# fi

# Subtitle Text, if any

Name1
Affiliation1
Email1

Name2 Name3
Affiliation2/3
Email2/3

### **Abstract**

This is the text of the abstract.

Categories and Subject Descriptors CR-number [subcategory]: third-level

General Terms term1, term2

Keywords intersection types, inheritance

#### 1. Introduction

- Compare Scala: merge[A,B] = new A with B- type IEval = eval : Int type IPrint = print : String
- F[\_]
  We present a polymorphic calculus containing intersection types and records, and show how this language can be used to

solve various common tasks in functional programming in a nicer way.

Intersection types provides a power mechanism for functional

Intersection types provides a power mechanism for functional programming, in particular for extensibility and allowing new forms of composition.

Prototype-based programming is one of the two major styles

Prototype-based programming is one of the two major styles of object-oriented programming, the other being class-based programming which is featured in languages such as Java and C#. It has gained increasing popularity recently with the prominence of JavaScript in web applications. Prototype-based programming supports highly dynamic behaviors at run time that are not possible with traditional class-based programming. However, despite its flexibility, prototype-based programming is often criticized over concerns of correctness and safety. Furthermore, almost all prototype-based systems rely on the fact that the language is dynamically typed and interpreted.

In summary, the contributions of this paper are:

 elaboration typing rules which given a source expression with intersection types, typecheck and translate it into an ordinary F term. Prove a type preservation result: if a term e has type τ in the source language, then the translated term [e] is well-typed and has type [τ] in the target language.

- present an algorithm for detecting incoherence which can be very important in practice.
- explores the connection between intersection types and object algebra by showing various examples of encoding object algebra with intersection types.

# 2. A Taste of fi

<sup>1</sup>hange the examples later to something very simple.

This section provides the reader with the intuition of fi, while we postpone the presentation of the details in later sections.

In short, **fi** generalizes System F by adding intersection polymorphism. **fi** terms are elaborated into **f**, a variant of System F. System F, or polymorphic lambda calculus lays the foundation of functional programming languages such as Haskell.

The type system of  $\hat{\mathbf{n}}$  permits a subtyping relation naturally and enables prototype-based inheritance. We will explore the usefulness of such a type system in practice by showing various examples.

# 2.1 Intersection Types

The central addition to the type system of  $\mathbf{f}$  in  $\mathbf{fi}$  is intersection types. What is an intersection type? One classic view is from settheoretic interpretation of types: A & B stands for the intersection of the set of values of A and B. The other view, adopted in this paper, regards types as a kind of interface: a value of type A & B satisfies both of the interfaces of A and B. For example, eval: Int is the interface that supports evaluation to integers, while eval: Int & print: String supports both evaluation and pretty printing. Those interfaces are akin to interfaces in Java or traits in Scala. But one key difference is that they are unnamed in  $\mathbf{fi}$ .

Intersection types provide a simple mechanism for ad-hoc polymorphism, similar to what type classes in Haskell achieve. The key constructs are the "merge" operator, denoted by ",,", at the value level and the corresponding type intersection operator, denoted by, "\&" at the type level.

For example, we can define an (ad-hoc)-polymorphic show function that is able to convert integers and booleans to strings. In **fi** such function can be given the type

(Int -> String) & (Bool -> String)

and be defined using the merge operator,, as

let show = showInt ,, showBool

where showString and showBool are ordinary monomorphic functions. Later suppose the integer 1 is applied to the show function, the first component showInt will be picked because the type of showInt is compatible with 1 while showBool is not.

[Copyright notice will appear here once 'preprint' option is removed.]

 $\overline{{}^{1}C}$ 

**fi** 1 2015/2/22

# 2.2 Encoding Records

In addition to introduction of record literals using the usual notation, **fi** support two more operations on records: record elimination and record update.

A record type of the form  $\{l:\tau\}$  can be thought as a normal type t tagged by the label 1.

e1 and e2 are two expressions that support both evaluation and pretty printing and each has type eval: Int, print: String. add takes two expressions and computes their sum. Note that in order to compute a sum, add only requires that the two expressions support evaluation and hence the type of the parameter eval: Int. As a result, the type of e1 and e2 are not exactly the same with that of the parameters of add. However, under a structural type system, this program should typecheck anyway because the arguments being passed has more information than required. In other words, eval: Int, print: String is a subtype of eval: Int.

How is this subtyping relation derived? In **fi**, multi-field record types are excluded from the type system because <code>eval</code>: Int, print: String can be encoded as <code>eval</code>: Int & print: String. And by one of subtyping rules derives that <code>eval</code>: Int & print: String is a subtype of <code>eval</code>: Int.

### 2.3 Parametric Polymorphism

The presence of both parametric polymorphism and intersection is critical, as we shall see in the next section, in solving modularity problems. Here is a code snippet from the next section (The reader is not required to understand the purpose of this code at this stage; just recognizing the two types of polymorphism is enough.)

```
type SubExpAlg E = (ExpAlg E) \& { sub : E -> E -> E };
let e2 E (f : SubExpAlg E) = f.sub (exp1 E f) (f.lit 2);
```

SubExpAlg is a type synonym (a la Haskell) defined as the intersection of ExpAlg E and sub : E -> E -> E, parametrized by a type parameter E. e2 exhibits parametric polymorphism as it takes a type argument E.

# 3. Application

### 3.1 F-bounded Polymorphism

```
let getName A (v : A & {name:String}) = v.name in
getName [{name:String}] {name = "Ek", age = 10}

    Scala
def getName[A <: { def name: String }](v: A) = v.name
case class Person(name: String)
getName[Person](Person("Ek"))</pre>
```

This section shows that the System F plus intersection types are enough for encoding extensible designs, and even beat the designs in languages with a much more sophisticated type system. In particular, **fi** has two main advantages over existing languages:

- 1. It supports dynamic composition of intersecting values.
- It supports contravariant parameter types in the subtyping relation.

Various solutions have been proposed to deal with the extensibility problems and many rely on heavyweight language features such as abstract methods and classes in Java.

These two features can be used to improve existing designs of modular programs.

The expression problem refers to the difficulty of adding a new operations and a new data variant without changing or duplicating existing code.

There has been recently a lightweight solution to the expression problem that takes advantage of covariant return types in Java. We show that FI is able to solve the expression problem in the same spirit. The A)

### 3.2 Object Algebras

Object algebras provide an alternative to *algebraic data types* (ADT). For example, the following Haskell definition of the type of simple expressions

```
data Exp where
  Lit :: Int -> Exp
  Add :: Exp -> Exp -> Exp
```

can be expressed by the *interface* of an object algebra of simple expressions:

```
trait ExpAlg[E] {
  def lit(x: Int): E
  def add(e1: E, e2: E): E
}
```

Similar to ADT, data constructors in object algebras are represented by functions such as lit and add inside an interface ExpAlg. Different with ADT, the type of the expression itself is abtracted by a type parameter E.

which can be expressed similarly in fi as:

```
type ExpAlg E = {
   lit : Int -> E,
   add : E -> E -> E
}
```

Scala supports intersection types via the with keyword. The type A with B expresses the combined interface of A and B. The idea is similar to

```
interface AwithB extends A, B \{\} in Java. ^2
```

The value level counterpart are functions of the type A => B => A with B.  $^3$ 

Our type system is a simple extension of System F; yet surprisingly, it is able to solve the limitations of using object algebras in languages such as Java and Scala. We will illustrate this point with an step-by-step of solving the expression problem using a source language built on top of **fi**.

Oliveira noted that composition of object algebras can be cumbersome and intersection types provides a solution to that problem.

We first define an interface that supports the evaluation operation:

```
type IEval = { eval : Int };
type ExpAlg E = { lit : Int -> E, add : E -> E -> E };
let evalAlg = {
   lit = \(x : Int). { eval = x },
   add = \(x : IEval). \(y : IEval). { eval = x.eval + y.
        eval }
};
```

The interface is just a type synonym IEval. In  ${\bf fi}$ , record types are structural and hence any value that satisfies this interface is of type IEval or of a subtype of IEval.  $^4$ 

In the following, ExpAlg is an object algebra interface of expressions with literal and addition case. And evalAlg is an object algebra for evluation of those expressions, which has type ExpAlg Int

```
type SubExpAlg E = (ExpAlg E) & { sub : E -> E -> E }; let subEvalAlg = evalAlg ,, { sub = \ (x : IEval). \ (y : IEval). \ { eval = x.eval - y.eval } };
```

fi 2 2015/2/22

<sup>&</sup>lt;sup>2</sup> However, Java would require the A and B to be concrete types, whereas in Scala, there is no such restriction.

<sup>&</sup>lt;sup>3</sup> FIXME

<sup>&</sup>lt;sup>4</sup> Should be mentioned in S2.

Next, we define an interface that supports pretty printing.

```
type IPrint = { print : String };
let printAlg = {
  lit = \(x : Int). { print = x.toString() },
  add = \(x : IPrint). \(y : IPrint). { print = x.print.
        concat(" + ").concat(y.print) },
  sub = \(x : IPrint). \(y : IPrint). { print = x.print.
        concat(" - ").concat(y.print) }
};
```

Provided with the definitions above, we can then create values using the appropriate algebras. For example: defines two expressions.

The expressions are unusual in the sense that they are functions that take an extra argument f, the object algebras, and use the data constructors provided by the object algebra (factory) f such as lit, add and sub to create values. Moreover, The algebras themselves are abstracted over the allowed operations such as evaluation and pretty printing by requiring the expression functions to take an extra argument E.

If we would like to have an expression that supports both evaluation and pretty printing, we will need a mechanism to combine the evaluation and printing algebras. Intersection types allows such composition: the merge function, which takes two expression algebras to create a combined algebra. It does so by constructing a new expression algebra, a record whose each field is a function that delegates the input to the two algebras taken.

```
let newAlg = merge IEval IPrint subEvalAlg printAlg in
let o1 = e1 (IEval & IPrint) newAlg in
o1.print
```

o1 is a single object created that supports both evaluation and printing, thus achieving full feature-oriented programming.

# 3.3 Visitors

Constructing instances seems clumsy!

The visitor pattern allows adding new operations to existing structures without modifying those structures. The type of expressions are defined as follows:

```
trait Exp[A] {
  def accept(f: ExpAlg[A]): A
}
trait SubExp[A] extends Exp[A] {
  override def accept(f: SubExpAlg[A]): A
}
```

The body of Exp and SubExp are almost the same: they both contain an accept method that takes an algebra f and returns a value of the carrier type A. The only difference is at f — SubExpAlg [A] is a subtype of ExpAlg[A]. Since f appear in parameter position of accept and function parameters are contravariant, naturally we would hope that SubExp[A] is a supertype of Exp[A]. However, such subtyping relation does not fit well in Scala because inheritance implies subtyping in such languages <sup>5</sup>. As SubExp[A] extends Exp[A], the former becomes a subtype of the latter.

Such limitation does not exist in  ${\bf fi}$ . For example, we can define the similar interfaces Exp and SubExp:

```
type Exp A = { accept: forall A. ExpAlg A \rightarrow A }; type SubExp A = { accept: forall A. SubExpAlg A \rightarrow A };
```

Then by the typing judgment it holds that SubExp is a supertype of Exp. This relation gives desired results. To give a concrete example:

A is called is the *interpretation*. It works for any interpretation you want.

First we define two data constructors for simple expressions:

Suppose later we decide to augment the expressions with subtraction:

One big benefit of using the visitor pattern is that programmers is able to write in the same way that would do in Haskell. For example, e2 = sub (lit 2)(lit 3) defines an expression.

Another important property that does not exist in Scala is that programmer is able to pass lit 2, which is of type Exp A, to sub, which expects a SubExp A because of the subtyping relation we have. After all, it is known statically that lit 2 can be passed into sub and nothing will go wrong.

#### 3.4 Yanlin Stuff

This subsection presents yet another lightweight solution to the Expression Problem, inspired by the recent work by Wang. It has been shown that contravariant return types allows refinement of the types of extended expressions.

First, we define the type of expressions that support evaluation and implement two constructors:

If we would like to add a new operation, say pretty printing, it is nothing more than refining the original Exp interface by *intersecting* the original type with the new print interface using the & primitive and *merging* the original data constructors using the , , primitive.

Now we can construct expressions using the constructors defined above:

```
let e1: ExpExt = addExt (litExt 2) (litExt 3)
let e2: Exp = add (lit 2) (lit 4)
```

e1 is an expression capable of both evaluation and printing, while e2 supports evaluation only.

We can also add a new variant to our expression:

Finally we are able to manipulate our expressions with the power of both subtraction and pretty printing.

```
(subExt e1 e1).print
```

<sup>&</sup>lt;sup>5</sup> It is still possible to encode contravariant parameter types in Scala but doing so would require some technique.

### 3.5 Mixins

Mixins are useful programming technique wildly adopted in dynamic programming languages such as JavaScript and Ruby. But obviously it is the programmers' responsibility to make sure that the mixin does not try to access methods or fields that are not present in the base class.

In Haskell, one is also able to write programs in mixin style using records. However, this approach has a serious drawback: since there is no subtyping in Haskell, it is not possible to refine the mixin by adding more fields to the records. This means that the type of the family of the mixins has to be determined upfront, which undermines extensibility.

**fi** is able to overcome both of the problems: it allows composing mixins that (1) extends the base behavior, (2) while ensuring type safety.

The figure defines a mini mixin library. The apostrophe in front of types denotes call-by-name arguments similar to the => notation in the Scala language.

We define a factorial function in mixin style and make a noisy mixin that prints "Hello" and delegates to its superclass. Then the two functions are composed using the mixin and extends combinators. The result is the noisyFact function that prints "Hello" every time it is called and computes factorial.

# 4. The fi calculus

6

This section formalizes the syntax, subtyping, and typing of  ${\bf fi}$ . In the next section, we will go through the type-directed translation from  ${\bf fi}$  to System F.

# 4.1 Syntax

The syntax of the **fi** calculus extends System F by adding the two features: intersection types and records. The formalization includes only single records because single record types as the multi-records can be desugared into the merge of multiple single records.

Types  $\tau$  have five constructs. The first three are standard (present in System F): type variable  $\alpha$ , function types  $\tau \to \tau$ , and type abstraction  $\forall \alpha.\tau$ ; while the last two, intersection types  $\tau \& \tau$  and record types  $\{l:\tau\}$ , are novel in **fi**. In record types, l is the label and  $\tau$  the type.

First five constructs of expressions are also standard: variables x and two abstraction-elimination pairs.  $\lambda(x:\tau).e$  abstracts expression e over values of type  $\tau$  and is eliminated by application e e;  $\Lambda\alpha.e$  abstracts expression e over types and is eliminated by type application e  $\tau$ .

The last four constructs are novel. e, e is the *merge* of two terms.  $\{l = e\}$  introduces a record literal having l as the label

```
Types
               \alpha \mid \tau \to \tau \mid \forall \alpha. \tau
                                            Intersection type
               \tau \& \tau
                \{l:\tau\}
                                            Record type
Terms
  e
       :=
               x
               \Lambda \alpha.e \mid e \tau
                                        Type abstraction/application
               \lambda(x:\tau).e \mid e \mid e \mid e
                                        Term abstraction/application
                                        Merge
               e, e
               \{l=e\}
                                        Record introduction
               e.l
                                        Record elimination
               e with \{l=e\}
                                        Record update
Contexts
               \epsilon \mid \Gamma, \alpha \mid \Gamma, x:\tau
Labels
 1.
```

Figure 1: Syntax of fi

for field containing expression e. e.l access the field with label l in e. Finally, e with  $\{l=e_1\}$  is a new record which is exactly the same as e except the field labelled l is updated to become e1. For simplicity, we omit other constructs in order to focus on the essence of the calculus. For example, fixpoints can be added in standard ways.

The field F is non-standard and introduced to deal with records. It is an associative list. Each item is a pair whose first item is either empty or a label and the second the types.

### 4.2 Subtyping

Thanks to intersection types, we have natural subtyping relations among types. For example, Int & Bool should be a subtype of Int, since the former can be viewed as either Int or Bool. The subtyping rules are standard except for three points listed below:

- 1.  $\tau_1 \& \tau_2$  is a subtype of  $\tau_3$ , if either  $\tau_1$  or  $\tau_2$  are subtypes of  $\tau_3$ ,
- 2.  $\tau_1$  is a subtype of  $\tau_2 \& \tau_3$ , if  $\tau_1$  is a subtype of both  $\tau_2$  and  $\tau_3$ .
- 3.  $\{l_1:\tau_1\}$  is a subtype of  $\{l_2:\tau_2\}$ , if  $l_1$  and  $l_2$  are identical and  $\tau_1$  is a subtype of  $\tau_2$ .

The first point is captured by two rules (S-And-1) and (S-And-2), whereas the second point by (S-And-3). Note that the last point means that record types are covariant in the type of the fields.

#### 4.3 Typing

The typing judgment for  ${\bf fi}$  is of the form:  $\Gamma \vdash e: \tau$ . This judgment uses the context  $\Gamma$ . The typing rules for our core languages are mostly standard ones for System F. In particular we introduce a (T-Merge) rule that applies to merge constructs.

dom reads: "the domain of". F(l) means the result of lookup for l inside the associative list F. The order of lookup can be either from left to right or from right to left but has to be consistent inside one implementation. We prefer the order from the right to the left because it make possible record overriding. For example,  $(\{count = 1\}, \{count = 2\}).count$  will evaluate to 2 in this case.

### 5. Type-directed Translation to System F

In this section we define the semantics of **fi** by means of a typedirected translation to System F. This translation removes the labels

<sup>&</sup>lt;sup>6</sup> Joshua Dunfield

$$\begin{array}{c} \alpha < \alpha & (\text{SVAR}) \\ \frac{\tau_3 < \tau_1}{\tau_1 \rightarrow \tau_2 < \tau_3 \rightarrow \tau_4} & (\text{SFun}) \\ \\ \frac{\tau_1 < [\alpha_2 \mapsto \alpha_1] \tau_2}{\forall \alpha_1. \tau_1 < \forall \alpha_2. \tau_2} & (\text{SForall}) \\ \\ \frac{\tau_1 < [\alpha_2 \mapsto \alpha_1] \tau_2}{\forall \alpha_1. \tau_1 < \forall \alpha_2. \tau_2} & (\text{SAnd1}) \\ \\ \frac{\tau_1 < \tau_3}{\tau_1 \& \tau_2 < \tau_3} & (\text{SAnd2}) \\ \\ \\ \frac{\tau_1 < \tau_2}{\tau_1 \& \tau_2 < \tau_3} & (\text{SAnd2}) \\ \\ \\ \frac{\tau_1 < \tau_2}{\tau_1 < \tau_2 \& \tau_3} & (\text{SAnd3}) \\ \\ \\ \frac{\tau_1 < \tau_2}{\{l : \tau_1\} < \{l : \tau_2\}} & (\text{SRcd}) \end{array}$$

Figure 2: Subtyping

$$\frac{\alpha \in \Gamma}{\Gamma \vdash \alpha} \qquad \qquad \text{(WF-VAR)}$$
 
$$\frac{\Gamma \vdash \tau_1 \qquad \Gamma \vdash \tau_2}{\Gamma \vdash \tau_1 \rightarrow \tau_2} \qquad \qquad \text{(WF-FUN)}$$
 
$$\frac{\Gamma, \alpha \vdash \tau}{\Gamma \vdash \forall \alpha. \tau} \qquad \qquad \text{(WF-FORALL)}$$
 
$$\frac{\Gamma \vdash \tau_1 \qquad \Gamma \vdash \tau_2}{\Gamma \vdash \tau_1 \& \tau_2} \qquad \qquad \text{(WF-AND)}$$
 
$$\frac{\Gamma \vdash \tau}{\Gamma \vdash \{l:\tau\}} \qquad \qquad \text{(WF-RCD)}$$

Figure 3: Well-formedness

Figure 4: Typing

of records and turns intersections into products, much like Dunfield's elaboration. But our translation also deals with parametric polymorphism and records.

### 5.1 Informal Discussion

To help the reader have a high-level understanding of how the translation works, in this subsection we present the translation informally. Take the **fi** expression for example:

First, multi-field record literals are desugared into merges of single-field record literals. Therefore  $\{eval=4, print="4"\}$  becomes  $\{eval=4\}, \{print="4"\}$ . Merges of two values are translated into just a pair of them by (TrMerge) and single-field record literals lose their field labels by (TrRcdIntro). Hence  $\{eval=4\}, \{print="4"\}$  becomes  $\{4, "4"\}$ .

Finally, e1 and e2 are both coerced by a projection function (x:(Int,String)).fst x.

### 5.2 Target Language

The target language is System F extended with pairs. The syntax and typing is completely standard. The syntax of Systm F is as follows:

$$\begin{array}{lll} \text{Types} & T & \coloneqq & \alpha \mid T \to T \mid \forall \alpha.T \mid \langle T,T \rangle \\ \text{Terms} & E,C & \coloneqq & x \mid \lambda(x:T).E \mid \Lambda\alpha.E \mid E \mid E \mid E \mid T \\ & \mid \langle E,E \rangle \mid \texttt{fst}E \mid \texttt{snd}E \end{array}$$

The dynamic semantics of System F can be found in ...

### Lemma 1. If

$$\Gamma \vdash \tau_1 \mathrel{<:} \tau_2 \hookrightarrow C$$

then

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau_1 \rrbracket \rightarrow \llbracket \tau_2 \rrbracket$$

The main translation judgment is  $\Gamma \vdash e : \tau \hookrightarrow E$  which states that with respect to the environment  $\Gamma$ , the **fi** expression e is of a **fi** type  $\tau$  and its translation is a System F expression E.

We also define the type translation function  $[\![\cdot]\!]$  from **fi** types  $\tau$  to System F types T.

The first three rules of the translation is standard. For the last two, the intersection of two types are translated into a product of them, and the label of record types are erased.

The translation consists of four sets of rules, which are explained below:

# 5.3 Subtyping (Coercion)

The coercion judgment  $\Gamma \vdash \tau_1 \mathrel{<\!\!\!:} \tau_2 \hookrightarrow C$  extends the subtyping judgment with a coercion on the right hand side of  $\hookrightarrow$ . A coercion C is an expression in the target language and has type  $\tau_1 \to \tau_2$ , as proved by Lemma 1. It is read "In the environment  $\Gamma$ ,  $\tau_1$  is a subtype of  $\tau_2$ ; and if any expression e has a type  $\tau_1$  that is a subtype of the type of  $\tau_2$ , the elaborated e, when applied to the corresponding coercion C, has exactly type  $[\![\tau]\!]_2$ ". For example,  $\Gamma \vdash Int\&Bool <: Bool \hookrightarrow fst$ , where fst is the projection of a tuple on the first element. The coercion judgment is only used in the (TrApp) case. As (SFun) supports contravariant parameter type and covariant return type, the coercion of the parameter types and that of the return types are used to create a coercion for the function type. (SAnd1), (SAnd2), and (SAnd3) deal with intersection types. The first two are complementary to each other. Take (SAnd1) for example, if we know  $\tau_1$  is a subtype of  $\tau_3$  and C is a coercion from  $\tau_1$  to  $\tau_3$ , then we can conclude that  $\tau_1 \& \tau_2$  is also a subtype of  $\tau_3$  and the new coercion is a function that takes a value x of type  $\tau_1 \& \tau_2$ , project x on the first item, and apply C to it. (SAnd3) uses both of two coercions and constructs a pair.

### 5.4 Typing (Translation)

In this subsection we now present formally the translation rules that convert  $\mathbf{fi}$  expressions into System F ones. This set of rules essentially extends those in the previous section with the light-blue part for the translation.

#### Coercion

Explained in the previous subsection.

#### Translation

The elaboration judgment  $\Gamma \vdash e : \tau \hookrightarrow E$  extends the typing judgment with an elaborated expression on the right hand side of  $\hookrightarrow$ . It is also standard, except for the case of (TrApp), in which a coercion from the inferred type of the argument,  $e_2$ , to the expected type of the parameter,  $\tau_1$ , is inserted before the argument; (TrMerge) translates merges into pairs. (TrRcdIntro) uses the same System F expression E for e as for e as for e and in (TrRcdEim) and (TrRcdUpd) the coercions generated by the "get" and "put" rules will be used to coece the main e fi expression.

(TrRcdEl im) typechecks e and use the "get" rule to return the type of the field  $\tau_1$  and the coecion C. The type of the whole expression is  $\tau_1$  and its translation of C E.

(TrRcdUpd) is similar to (TrRcdElim) in that it uses the auxiliary "put" rule. This rule typechecks e and  $e_1$ , and uses the "put" rule. Note that it allows refining of types by an  $e_1$  that is of a subtype of  $\tau_1'$ , which is the type of the field l in e. The type of the updated expression then takes the type  $\tau'$  returned by the "put" rule, while its translation is E, applied to the coercion generated by the "put" rule, C.

The two set of rules are explained below.

### • "get" rules

The "get" judgment deals spefically with record elimination and yields a coercion can be thought as a field accessor. For example:

$$\Gamma \vdash_{get} (\{eval : Int\}, eval) : \{eval : Int\} \hookrightarrow \lambda(x:Int).x$$

The lambda is the field accessor and when applied to a translated expression of type  $\{eval:Int\}$ , it is able to give the desired field. (GetBase) is the base case: the type of the field labelled l in a  $\{l:\tau\}$  is just  $\tau$  and the coercion is an identity function specialized to type  $[\![\{l:\tau\}]\!]$  (GetLeft) and (GetRight) are complementary to each other.

Consider the source program:

```
({ name = ''Isaac'', age = 10 }).name
```

Multi-field records are desugared into merge of single-field records:

By (GetBase),

$$\vdash_{qet} (\{name:String\}; name) : String$$

we have the coercion

$$\lambda(x:[[name:String]]).x$$

which is just  $\lambda(x:String).x$  according to type translation. By (GetLeft),

$$\vdash_{qet} (\{name:String\} \& \{age:Int\}; name) : String$$

By typing rules, the translation of the program is

Figure 5: Type translation

$$\boxed{ \begin{aligned} & \alpha < \alpha \hookrightarrow \lambda(x \llbracket \alpha \rrbracket).x & \text{(SVAR)} \\ & \frac{\tau_3 < \tau_1 \hookrightarrow C_1}{\tau_1 \to \tau_2 < \tau_3 \to \tau_4 \hookrightarrow C_2} & \text{(SFun)} \\ & \frac{\tau_1 < \tau_2 < \tau_3 \to \tau_4 \hookrightarrow \lambda(f : \llbracket \tau_1 \to \tau_2 \rrbracket).\lambda(x : \llbracket \tau_3 \rrbracket).C_2 \left( f \left( C_1 \, x \right) \right)}{\nabla_1 \to \tau_2 < \tau_3 \to \tau_4 \hookrightarrow \lambda(f : \llbracket \tau_1 \to \tau_2 \rrbracket).\lambda(x : \llbracket \tau_3 \rrbracket).C_2 \left( f \left( C_1 \, x \right) \right)} & \text{(SForall)} \\ & \frac{\tau_1 < \llbracket \alpha_2 \hookrightarrow \alpha_1 \rrbracket \tau_2 \hookrightarrow C}{\forall \alpha_1.\tau_1 < \forall \alpha_2.\tau_2 \hookrightarrow \lambda(f : \llbracket \forall \alpha.\tau_1 \rrbracket).\Lambda\alpha.C \left( f \, \alpha \right)} & \text{(SAnd1)} \\ & \frac{\tau_1 < \tau_3 \hookrightarrow C}{\tau_1 \& \tau_2 < \tau_3 \hookrightarrow \lambda(x : \llbracket \tau_1 \& \tau_2 \rrbracket).C \left( f \, \text{st} \, x \right)} & \text{(SAnd2)} \\ & \frac{\tau_2 < \tau_3 \hookrightarrow C}{\tau_1 \& \tau_2 < \tau_3 \hookrightarrow \lambda(x : \llbracket \tau_1 \& \tau_2 \rrbracket).C \left( \, \text{snd} \, x \right)} & \text{(SAnd2)} \\ & \frac{\tau_1 < \tau_2 \hookrightarrow C_1}{\tau_1 < \tau_2 & \Delta C_3 \hookrightarrow \lambda(x : \llbracket \tau_1 \rrbracket).C \left( \, x , C_2 \, x \right)} & \text{(SAnd3)} \\ & \frac{\tau_1 < \tau_2 \hookrightarrow C}{\{l \colon \tau_1 \} \hookrightarrow \{l \colon \tau_2 \} \hookrightarrow \lambda(x : \llbracket \{l \colon \tau_1 \} \rrbracket).C \, x} & \text{(SRcD)} \end{aligned}$$

Figure 6: Coercion

. If we apply the coercion to it, we get

"Isaac''

# • "put" rules

The "put" judgment deals spefically with record update can be thought as producing a field updater. Compared to the "get" rules, the "put" rules take an extra input e, which is the desired expression to replace the field lablled l in values of type  $\tau$ . (PutBase) is the base case. This rule allows refinement of record fields in the sense that the type of e can be a subtype of the type of the field labelled by l. The resulting type is  $\{l:\tau'\}$  and the generated coercion is a constant function that always returns E. (PutLeft) and (PutRight) are complementary to each other: the idea is exactly the same as (GetLeft) and (GetRight) except that the refined type  $\tau_1'$  and  $\tau_2'$  is used.

# 5.5 Meta-theory

**Lemma 2.** Subtyping is reflexive Given a type  $\tau$ ,  $\tau <: \tau$ .

**Lemma 3.** Subtyping is transitive If  $\tau_1 < \tau_2$  and  $\tau_2 < \tau_3$ , then  $\tau_1 < \tau_3$ .

**Lemma 4** (Get rules produce the type-correct coercion). *If* 

$$\Gamma \vdash_{qet} \tau; l = C; \tau_1$$

then

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \rightarrow \llbracket \tau_1 \rrbracket$$

*Proof.* By induction on the given derivation.

Lemma 5 (Put rules produce the type-correct coercion). If

$$\Gamma \vdash_{put} \tau; l; E = C; \tau_1$$

then

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \to \llbracket \tau \rrbracket$$

*Proof.* By induction on the given derivation.

Lemma 6 (Translation preserves well-formedness). If

 $\Gamma \vdash \tau$ 

then

$$[\![\Gamma]\!]\vdash [\![\tau]\!]$$

*Proof.* By induction on the given derivation.

**Theorem 1** (Type preserving translation). *If* 

$$\Gamma \vdash e : \tau \hookrightarrow E$$

$$\begin{array}{c} \Gamma \vdash e : \tau \rightarrow E \\ \\ \Gamma \vdash x : \tau \leftrightarrow x \end{array} & (TRVAE) \\ \\ \Gamma \vdash x : \tau \leftrightarrow x \end{array} & (TRVAE) \\ \\ \Gamma \vdash x : \tau \leftrightarrow x \end{array} & (TRVAE) \\ \\ \Gamma \vdash x : \tau \leftrightarrow x \end{array} & (TRVAE) \\ \\ \Gamma \vdash \lambda (x\tau) \cdot e : \tau \rightarrow \tau_1 \leftrightarrow \lambda (x[\tau]) \cdot E \\ \\ \Gamma \vdash \lambda (x\tau) \cdot e : \tau \rightarrow E \\ \Gamma \vdash \lambda (x\tau) \cdot e : \tau \rightarrow E \\ \Gamma \vdash \lambda (x\tau) \cdot e : \tau \rightarrow E \\ \Gamma \vdash \lambda (x\tau) \cdot e : \tau \rightarrow E \\ \Gamma \vdash e : \tau_2 \leftrightarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_3 \leftrightarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash e : \tau_4 \leftarrow \lambda e \cdot E \\ \Gamma \vdash$$

Figure 7: Type system

$$\llbracket \Gamma \rrbracket \vdash E : \llbracket \tau \rrbracket$$

*Proof.* (Sketch) By structural induction on the expression and the corresponding inference rule. The full proof can be found in the appendix.  $\Box$ 

Type-Directed Translation to System F. Main results: type-preservation + coherence.

# 6. Implementation

# 6.1 Type Synonyms

We extend the implementation of the type system extended with type synonyms and lazy arguments.

type T A1 A2 = 
$$\dots$$
 in

# 6.2 Optimization

# 7. Related Work

Strictly speaking, the system we have described is not intersection in the set-theoretic sense, but rather "implicit pairs" whose introduction is explicit by using the merge operator and elimination is implicit.

### 7.1 Intersection Types with Parametric Polymorphism

Our type system combines intersection types and polymorphism. The closest work to ours is Pierce [3]'s work on a prototype compiler for a language with both intersection types, union types, and parametric polymorphism. The important difference with our system is that in his language there is no explicit introduction construct like our merge operator. However, as shown in Section 3, this feature is critical in supporting modularity and extensibility because it allows dynamic composition of values. Pierce does not include any proof.

# 7.2 Other Type Systems with Intersection Types

Dunfield [2] describes a similar approach to ours: compiling a system with intersection types into ordinary  $\lambda$ -calculus terms. But his system does not include parametric polymorphism, while ours does not include unions. Although similar in this spirit, our translation uses coercion explicitly in the case of term applications. Besides, our translation does not have the undeterministic merge rule. Instead, it is captured by the deteterministic (TrMerge) rule.

Reynolds invented Forsythe in the 1980s. Our merge operator is analogous to  $p_1, p_2$ . Castagna, and Dunfield describe elaborating multi-fields records into merge of single-field records. As Dunfield has noted, in Forsythe merges can be only used unambiguously. For instance, it is not allowed in Forsythe to merge two functions. Reynolds and Castagna do not consider elaboration and Dunfield do not formalize elaborating records.

Both Pierce and Dunfield's system include a subsumption rule, which states that if an expression has been inferred of type  $\tau$ , then it is also of any supertype of  $\tau$ . Our system does not have this rule.

# 7.3 Type Systems for Modularity

# 7.4 Extensible Records and Row Types

[1]

The idea of encoding records using intersection types is due to Castagna and Reynolds (1996). Although Dunfield also discusses this idea in his paper [2] and provides an implementation, he does not proceed to propose a set of elaboration typing rules for records.

### 7.5 Other Related Work

# • Elaborating simply-typed lambda calculus

Dunfield has introduced a type system with intersection polymorphism but no parametric polymorphism.

Nystrom et. al. OOPSLA 06

Applications:

- Object/Fold Algebras. How to support extensibility in an easier way.

See Datatypes a la Carte

- Mixins
- Lenses? Can intersection types help with lenses? Perhaps making the types more natural and easy to understand/use?
- Embedded DSLs? Extensibility in DSLs? Composing multiple DSL interpretations?

http://www.cs.ox.ac.uk/jeremy.gibbons/publications/embedding.pdf

# 8. Conclusion

### A. Proofs

*Proof.* By structural induction on the types and the corresponding inference rule.

(SubVar) (SubFun) (SubForall) (SubAnd1) (SubAnd2) (SubAnd3) (SubRcd)

# Lemma 7. If

$$\Gamma \vdash_{aet} \tau; l = C; \tau_1$$

then

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \rightarrow \llbracket \tau_1 \rrbracket$$

*Proof.* By structural induction on the type and the corresponding inference rule.

(Get-Base) 
$$\Gamma \vdash_{qet} \{l:\tau\}; l = \lambda(x:[[\{l:\tau\}]]).x; \tau$$

By the induction hypothesis

$$\llbracket \Gamma \rrbracket \vdash \lambda(x: \llbracket \{l:\tau\} \rrbracket) . x : \llbracket \{l:\tau\} \rrbracket \to \llbracket \tau \rrbracket$$

(Get-Left) (Get-Right)

# Lemma 8. If

$$\Gamma \vdash_{put} \tau; l; E = C; \tau_1$$

then

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \to \llbracket \tau \rrbracket$$

*Proof.* By structural induction on the type and the corresponding inference rule.

П

$$\Gamma \vdash \tau$$

then

$$[\![\Gamma]\!]\vdash [\![\tau]\!]$$

Proof. Since

$$\Gamma \vdash \tau$$

It follows from (FI-WF) that

$$ftv(()\tau) \subseteq ftv(()\Gamma)$$

And hence

$$ftv(()[\![\tau]\!]) \subseteq ftv(()[\![\Gamma]\!])$$

By (F-WF) we have

$$\Gamma \vdash \tau$$

Theorem 2 (Type preserving translation). If

$$\Gamma \vdash e : \tau \hookrightarrow E$$

then

$$\llbracket \Gamma \rrbracket \vdash E : \llbracket \tau \rrbracket$$

*Proof.* By structural induction on the expression and the corresponding inference rule.

(TrVar) 
$$\Gamma \vdash x : \tau \hookrightarrow x$$

It follows from (TrVar) that

$$(x:t) \in \Gamma$$

Based on the definition of  $[\![\cdot]\!]$ ,

$$(x: [t]) \in [\Gamma]$$

Thus we have by (F-Var) that

$$\llbracket \Gamma \rrbracket \vdash x : \llbracket \tau \rrbracket$$

(TrAbs) 
$$\Gamma \vdash \lambda(x:\tau_1).e:\tau_1 \to \tau_2 \hookrightarrow \lambda(x:\llbracket \tau_1 \rrbracket).E$$

It follows from (TrAbs) that

$$\Gamma, x : \tau_1 \vdash e : \tau_2 \hookrightarrow E$$

And by the induction hypothesis that

$$[\![\Gamma]\!], x : [\![\tau_1]\!] \vdash E : [\![\tau_2]\!]$$

By (TrAbs) we also have

$$\Gamma \vdash \tau_1$$

It follows from Lemma 9 that

$$\llbracket \Gamma \rrbracket \vdash \llbracket \tau_1 \rrbracket$$

Hence by (F-Abs) and the definition of  $[\cdot]$  we have

$$\llbracket \Gamma \rrbracket \vdash \lambda(x:\llbracket \tau_1 \rrbracket).E : \llbracket \tau_1 \to \tau_2 \rrbracket$$

(TrApp) 
$$\Gamma \vdash e_1 \ e_2 : \tau_2 \hookrightarrow E_1(C \ E_2)$$

From (TrApp) we have

$$\Gamma \vdash \tau_3 \mathrel{<:} \tau_1 \hookrightarrow C$$

Applying Lemma 1 to the above we have

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau_3 \rrbracket \rightarrow \llbracket \tau_1 \rrbracket$$

Also from (TrApp) and the induction hypothesis

$$[\![\Gamma]\!] \vdash E_1 : [\![\tau_1]\!] \to [\![\tau_2]\!]$$

Also from (TrApp) and the induction hypothesis

$$\llbracket \Gamma \rrbracket \vdash E_2 : \llbracket \tau_3 \rrbracket$$

Assembling those parts using (F-App) we come to

$$\llbracket \Gamma \rrbracket \vdash E_1(C E_2) : \llbracket \tau_2 \rrbracket$$

(TrTAbs)  $\Gamma \vdash \Lambda \alpha.e : \forall \alpha.\tau \hookrightarrow \forall \alpha.E$ 

From (TrTAbs) we have

$$\Gamma \vdash e : \tau \hookrightarrow E$$

By the induction hypothesis we have

$$\llbracket \Gamma \rrbracket \vdash E : \llbracket \tau \rrbracket$$

Thus by (F-TAbs) and the definition of  $\llbracket \cdot \rrbracket$ 

$$\Gamma \vdash \Lambda \alpha . E : \llbracket \forall \alpha . \tau \rrbracket$$

(TrTApp) 
$$\Gamma \vdash e \ \tau : [\alpha := t] \tau_1 \hookrightarrow E \ [\![\tau]\!]$$

From (TrTApp) we have

$$\Gamma \vdash e : \forall \alpha.\tau_1 \hookrightarrow E$$

And by the induction hypothesis that

$$\llbracket \Gamma \rrbracket \vdash E : \forall \alpha. \llbracket \tau_1 \rrbracket$$

Also from (TrTApp) and Lemma 9 we have

$$\llbracket \Gamma \rrbracket \vdash \llbracket \tau \rrbracket$$

Then by (F-TApp) that

$$\llbracket \Gamma \rrbracket \vdash E \llbracket \tau \rrbracket : [\alpha := \llbracket \tau \rrbracket] \llbracket \tau_1 \rrbracket$$

Therefore

П

$$[\![\Gamma]\!] \vdash E [\![\tau]\!] : [\![\alpha := \tau]\!] [\![\tau_1]\!]]$$

From (TrMerge) and the induction hypothesis we have

$$\llbracket \Gamma \rrbracket \vdash E_1 : \llbracket \tau_1 \rrbracket$$

and

$$\llbracket \Gamma \rrbracket \vdash E_2 : \llbracket \tau_2 \rrbracket$$

Hence by (F-Pair)

$$\llbracket \Gamma \rrbracket \vdash \langle E_1, E_2 \rangle : \langle \llbracket \tau_1 \rrbracket, \llbracket \tau_2 \rrbracket \rangle$$

Hence by the definition of  $[\cdot]$ 

$$\llbracket \Gamma \rrbracket \vdash \langle E_1, E_2 \rangle : \llbracket \tau_1 \& \tau_2 \rrbracket$$

(TrRcdIntro) 
$$\Gamma \vdash \{l = e\} : \{l:\tau\} \hookrightarrow E$$

From (TrRcdIntro) we have

$$\Gamma \vdash e : \tau \hookrightarrow E$$

And by the induction hypothesis that

$$[\![\Gamma]\!] \vdash E : [\![\tau]\!]$$

Thus by the definition of  $[\![\cdot]\!]$ 

$$\llbracket \Gamma \rrbracket \vdash E : \llbracket \{l : \tau \} \rrbracket$$

(TrRcdElim) 
$$\Gamma \vdash e.l : \tau_1 \hookrightarrow C \ E$$

From (TrRcdElim)

$$\Gamma \vdash e : \tau \hookrightarrow E$$

And by the induction hypothesis that

$$[\![\Gamma]\!] \vdash E : [\![\tau]\!]$$

Also from (TrRcdEim)

$$\Gamma \vdash_{get} e; l = C; \tau_1$$

Applying Lemma 4 to the above we have

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \to \llbracket \tau_1 \rrbracket$$

Hence by (F-App) we have

$$\llbracket \Gamma \rrbracket \vdash C \ E : \llbracket \tau_1 \rrbracket$$

From (TrRcdUpd)

$$\Gamma \vdash e : \tau \hookrightarrow E$$

And by the induction hypothesis that

$$[\![\Gamma]\!] \vdash E : [\![\tau]\!]$$

Also from (TrRcdUpd)

$$\Gamma \vdash_{put} \tau; l; E = C; \tau_1$$

Applying Lemma 5 to the above we have

$$\llbracket \Gamma \rrbracket \vdash C : \llbracket \tau \rrbracket \to \llbracket \tau \rrbracket$$

Hence by (F-App) we have

$$[\![\Gamma]\!] \vdash C \ E : [\![\tau]\!]$$

# Acknowledgments

Acknowledgments, if needed.

# References

- [1] L. Cardelli. Extensible records in a pure calculus of subtyping. Digital. Systems Research Center, 1992.
- [2] J. Dunfield. Elaborating intersection and union types. *Journal of Functional Programming*, 24(2-3):133–165, 2014.
- [3] B. C. Pierce. Programming with intersection types and bounded polymorphism. PhD thesis, Carnegie Mellon University Pittsburgh, PA, 1991