a, c, d) -> (b, c, d)
first f (a, b, c) = (f a, b, c)
second :: (a -> b) -> (c, a, d) -> (c, b, d)
second f (a, b, c) = (a, f b, c)
third :: (a -> b) -> (c, d, a) -> (c, d, b)
third f (a, b, c) = (a, b, f c)
```
7.2.2 Building PakaCode

We can add new C includes:

```
include :: String → PakaBuilding → PakaBuilding
include id = second $ include' id
    where include' id globalEnv
        = case id 'Map.lookup' incls of
            Nothing → globalEnv { includes = Map.insert id decl incls }
            Just _ → globalEnv
    where incls = includes globalEnv
        decl = text "#include" < + > text id
```

We can declare new C types:

```
declare :: String → Doc → PakaBuilding → PakaBuilding
declare id typ decl = second $ declare' id typ decl
    where declare' id typ decl globalEnv =
        case id 'Map.lookup' typs of
            Nothing → globalEnv { declarations = (id, decl) :decls,
                types = Map.insert id typ typs }
            Just _ → globalEnv
    where decls = declarations globalEnv
        typs = types globalEnv
```

We can declare global variables:

```
globalVar :: String → Doc → PakaBuilding → PakaBuilding
globalVar id def = second $ globalVar' id def
    where globalVar' id def globalEnv =
        case id 'lookup' vars of
            Nothing → globalEnv { globalVars = (id, def) : vars }
            Just _ → globalEnv
    where vars = globalVars globalEnv
```

We can add function prototypes:

```
prototype :: String → Doc → PakaBuilding → PakaBuilding
prototype id proto = second $ prototype' id proto
    where prototype' id proto globalEnv =
        case id 'Map.lookup' protos of
            Nothing → globalEnv { prototypes = Map.insert id proto protos }
            Just _ → globalEnv
    where protos = prototypes globalEnv
```

And we can define new functions:

```
function :: Doc → Doc → String → Doc → PakaIntra → ILPaka → PakaBuilding → PakaBuilding
function returnT attrs funName funArgs lEnv body =
    second $ function' returnT attrs funName funArgs lEnv body
    where function' returnT attrs funName funArgs lEnv body gEnv =
        case funName 'Map.lookup' functions' of
            Nothing → gEnv { functions = Map.insert funName (returnT, attrs, funName, funArgs, lEnv, body) functions' }
            Just _ → gEnv
    where functions' = functions gEnv
```
7.2.3 Building PakaIntra

As for global variables in the PakaCode, we can add local variables in the PakaIntra environment:

\[
\text{localVar} :: \text{String} \to \text{Doc} \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{localVar id def = third $ localVar' id def}
\]

\[
\text{where localVar' id def localEnv} = \begin{case}
\text{Nothing \to localEnv \{localVars = Map.insert id def vars\}}
\text{Just _ \to localEnv}
\end{case}
\]

\[
\text{where vars = localVars localEnv}
\]

And we can bring a constant in the PakaIntra:

\[
\text{constant} :: \text{PureExpr} \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{constant e = third}$\text{constant' e}
\]

\[
\text{where constant' e lEnv} = \begin{case}
\text{Nothing \to lEnv \{expr = Just e\}}
\end{case}
\]

7.2.4 Building ILPaka

Obviously, the serious stuff happens in ILPaka, or more precisely ILPaka \to ILPaka: this code is seriously continuation-passing. The plan is that we want to build a PStatement value. However, we note that, for instance, to build a PStatement value, we need to know the remaining code. But we don’t know it yet, as we are compiling it! So, we return a continuation that waits for that uncompiled chunk and plug it in the right place. Continuation-passing style, yay!

As an example of that technique in action, take a look at instr and assgn below. Apart from that CPS detail, they are computationally trivial, bringing their arguments in the right place of the constructor and returning by calling the continuation.

\[
\text{instr} :: \text{Term} \to [\text{PakaVarName}] \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{instr instruction vars = first $ instr' instruction vars}
\]

\[
\text{where instr' instruction varNames k} = \lambda c \to
\]

\[
\begin{case}
\text{k \& PStatement (PInstruction instruction varNames) c}
\end{case}
\]

\[
\text{assgn} :: \text{PakaVarName} \to \text{Term} \to [\text{PakaVarName}] \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{assgn wVarName assgnmt rVarNames = first $ assgn' wVarName assgnmt rVarNames}
\]

\[
\text{where assgn' wVarName assgnmt rVarNames k} = \lambda c \to
\]

\[
\begin{case}
\text{k \& PStatement (PAssign wVarName assgnmt rVarNames) c}
\end{case}
\]

As you can expect, we need to stop “continuing” at some point. This naturally fits with the role of closing terms:

\[
\text{close} :: \text{PakaClosing} \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{close c = first $ close' c}
\]

\[
\text{where close' c = \lambda k \_ \to k (PClosing c)}
\]

Similarly, the control-flow operators closes all their branches and only continue downward:

\[
\text{pif} :: \text{ILPaka} \to \text{PureExpr} \to \text{ILPaka} \to \text{ILPaka} \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{pif cond test ifTrue ifFalse = first $ pif' cond test ifTrue ifFalse}
\]

\[
\text{where pif' cond test ifTrue ifFalse cont} = \lambda c \to
\]

\[
\begin{case}
\text{cont \& PIf cond test ifTrue ifFalse c}
\end{case}
\]

\[
\text{pwhile} :: \text{ILPaka} \to \text{PureExpr} \to \text{ILPaka} \to \text{PakaBuilding} \to \text{PakaBuilding}
\]

\[
\text{pwhile cond test loop = first $ pwhile' cond test loop}
\]
\[ \text{where } \text{while}' \text{ cond test loop cont } = \lambda c \rightarrow \]
\[ \text{cont } \text{ PWhile cond test loop c} \]
\[ \text{pdoWhile :: ILPaka } \rightarrow \text{ ILPaka } \rightarrow \text{ PureExpr } \rightarrow \text{ PakaBuilding } \rightarrow \text{ PakaBuilding} \]
\[ \text{pdoWhile loop cond test } = \text{ first } \$ \text{ pdoWhile'} \text{ loop cond test} \]
\[ \text{where pdoWhile' loop cond test cont } = \lambda c \rightarrow \]
\[ \text{cont } \text{ PDOWhile loop cond test c} \]
\[ \text{pswitch :: PureExpr } \rightarrow \text{ [(PureExpr, ILPaka)] } \rightarrow \text{ ILPaka } \rightarrow \text{ PakaBuilding } \rightarrow \text{ PakaBuilding} \]
\[ \text{pswitch test cases defaultCase } = \text{ first } \$ \text{ pswitch'} \text{ test cases defaultCase} \]
\[ \text{where pswitch'} \text{ test cases defaultCase cont } = \lambda c \rightarrow \]
\[ \text{cont } \text{ PSwitch test cases defaultCase c} \]

### 7.3 Translating IL.FoF to IL.Paka

To translate IL.FoF code, we simply iterate over it and build the corresponding IL.Paka term.

\[
\text{compileFoFtoPaka } :: \text{ ILFoF } \rightarrow \text{ PakaCode} \\
\text{compileFoFtoPaka code } = \text{ ccode} \\
\text{where } (\_, \text{ ccode, } \_) = \text{ compileFoFtoPaka'} \text{ code } (\text{id, emptyCode, emptyIntra})
\]

The translation is often trivial, because both languages are very similar in structure. The major novelty is that intra-procedural and extra-procedural code are translated into different data-structures: building an ILPaka term for the former, defining a PakaCode record for the latter. At the same time, we carry a PakaIntra environment during intra-procedural compilations. All these details are abstracted away by the builders we have defined in the previous section and that we abuse in this section.

At this stage, the compiler simply dispatches to construct-specific compilers. Hence the following code:

\[
\text{compileFoFtoPaka'} :: \text{ ILFoF } \rightarrow \text{ PakaBuilding } \rightarrow \text{ PakaBuilding} \\
\text{compileFoFtoPaka'} (\text{FStatement stmt k}) = \text{ compileFoFtoPakaStmt stmt k} \\
\text{compileFoFtoPaka'} (\text{FIf _ _ _}) = \text{ compileFoFtoPakaIf t} \\
\text{compileFoFtoPaka'} (\text{FClosing c}) = \text{ compileFoFtoPakaClosing c} \\
\text{compileFoFtoPaka'} (\text{FNewDef _ _ _}) = \text{ compileFoFtoPakaFunDef t} \\
\text{compileFoFtoPaka'} (\text{FWhile _ _ _}) = \text{ compileFoFtoPakaWhile t} \\
\text{compileFoFtoPaka'} (\text{FDoWhile _ _ _}) = \text{ compileFoFtoPakaDoWhile t} \\
\text{compileFoFtoPaka'} (\text{FFor _ _ _}) = \text{ compileFoFtoPakaFor t} \\
\text{compileFoFtoPaka'} (\text{FSwitch _ _ _}) = \text{ compileFoFtoPakaSwitch t} \\
\text{compileFoFtoPaka'} (\text{FConstant e}) = \text{ compileFoFtoPakaCst e}
\]

### 7.3.1 Compiling Function definition

The compilation of a function definition consists in building a prototype, compiling the body of the function, building it, and pursuing with the next definition.

\[
\text{compileFoFtoPakaFunDef } :: \text{ ILFoF } \rightarrow \text{ PakaBuilding } \rightarrow \text{ PakaBuilding} \\
\text{compileFoFtoPakaFunDef (FNewDef funAttrs} \\
\text{ funName} \\
\text{ body} \\
\text{ returnT} \\
\text{ args} \\
\text{ k)} = \text{ prototype funName } (\text{ attr } + > \text{ returnType } + > \text{ text funName } <> \text{ parens functionArgs } <> \text{ semi})
\]
7.3.2 Compiling Constant

This one is directly handled by the so-called builder:

```
compileFoFtoPakaCst :: PureExpr → PakaBuilding → PakaBuilding
compileFoFtoPakaCst = constant
```

7.3.3 Compiling Closing statements

As for closing statements, this is not much more difficult:

```
compileFoFtoPakaClosing :: FClosing → PakaBuilding → PakaBuilding
compileFoFtoPakaClosing (FReturn expr) = close $ PReturn expr
compileFoFtoPakaClosing (FBreak) = close PBreak
compileFoFtoPakaClosing (FContinue) = close PContinue
```

7.3.4 Compiling control-flow operators

The mechanics of control-flow operators does not vary much between operators, so they are all here, together.

Some points worth mentioning. First, sub-branches are compiled down with `compileFoFtoPaka'`, as one would expect. Second, to get a ILPaka value out of an ILPaka → ILPaka continuation k, we call `k pVoid`: `void` is the ultimate closing statement, after all. Third, an expression computing a tested value must return a pure expression, which we can grab fromJust $ expr intraEnv. This is an invariant, if not respected fromJust will blow up.

Finally, it’s all fine and good to compile sub-branches privately (inside where statements) but don’t forget to bring the resulting global and local environments in the public setting. This corresponds to the use of `second (const globalEnv)` and `third (const localEnv)` in the public flow. Also, don’t forget to thread these environments in your private compilations, too. Someone should think of a less error-prone solution.

```
compileFoFtoPakaIf :: ILFoF → PakaBuilding → PakaBuilding
compileFoFtoPakaIf (FIf cond
  ifTrue
  ifFalse
  k) (cont, gEnv, lEnv) =
  pif pcond test cifTrue cifFalse
  # second (const gEnv3)
```
third (const lEnv3)

$ (cont, gEnv3, lEnv3)$

where (cond_, gEnv1, lEnv1) = compileFoFtoPaka’ cond (id, gEnv, lEnv)
  ccond = ccond_ PVoid
  test = fromJust $ expr lEnv1
  (cifTrue_, gEnv2, lEnv2) = compileFoFtoPaka’ ifTrue (id, gEnv1, lEnv1)
  cifTrue = cifTrue_ PVoid
  (cifFalse_, gEnv3, lEnv3) = compileFoFtoPaka’ ifFalse (id, gEnv2, lEnv2)
  cifFalse = cifFalse_ PVoid

compileFoFtoPakaWhile (FWhile cond loop)
  k) (cont, gEnv, lEnv) =
      while cond test cloop
      # second (const gEnv2)
      # third (const lEnv2)
      # compileFoFtoPaka’ k
      $(cont, gEnv2, lEnv2)$

where (cond_, gEnv1, lEnv1) = compileFoFtoPaka’ cond (id, gEnv, lEnv)
  ccond = ccond_ PVoid
  test = fromJust $ expr lEnv1
  (cloop_, gEnv2, lEnv2) = compileFoFtoPaka’ loop
    # compileFoFtoPaka’ cond
    $(id, gEnv1, lEnv1)$
  cloop = cloop_ PVoid

compileFoFtoPakaDoWhile (FDoWhile loop cond)
  k) (cont, gEnv, lEnv) =
      pdoWhile cloop ccond test
      # second (const gEnv2)
      # third (const lEnv2)
      # compileFoFtoPaka’ k
      $(cont, gEnv2, lEnv2)$

where (cond_, gEnv1, lEnv1) = compileFoFtoPaka’ cond (id, gEnv, lEnv)
  ccond = ccond_ PVoid
  test = fromJust $ expr lEnv1
  (cloop_, gEnv2, lEnv2) = compileFoFtoPaka’ loop
    # compileFoFtoPaka’ cond
    $(id, gEnv1, lEnv1)$
  cloop = cloop_ PVoid

compileFoFtoPakaSwitch (FSwitch test cases)
  k) (cont, gEnv, lEnv) =
      pswitch test cases ddefaultCase
      # second (const gEnv2)
      # third (const lEnv2)
      # compileFoFtoPaka’ k
      $(cont, gEnv, lEnv)$

where (defaultCase_, gEnv1, lEnv1) = compileFoFtoPaka’ defaultCase (id, gEnv, lEnv)
  ddefaultCase = ddefaultCase_ PVoid
  (codes, fcases) = unzip cases
  (ccases_, gEnv2, lEnv2) = compileCases fcases gEnv1 lEnv1
  ccases = zip codes ccases_
  compileCases [] x y = ([], x, y)
compileCases (fcase : fcases) gEnv lEnv =
    - cfcase 'deepSeq' codes 'deepSeq'
    (cfcase : codes, gEnv2, lEnv2)
    where (cfcase, gEnv1, lEnv1) = compileFoFtoPaka' fcase (id, gEnv, lEnv)
        cfcase = cfcase' PVoid
    (codes, gEnv2, lEnv2) = compileCases fcases gEnv1 lEnv1

For my personal convenience, for loops are compiled into while loops. If you’re not happy with that, go ahead and implement that. However, I have to warn you that dealing with computations inside the indices is not a joy.

compileFoFtoPakaFor (FFor init
  test
  inc
  loop
  k) (cont, gEnv, lEnv) =
    while ccond etest cloop
    # second (const gEnv2)
    # third (const lEnv2)
    # compileFoFtoPaka' k
    $ (cont, gEnv2, lEnv2)
    where (cond, gEnv1, lEnv1) = compileFoFtoPaka' init
    # compileFoFtoPaka' test
    $ (id, gEnv, lEnv)
    ccond = ccond' PVoid
    etest = fromJust $ expr lEnv1
    (cloop, gEnv2, lEnv2) = compileFoFtoPaka' loop
    # compileFoFtoPaka' inc
    # compileFoFtoPaka' test
    $ (id, gEnv1, lEnv1)
    cloop = cloop' PVoid

7.3.5 Compiling statements

The real stuff happens below: compiling these damned statements. And there is a lot of them. That was for the bad news. The good news is that, individually, these functions are quite easy to understand.

The careful reader will notice that Terms are not using all their arguments. Honestly, I just wanted the basic Optimizer to be done, so I dropped everything not necessary. So, you have the architecture, now fill the holes if you want to do something clever. Therefore, when you see a term defined with \((\lambda [xs, \_, xss] \to ...)\), this means that the ignored variable is hard-coded in the term, and cannot be actually replaced. This is ok with my simple optimizer, that would probably need to be changed if you are to do something more clever.

Compiling References

As a starting, non frightening example, here is the code to compile references. Honestly, it is self-explanatory, isn’t it?

compileFoFtoPakaStmt (FNewRef varName dat) k =
    localVar mvarName (toC (typeOf dat) < + > toC varName <> semi)
    # assgn pvarName \(\lambda [\_, e] \to toC varName < + > \text{char '='} < + > e <> \text{semi})
    [pakaVarName dat]
    # compileFoFtoPaka' k
where mvarName = mkPakaVarName varName

where mvarName = mkPakaVarName varName

compileFoFtoPakaStmt (FReadRef varName ref) k =
localVar mvarName (toC (unfoldPtrType ref) < + > toC varName <> semi)
  # assign pvarName (\[\_ e] ->
  toC varName < + > char '=' < + > e <> semi)
  [pakaValName ref]
  # compileFoFtoPaka' k
where mvarName = mkPakaVarName varName

compileFoFtoPakaStmt (FWriteRef ref dat) k =
assign (pakaValName ref)
  (\[\_ e] -> toC ref < + > char '=' < + > e <> semi)
  [pakaValName dat]
  # compileFoFtoPaka' k

Compiling Arrays

Similarly, compiling arrays work the same way. There is minor nitpick in the current implementation: it doesn’t support dynamic array (that is, malloc’ed arrays).

Actually, I suspect that if you are reading this file, it is because your code is using a dynamic array and the compiler blew up when you use it. Well, the code needs to be written. It is remotely similar to static arrays, with the additional need to malloc memory and initialize the data. If you are looking for a word to describe your situation, I think that “screwed” is appropriate. Hint: a dynamic array should be defined by a single initial element and an integer variable specifying (at run time) the length of the array.

compileFoFtoPakaStmt (FNewArray varName alloc@ (StaticArray size) dat) k =
globalVar mvarName (toC typeOfDat < + > toC varName <> brackets Pprinter.empty
  < + > char '=' < + > braces (nest 4 $
  fsep (punctuate comma
    [text (deref val) <> toC val
     | val ← dat]|)
  ) <> semi)
  # compileFoFtoPaka' k
where mvarName = mkPakaVarName varName
typeOfDat = typeOf $ head dat

compileFoFtoPakaStmt (FReadArray varName
  (CLRef origin
   (TArray (StaticArray size) typ)
   xloc)
  index) k =
localVar mvarName (toC typ < + > toC varName <> semi)
  # (case symbEval index of
  CLInteger _ x ->
  if x < (toInteger size) then
    assign pvarName (\[\_ \_ ] ->
      toC varName < + > char '='
      < + > toC xloc <> brackets (toC index) <> semi)
      [Complex $ Var $ mkPakaVarName xloc]
else

    instr (\_, \_ \rightarrow
        text "assert" <> parens (text "! "ReadArray: Out of bound\"") <> semi)
    []
    _ \rightarrow

    assign pvarName (\[\_, \_ \rightarrow
        text "if" < + > parens (e
            < + > char '<'
            < + > int size) <> brace
        § + §
        nest 4 (toC varName < + > char '='
            < + > toC zloc <> brackets e <> semi)
        § + §
        rbrace < + > text "else" < + > brace § + §
        nest 4 (text "assert" <> parens (text "! "ReadArray: Out of bound\"") <> semi
            § + § toC varName < + > char '=' < + > text "NULL" <> semi)
        § + §
        rbrace)


    [ Complex $ Var $ mkPakaVarName xloc,
      pakaValName index]]
    # compileFoFtoPaka k
    where mvarName = mkPakaVarName varName
          pvarName = Var $ mkPakaVarName varName
    compileFoFtoPakaStmt (fwriteArray ref @(CLRef origin
                      (TArray (StaticArray size) typ)
                      xloc)

    index
dat) k =
    assign pxloc (\[\_, e, f] \rightarrow
        text "if" < + > parens (f < + > char '<' < + > int size) <> brace
        § + § nest 4 (toC zloc <> brackets f
            < + > char '=' < + > e <> semi)
        § + § rbrace < + > text "else" < + > brace
        § + § nest 4 (text "assert" <> parens (text "! "Out of bound \"") <> semi)
        § + § rbrace) [pakaValName dat, pakaValName index]
    # compileFoFtoPaka k
    where pxloc = Var $ mkPakaVarName zloc

Compiling Strings

Building a new string is as simple as building a new static array:

    compileFoFtoPakaStmt (FNewString varName dat) k =
    globalVar mvarName (toC TChar < + > toC varName <> text "[]"
        < + > char '=',
        < + > doubleQuotes (text dat) <> semi)
    # compileFoFtoPaka k
    where mvarName = mkPakaVarName varName

Compiling Function call

As for function call, there is no black magic either:
**Compiling Enumerations**

We can safely compile enumerations:

```plaintext
compileFoFtoPakaStmt (FNewEnum varName
  nameEnum
  fields
  initVal) k =
  declareEnum nameEnum fields
  # compileFoFtoPaka' k
  where mvarName = mkPakaVarName varName
       pvarName = Var $ mkPakaVarName varName
```

**Compiling Union**

It is not a big deal to compile union operations either:

```plaintext
compileFoFtoPakaStmt (FNewUnion name
  DynamicUnion
  nameUnion
  fields
  (initField, initData)) k =
  declareRecursive (TUnion DynamicUnion nameUnion fields)
  # localVar (mkPakaVarName name) (text "union" < + > text nameUnion <> char ’*’ < + > toC name <> semi
  # assgn varName (λ[] →
    toC name < + > char ’*’ < + >
    parens (text "union" < + > text nameUnion <> char ’*’)
    < + > text "malloc" <> parens (text "sizeof" <> parens (text "union" < + > text nameUnion))
    <> semi) []
```
Compiling Structs

Quite the same goes for structure operations:

```haskell
compileFoFtoPakaStmt (FNewStruct varName DynamicStruct nameStruct)
```
declareRecursive (TStruct DynamicStruct nameStruct fieldsTypeStr)

# localVar mVarName (text "struct" < + > text nameStruct < + > toC varName <> semi)
# (assign pVarName (\[_."] \rightarrow
toC varName < + > char '=',
< + > parens (text "struct" < + > text nameStruct < + > char '•')
< + > text "malloc"
< > parens (text "sizeof")
< > parens (text "struct" < + > text nameStruct)
< > semi) [])

# foldl' (#) id [assign pVarName (\[_."] \rightarrow
toC varName <> text "->" <> text field
< + > char '=',
< + > e <> semi) [pakaVarName val]
| (field, (typ, val)) \< fields]

where mVarName = mkPakaVarName varName
pVarName = Var \$ mkPakaVarName varName
fieldsTypeStr = [(field, typ)
| (field, (typ, _)) \< fields]

compileFoFtoPakaStmt (FNewStruct varName
StaticStruct
nameStruct
fields) k =
declareRecursive (TStruct StaticStruct nameStruct fieldsTypeStr)

# localVar mVarName (text "struct" < + > text nameStruct < + > toC varName
< + > char '=',
< + > braces (nest 4 $)
hecat (punctuate comma
| text (deref val) <> toC val
| (_. (._, val)) \< fields)])
< > semi)

where mVarName = mkPakaVarName varName
pVarName = Var \$ mkPakaVarName varName
fieldsTypeStr = [(field, typ)
| (field, (typ, _)) \< fields]

compileFoFtoPakaStmt (FReadStruct varName
ref@(CLRef origin
typeS@(TStruct alloc
nameStruct
fields)
xloc)
field) k =
declareRecursive typeS

# localVar mVarName (toC typeField < + > toC varName <> semi)
# (assign pVarName (\[_."] \rightarrow
toC varName < + > char '=',
< + > toC xloc <> ptrSigStruct alloc <> text field <> semi)
| Complex $ Var $ mkPakaVarName xloc]

where typeField = fromJust $ field 'lookup' fields
mVarName = mkPakaVarName varName
pVarName = Var $ mkPakaVarName varName

compileFoFtoPakaStmt (FWriteStruct ref@(CLRef origin
typeS@(TStruct alloc
nameStruct

Barrelfish TN-024
FileOfFish - 69
fields) zloc) field value) k =
declareRecursive typeS
# assign pxloc (λ[, e] →
toC zloc <> ptrSigStruct alloc <> text field
  <+ > char '+=' <+ > e <> semi)
  [pakaVarName value]
# compileFoFtoPaka' k
where pxloc = Var $ mkPakaVarName zloc

Compiling Typedef

And we can even get typedefs:

compileFoFtoPakaStmt (FTypedef typ aliasName) k =
declareRecursive typ
# declare aliasName Pprinter.empty
  (text "typedef" <+ > toC typ <+ > text aliasName <> semi)
# compileFoFtoPaka' k
compileFoFtoPakaStmt (FTypedefE inclDirective (TTypedef typ aliasName)) k =
  include inclDirective
# compileFoFtoPaka' k

Compiling Foreign calls

It is always the same story for foreign function calls. If you have extended Filet-o-Fish with a new
foreign-function, don’t look further: you should put your foreign call here!

So, as often, we have an inoffensive dispatcher. Don’t touch it.

compileFoFtoPakaStmt (FFFICall nameCall args) k =
  compileFFI nameCall args
# compileFoFtoPaka' k

And the dispatched function, in which you should add your foreign code generator. This is just like
writing C code, so don’t be shy.

compileFFI nameCall params | nameCall ≡ "printf" =
  include "<stdio.h>"
# instr (λ → text "printf" <> parens (heat (punctuate comma (map toC params))) <> semi)
  (map (Complex ◦ pakaVarName) params)
compileFFI nameCall [e] | nameCall ≡ "assert" =
  include "<assert.h>"
# instr (λ[e] → text "assert" <> parens e <> semi) [pakaValName e]
compileFFI nameCall [varName, param] | nameCall ≡ "has_descendants" =
  include "<mdb/mdb.h>"
# include "<capabilities.h>"
# include "<stdbool.h>"
# localVar (show $ toC varName)
  (text "bool" <+ > toC varName <> semi)
Declaring types

Above, we have dealt with the compilation of operations on complex structures, such as enums, structs, and unions. When compiling a code operating on such structure, we need to make sure that the corresponding type is defined.

Hence, we provide an advanced builder to declare a struct or an union:

\[
\text{declareStructUnion kind name fields =}
\]
\[
\text{declare name (text kind } < + > \text{ text name } < > \text{ semi)}
\]
\[
\text{nest 4 (vcat'}\text{toC typ } < + > \text{ text field } < > \text{ semi)
- special case for static array?
| (field, typ) } \leftarrow \text{ fields}) \text{)} \text{ } < > \text{ semi)
\]

And similarly for declaring an enum, however without the forward declaration:

\[
\text{declareEnum nameEnum fields =}
\]
\[
\text{declare nameEnum empty}
\]
\[
\text{nest 4 (vcat'}\text{int val } < + > \text{ text name } < > \text{ semi)
| (name, val } \leftarrow \text{ fields})}) \text{)} \text{ } < > \text{ semi)
\]
However, that does not solve the problem: a structure or an union might be defined in terms of other structures or unions. Hence, we need to declare the dependencies before defining the concerned object. This is handled by \texttt{declareRecursive}:

\begin{verbatim}
declareRecursive = declareRecursive'
  where declareRecursive' (TStruct _ name fields) (code, gEnv, lEnv) =
    case name `Map.lookup' types gEnv of
    Just _ -> (code, gEnv, lEnv)
    Nothing ->
    foldl' (#) id [declareRecursive' typ | (_, typ) <- fields]
    # declareStructUnion "struct" name fields
    $ (code, gEnv, lEnv)

declareRecursive' (TUnion _ name fields) (code, gEnv, lEnv) =
  case name `Map.lookup' types gEnv of
  Just _ -> (code, gEnv, lEnv)
  Nothing ->
  foldl' (#) id [declareRecursive' typ | (_, typ) <- fields]
  # declareStructUnion "union" name fields
  $ (code, gEnv, lEnv)

declareRecursive' (TEnum name fields) t =
  declareEnum name fields $ t

declareRecursive' _ t = id t
\end{verbatim}

These two functions have also been handy above, even though they are not fundamentally clever. Depending on the allocation policy of the data-structure, they choose to dereference and access it, or directly access it.

\begin{verbatim}
ptrSigUnion :: AllocUnion -> Doc
ptrSigUnion DynamicUnion = text "->"
ptrSigUnion StaticUnion = char '.'
ptrSigStruct :: AllocStruct -> Doc
ptrSigStruct DynamicStruct = text "->"
ptrSigStruct StaticStruct = char '.'
\end{verbatim}

\section*{7.4 Translating IL.Paka to C}

This file could as well be called ./IL/C/C.lhs but I felt guilty of introducing yet another confusing IL. So, it is here but feel free to move it around.

\subsection*{7.4.1 Printing types and expressions}

Because we are good kids, we create a type-class called \texttt{Compileable}. A data-type satisfying \texttt{Compileable} can be pretty-printed to something vaguely looking like a bunch of C code.

\begin{verbatim}
class Compileable a where
toC :: a -> Doc
\end{verbatim}

Part of the \texttt{Compileable} class are FoF’s types \texttt{TypeExpr} and FoF’s pure expressions \texttt{PureExpr}.

There is nothing but boiler plate code to get the job done for pure expressions:

\begin{verbatim}
instance Compileable PureExpr where
  toC (Quote s) = doubleQuotes $ text s
\end{verbatim}
instance Compileable UnaryOp where
  toC Minus = char ' -'
  toC Complement = char '~'
  toC Negation = char ' !'

instance Compileable BinaryOp where
  toC Plus = text " +"
  toC Sub = text " -"
  toC Mul = text " *"
  toC Div = text " /"
  toC Mod = text " %"
  toC Shl = text " <<"
  toC Shr = text " >>"
  toC AndBit = text " &"
  toC OrBit = text " |
  toC XorBit = text " ^
  toC Le = text " <=
  toC Leq = text " <=
  toC Ge = text " >=
  toC Geq = text " >=
  toC Eq = text " ==
  toC Neq = text " !=

And similarly for types:

instance Compileable TypeExpr where
  toC TInt8 = text " int8_t"
  toC TInt16 = text " int16_t"
  toC TInt32 = text " int32_t"
  toC TInt64 = text " int64_t"
  toC TInt8_t = text " uint8_t"
  toC TInt16_t = text " uint16_t"
  toC TInt32_t = text " uint32_t"
  toC TInt64_t = text " uint64_t"
  toC TFloat = text " float"
  toC TVoid = text " void"
  toC TChar = text " char"
  toC TArray DynamicArray typ = toC typ <> char '*'
  toC TArray (StaticArray size) typ = toC typ <> char '*'
  toC TPointer x = toC x <> char '*'
  toC TStruct DynamicStruct name fields = text " struct " <> text name <> char '*'
  toC TStruct StaticStruct name fields = text " struct " <> text name
  toC TUnion DynamicUnion name fields = text " union " <> text name <> char '*"
toC (TUnion StaticUnion name fields) = text "union " < + > text name
toC (TCompPointer name) = text "uintptr_t"
toC (TTypedef typ name) = text name
toC (TEnum name _) = text "enum" < + > text name

The picky reader will have noticed the absence of printer for function types. This is hardly a problem at
the moment because we do not support function pointers, so we are not going to declare function types
anytime soon. Note that this argument might well be circular: if we do not support function pointers,
it is because it is a pain to write their type, among other things (if I remember correctly). Oh well.

Printing variable names is dead easy:

\[
\text{instance Compileable VarName where}
\text{toC x = text $ mkPakaVarName x}
\]

7.4.2 Names, everywhere

I am not very proud of that section, and of the way I abused these functions in IL/Paka/Paka.lhs. I
beg your pardon for that. There must some abstraction to bust here but I was not able to catch it.

Provided a FoF VarName, we turn it into a string with a bit of Hungarianism, but very little. Why this
function is called mkPakaVarName when it does not deal with PakaVarName? I have no clue.

\[
\text{mkPakaVarName :: VarName -> String}
\text{mkPakaVarName (Generated x) = "_" + x}
\text{mkPakaVarName (Provided x) = x}
\text{mkPakaVarName (Inherited y x) = mkPakaVarName x + "_" + show y}
\]

Then, we have to functions turning a PureExpr into a PakaVarName. PakaValName provides you with
the value described by the PureExpr. On the other hand, PakaVarName works one level below and
gives you the value contained in the PureExpr.

I have to admit that I am not myself convinced by this explanation. Basically, I would have to look at
the former code, the typeOf, deref, readOf functions, and the new code. Then, I might be able to make
more sense of that. However, intrinsically, references are a non-sense.

\[
pakaValName :: PureExpr -> PakaVarName
\text{pakaValName (CLRef origin (TPointer _ Avail) loc) = Var $! mkPakaVarName loc}
\text{pakaValName (CLRef origin (TPointer _ Read) loc) = Ptr $! Var $ mkPakaVarName loc}
\text{pakaValName (CLRef _ loc) = Var $! mkPakaVarName loc}
\text{pakaValName x = K x}
\]

\[
pakaVarName :: PureExpr -> PakaVarName
\text{pakaVarName (CLRef origin (TPointer _ Avail) loc) = Deref (Var $ mkPakaVarName loc)}
\text{pakaVarName (CLRef origin (TPointer _ Read) loc) = Var $ mkPakaVarName loc}
\text{pakaVarName (CLRef _ loc) = Var $ mkPakaVarName loc}
\text{pakaVarName x = K x}
\]

Finally, we need to be able to print these PakaVarName into meaning C code. Here you go.

\[
\text{instance Compileable PakaVarName where}
\text{toC (Deref x) = char '|' &! <> toC x}
\text{toC (Var x) = text x}
\text{toC (Ptr x) = char '*.' <> toC x}
\text{toC (Complex _) = error "Cannot convert a Complex var name to C"}
\text{toC (K x) = toC x}
\]
7.4.3 Generating C

The following is a small addendum to the pretty-printer library. We don’t know why it is not defined there.

\[
\text{vcat'} : \text{[Doc]} \rightarrow \text{Doc} \\
\text{vcat'} [] = \text{empty} \\
\text{vcat'} (x : xs) = l \cdot seq' r \cdot seq' r \\
\text{where} \\
l = \text{vcat'} xs \\
r = x \cdot + \cdot $ + $ l
\]

For once, I will do a bottom-up presentation. So, I will describe the implementation of pretty-printers from Paka code to C.

The first step consists in printing closing terms:

\[
\text{pprintClosing} : \text{PakaClosing} \rightarrow \text{Doc} \\
\text{pprintClosing} (\text{PReturn} e) = \text{text "return" < + > parens (toC e) <> semi} \\
\text{pprintClosing} (\text{PBreak}) = \text{text "break"} \\
\text{pprintClosing} (\text{PContinue}) = \text{text "continue"}
\]

Then, we print statements. As you remember, we need to build the final code by applying the variables to the term:

\[
\text{pprintStmt} : \text{PakaStatement} \rightarrow \text{Doc} \\
\text{pprintStmt} (\text{PAssign} dst x srcs) = x \cdot (\text{toC dst : map toC srcs}) \\
\text{pprintStmt} (\text{PInstruction} x srcs) = x \cdot (\text{map toC srcs})
\]

The next step consists in compiling intra-procedural code. This is rather simple and quite directly follows from the Paka definitions:

\[
\text{pprintPaka} : \text{ILPaka} \rightarrow \text{Doc} \\
\text{pprintPaka} (\text{PVoid}) = \text{empty} \\
\text{pprintPaka} (\text{PClosing} c) = \text{pprintClosing c} \\
\text{pprintPaka} (\text{PStatement} stmt k) = \\
\text{pprintStmt} stmt \cdot + \cdot \text{pprintPaka k} \\
\text{pprintPaka} (\text{PIf} cond test ifTrue ifFalse k) = \\
\text{pprintPaka} cond \cdot + \cdot \text{text "if" < + > parens (toC test) <> lbrace $ + $} \\
\text{(nest 4 $!\cdot \text{pprintPaka ifTrue} $ + $} \\
\text{rbrace < + > text "else" < + > lbrace $ + $} \\
\text{(nest 4 $!\cdot \text{pprintPaka ifFalse} $ + $} \\
\text{rbrace $ + $} \\
\text{pprintPaka k} \\
\text{pprintPaka} (\text{PWhile} cond test loop k) = \\
\text{pprintPaka} cond \cdot + \cdot \text{text "while" <> parens (toC test) <> lbrace $ + $} \\
\text{(nest 4 $!\cdot \text{pprintPaka loop} $ + $} \\
\text{rbrace $ + $} \\
\text{pprintPaka k} \\
\text{pprintPaka} (\text{PDoWhile} loop cond test k) = \\
\text{text "do" < + > lbrace $ + $} \\
\text{(nest 4 $!\cdot \text{pprintPaka loop} $ + $} \\
\text{rbrace < + > text "while" < + > parens (toC test) <> semi $ + $} \\
\text{pprintPaka k} \\
\text{pprintPaka} (\text{PSwitch} test cases defaultCase k) = \\
\text{text "switch" < + > parens (toC test) < + > lbrace $ + $}
\]
Finally, we can pretty-print a complete PakaCode by iterating other each section, and, in each section, pretty-printing each element.

```haskell
nest 4 $ vcat' $ map compileCase cases $ + $
(nest 4 (text "default:" $ + $
  (nest 4! pprintPaka defaultCase) $ + $
  rbrace)) $ + $
 rbrace $ + $
pprintPaka k
where compileCase (i, code) =
  text "case" $ + > toC i $ + $
  colon $ + $
  (nest 4! (pprintPaka code $ + $
    text "break" $ + $
    semicolon)) $ + $
  rbrace

Finally, we can pretty-print a complete PakaCode by iterating other each section, and, in each section, pretty-printing each element.

```
We note the use of two extraction functions: these functions remove the keys from the associative structure in use, and simply return the content. When an order was maintained, i.e. an associative list was used, the definition order is carefully restored by reversing the list.

\[
\begin{align*}
extractL & :: \text{Eq } a \Rightarrow [(a, b)] \rightarrow [b] \\
extractL &= (map \text{ snd}) \circ \text{reverse} \\
extractM & :: \text{Map.Map } a \ b \rightarrow [b] \\
extractM &= \text{Map.elems}
\end{align*}
\]

Because we have worked very hard, we are rewarded by the right to instantiate these \textit{PakaCode} in the \textit{Show}.

\[
\text{instance } \text{Show } \text{PakaCode} \text{ where} \\
\text{show} = \text{render} \circ \text{pprint}
\]

7.5 \textbf{IL. Paka Code Optimizer}

The currently implemented optimizer is a naive redundant assignment simplifier, which happens to do constant propagation at the same time. It is naive in the several dimensions. An important one is that it is entirely hard-coded, while we all know that optimization is simply a matter of dataflow analysis. So, at some point, we should use a more generic framework for that. It is also naive because it does not try to reach a fix-point: it is single phase, while it is obvious that more assignments could still be eliminated in subsequent phases. Finally, it is naive because any case that was not easy to deal with have been discarded: more redundant assignments could be removed if the logic were more precise.

The purpose of that module is to show that “it is possible to do optimization”. It is a proof of concept. Now, it is Future Work (Chapter A) to get a clever optimization framework. The ease I had in implementing that stuff convince me that we are not far from this heaven.

So, if you want optimized Paka code, you will only get a slightly less redundant code:

\[
\begin{align*}
\text{optimizePaka} & :: \text{PakaCode } \rightarrow \text{PakaCode} \\
\text{optimizePaka} &= \text{optimizeAssgmtElim}
\end{align*}
\]

Because this analysis is intra-procedural, we go over each function and apply an intra-procedural optimizer:

\[
\begin{align*}
\text{optimizeAssgmtElim} & :: \text{PakaCode } \rightarrow \text{PakaCode} \\
\text{optimizeAssgmtElim} \ \text{code} &= \text{code} \ \{ \text{functions} = \text{optFunc} \} \\
\text{where} \ \text{funcs} &= \text{functions} \ \text{code} \\
\text{optFunc} &= \text{Map.mapMaybe} \ (\lambda(b, c, d, e, f, \text{fun}) \rightarrow \text{Just } (b, c, d, e, f, \text{assgmtElim fun})) \ \text{funcs}
\end{align*}
\]

7.5.1 \textbf{Implementation}

This optimizer is quite easy to implement, assuming we have the right tools at hand. That is, assuming that we are able to replace a variable \(x\) by a variable \(y\) in a code \(k\) – using \textit{replace} \ ((\text{Var} \ x) \ (\text{Var} \ y) \ k) \rightarrow \), that we are able to say if a variable \(y\) is either a constant or never used in a code \(k\) – using \textit{isUsed flatten} \ y \ k \rightarrow, and that we are able to say if a variable \(x\) is used without side-effects in a code \(k\) – using \textit{isUsed flattenS} \ x \ k.
The intra-procedural optimizer will turn an ILPaka into a better ILPaka:

\[
\text{assgmtElim} :: \text{ILPaka} \rightarrow \text{ILPaka}
\]

The interesting case is obviously the variable assignment: a value \( y \) is assigned to a variable \( x \). We remove that assignment and replace \( x \) by \( y \) if and only if \( y \) is never used again and \( x \) is not involved in some weird computation. Otherwise, we go ahead.

A small issue here is that we ask for \( y \) to be never used again. That's quite restrictive. This results in being able to carry only 4 assignment eliminations on today's Hamlet and Fugu inputs. This is shame, compared to the numerous opportunities. To solve that issue, we would have to extend or re-design \( \text{isUsed} \) to allow the definition of more fine-grained predicates, such as "is overwritten".

\[
\begin{align*}
\text{assgmtElim} \ (PStatement \:\: a \oplus (PAssign \ (Var \ x) \ _ \ [Var \ y]) \: k) &= \\
& \quad \text{if } (\neg (\text{isUsed flatten} \ y \ k)) \\
& \quad \land (\neg (\text{isUsed flattenS} \ x \ k)) \ \text{then} \\
& \quad \text{assgmtElim} \Replacement \ (Var \ x) \ (Var \ y) \: k \\
& \quad \text{else} \\
& \quad \ PStatement \ a \ \Replacement \ k
\end{align*}
\]

All other assignments that do not fit this scheme, or the instructions are skipped:

\[
\text{assgmtElim} \ (PStatement \ a \: k) = \\
\quad \ PStatement \ a \ \Replacement \ k
\]

Finally, control-flow operators are simply iterated over:

\[
\begin{align*}
\text{assgmtElim} \ (PIf \ c \ t \ ifT \ ifF \: k) &= \\
& \quad \ PIf \\
& \quad (\text{assgmtElim} \ c) \\
& \quad t \\
& \quad (\text{assgmtElim} \ ifT) \\
& \quad (\text{assgmtElim} \ ifF) \\
& \quad (\text{assgmtElim} \ k) \\
\text{assgmtElim} \ (PWhile \ c \ t \ l \: k) &= \\
& \quad \ PWhile \\
& \quad (\text{assgmtElim} \ c) \\
& \quad t \\
& \quad (\text{assgmtElim} \ l) \\
& \quad (\text{assgmtElim} \ k) \\
\text{assgmtElim} \ (PDoWhile \ l \ c \ t \: k) &= \\
& \quad \ PDoWhile \\
& \quad (\text{assgmtElim} \ l) \\
& \quad (\text{assgmtElim} \ c) \\
& \quad t \\
& \quad (\text{assgmtElim} \ k) \\
\text{assgmtElim} \ (PSwitch \ t \ cases \ d \: k) &= \\
& \quad \ PSwitch \\
& \quad t \\
& \quad (map \ (\lambda (a, b) \rightarrow (a, \text{assgmtElim} \ b)) \ cases) \\
& \quad (\text{assgmtElim} \ d) \\
& \quad (\text{assgmtElim} \ k) \\
& \quad \text{assgmtElim} \ x = x
\end{align*}
\]
7.5.2 Code predication

First, if I correctly remember my Software Testing lecture, a use site is a place where a variable is read. In opposition to a def site where a variable is written to. Well, then the following is misleading.

\[ isUsed f x k \] tells you that \( x \) has been found in a use or def site of \( k \) in a situation where it played a role caught by \( f \). To simplify, \( isUsed flatten \) will catch any kind of use or def. \( isUsed flattenS \) will catch a use or def in a \( Complex \) state.

As for the implementation, it is simply going over \( ILPaka \) terms and doing the necessary on \( PStatement \).

\[
isUsed :: (PakaVarName \rightarrow Maybe String) \rightarrow String \rightarrow ILPaka \rightarrow Bool
\]
\[
isUsed p var PVoid = False
\]
\[
isUsed p var (PClosing (PReturn k)) = Just var \equiv (flatten $ pakaValName k)
\]
\[
isUsed p var (PClosing _) = False
\]
\[
isUsed p var (PStatement s k) = isUsedStmt s \lor isUsed p var k
\]
\[
where
\[
isUsedStmt :: (PAssign t ls) =
\]
\[
Just var \in \text{map flatten (t : ls)}
\]
\[
isUsedStmt :: (PInstruction ls) =
\]
\[
Just var \in \text{map flatten ls}
\]
\[
isUsed p var (PIf c t ifT ifF k) =
\]
\[
(Just var \equiv (flatten $ pakaValName t)) \lor
\]
\[
isUsed p var c \lor isUsed p var ifT \lor isUsed p var ifF \lor isUsed p var k
\]
\[
isUsed p var (PWhile c t l k) =
\]
\[
(Just var \equiv (flatten $ pakaValName t)) \lor
\]
\[
isUsed p var c \lor isUsed p var l \lor isUsed p var k
\]
\[
isUsed p var (PDoWhile l c t k) =
\]
\[
(Just var \equiv (flatten $ pakaValName t)) \lor
\]
\[
isUsed p var c \lor isUsed p var l \lor isUsed p var k
\]
\[
isUsed p var (PSwitch t c d k) =
\]
\[
(Just var \equiv (flatten $ pakaValName t)) \lor
\]
\[
foldl′ \lambda a (b, c) \rightarrow a \lor isUsed p var b) False c
\]
\[
\lor isUsed p var d \lor isUsed p var k
\]

In light of the explanation above, the definition of \( flatten \) and \( flattenS \) should be obvious. Aren’t they?

\[
flatten :: PakaVarName \rightarrow Maybe String
\]
\[
flatten (Var s) = Just $ s
\]
\[
flatten (Ptr x) = flatten x
\]
\[
flatten (Deref x) = flatten x
\]
\[
flatten (Complex x) = flatten x
\]
\[
flatten (K _) = Nothing
\]
\[
flattenS :: PakaVarName \rightarrow Maybe String
\]
\[
flattenS (Var s) = Nothing
\]
\[
flattenS (Ptr x) = Nothing
\]
\[
flattenS (Deref x) = Nothing
\]
\[
flattenS (Complex x) = flatten x
\]
\[
flattenS (K _) = Nothing
\]

7.5.3 Code transformation

As for replace, it is by now standard: go over the terms, hunt the \( dest \), and kill it with \( source \). It is surgical striking, in its full glory.
replace :: PakaVarName → PakaVarName → ILPaka → ILPaka

replace dest source (PStatement (PAssign dst stmt srcs) k) =
  PStatement (PAssign dst stmt srcs′)
  (replace dest source k)

where srcs′ = replaceL dest source srcs

replace dest source (PStatement (PInstruction stmt srcs) k) =
  PStatement (PInstruction stmt srcs′)
  (replace dest source k)

where srcs′ = replaceL dest source srcs

replace dest source (PIf c t ifT ifF k) =
  PIf (replace dest source c) t
  (replace dest source ifT)
  (replace dest source ifF)
  (replace dest source k)

replace dest source (PWhile c t l k) =
  PWhile (replace dest source c)
  t
  (replace dest source l)
  (replace dest source k)

replace dest source (PDoWhile l c t k) =
  PDoWhile (replace dest source l)
  (replace dest source c)
  t
  (replace dest source k)

replace dest source (PSwitch t cases d k) =
  PSwitch t
  (map (λ(a, b) → (a, replace dest source b)) cases)
  (replace dest source d)
  (replace dest source k)

replace dest source x = x

replaceL x y = map (λz → if z ≡ x then y else z)
Part III

Appendix
Appendix A

Future Work

Follow! But! follow only if ye be men of valor, for the entrance to this cave is guarded by a creature so foul, so cruel that no man yet has fought with it and lived! Bones of four fifty men lie strewn about its lair. So, brave knights, if you do doubt your courage or your strength, come no further, for death awaits you all with nasty big pointy teeth.

Monty Python

This is going to look like a brain dump, despite any effort to make it understandable by the Outside World.

**Module import clean-up:** for historical reasons, some imports might be completely useless now. Similarly, imports such as —Debug.Trace— should disappear too;

**Paka terms with real holes:** in Section 7.3, we have seen that Paka terms are ignoring most of their holes by using hard-coded values;

**More efficient redundant assignment optimizer:** in Chapter 7.5, we have seen that the optimizer is quite conservative, making it quite useless in practice;

**Supporting function pointers:** preventing Filet-o-Fish users to abuse function pointers is a violation to Geneva convention. I do not think that there is some deep technical difficulty to get that. But printing the type of such pointer was a first trouble, if I remember correctly;

**Implementing the interpreter in the Agda language:** this was already one of my goal initially, but the NICTA people insisted that without an in-theorem-prover semantics, the dependability argument is just bullsh*t. Ha, these Australians...;

**Code generator back-back-end:** following the steps of FoF and Paka, we need a more principled back-back-end, generating (correct) out of —FoFCode—;

**Hoopl-based optimization framework:** the Hoopl [4] framework is a promising tool to implement any kind of data-flow analysis and optimization. Instead of developing our own crappy optimizer, we should use that stuff, when the source is released. This is the reason why @IL.Paka.Optimizer@ is such a joke: it must be dropped asap;

**Translation validation infrastructure:** because we claim dependability but our compiler is such a tricky mess, we need a good bodyguard. Translation validation [3] is an affordable technique that tells you, when you run your compiler, if it has barfed (and where), or not. If it has not failed, then you know for sure that the generated code is correct;
More stringent syntactic tests: it is very easy to build ill-formed Filet-o-Fish terms, because the types of constructs have not been engineered to ensure their invariants, and there is little or no run-time checks. It is just a matter of putting more run-time checks, a lot more;

Compiling to macros: that’s an interesting topic: we are able to generate C code. We might need to generate C macro at some point. How would that fit into Filet-o-Fish?

Compiling with assertions: assuming that Filet-o-Fish-generated C code is correct, we are ensured that it must never failed at run-time, except if it is provided with bogus input data. Being able to specify what is a valid input data and translating that into assertions might be useful. Similarly, when reading in an array, for example, we probably want to ensure that we are not going out of bounds, and an assert should fail if this is the case.
References