Intensional Polymorphism in Type-Erasure Semantics

Karl Crary Stephanie Weirich Greg Morrisett
Cornell University

Abstract

Intensional polymorphism, the ability to dispatch to different routines based on types at run time, enables a variety of advanced implementation techniques for polymorphic languages, including tag-free garbage collection, unboxed function arguments, polymorphic marshaling, and flattened data structures. To date, languages that support intensional polymorphism have required a type-passing (as opposed to type-erasure) interpretation where types are constructed and passed to polymorphic functions at run time. Unfortunately, type-passing suffers from a number of drawbacks; it requires duplication of constructs at the term and type levels, it prevents abstraction, and it severely complicates polymorphic closure conversion.

We present a type-theoretic framework that supports intensional polymorphism, but avoids many of the disadvantages of type passing. In our approach, run-time type information is represented by ordinary terms. This avoids the duplication problem, allows us to recover abstraction, and avoids complications with closure conversion. In addition, our type system provides another improvement in expressiveness; it allows unknown types to be refined in place thereby avoiding certain beta-expansions required by other frameworks.

1 Introduction

Type-directed compilers use type information to enable optimizations and transformations that are impossible (or prohibitively difficult) without such information [12, 10, 16, 1, 22, 24, etc.]. However, type-directed compilers for some languages such as Modula-3 and ML face the difficulty that some type information cannot be known at compile time. For example, polymorphic code in ML may operate on inputs of type $\alpha$ where $\alpha$ is not only unknown, but may in fact be instantiated by a variety of different types.

In order to use type information in contexts where it cannot be provided statically, a number of advanced implementation techniques process type information at run time [10, 16, 27, 18, 24]. Such type information is used in two ways: behind the scenes, typically by tag-free garbage collectors [27], and explicitly in program code, for a variety of purposes such as efficient data representation and marshalling [16, 10]. In this paper we focus on the latter area of applications.

To lay a solid foundation for programs that analyze types at run time, Harper and Morrisett devised an internal language, called $\lambda^i_{ML}$, that supports the first-class intensional analysis of types (following earlier work by Constable [2, 3]). The $\lambda^i_{ML}$ language and its derivatives were then used extensively in the high-performance ML compilers TIL/ML [26, 19] and FLINT [25]. The primary novelty of $\lambda^i_{ML}$ is the presence of “typecase” operators at the level of terms and types, that allow computations and type expressions to depend upon the values of other type expressions at run time.

Supporting intensional type analysis (and the use of type information at run time in general) seems to require semantics where types are constructed and passed to polymorphic functions during computation. However, there are a number of practical and theoretical reasons why type-passing is unattractive:

- A type-passing language such as $\lambda^i_{ML}$ requires that type information always be constructed and passed to polymorphic functions. This can result in considerable overhead if types are rarely examined at run time, and, as we discuss later, it makes abstraction impossible.
- Type passing results in considerable complexity in language semantics, due in large part to the number of semantic devices that must be duplicated for both terms and types. For example, in semantics that make memory allocation explicit [17, 18] a central device is a formal heap in which data is stored; in a type-erasure framework one such heap

---

1 “Intensional” since types are analyzed by the structure of their names, rather than by what terms they contain. This is critical for practicality.
In this paper we propose a typed calculus, called $\lambda_R$, that ameliorates the problems of type passing without sacrificing intensional type analysis. If run-time type dispatch is to be supported, then clearly on some level types must be passed. The fundamental idea behind our approach is to construct and pass terms that represent types instead of the types themselves. The connection between a type $\tau$ and its term representation $e$ is made in the static semantics by assigning $e$ the special type $R(\tau)$. Semantically, we may interpret $R(\tau)$ as a singleton type that contains only the representation of $\tau$.

This framework resolves the difficulties with type passing semantics discussed above. In particular, as representations of types are simply terms, we can use the pre-existing term operations to deal with run-time type information in languages and their semantics. Furthermore, we can eliminate the difficulties associated with polymorphic closure conversion, as we show in Section 4. Finally, our approach enables the choice not to pass representations. In turn, this allows us to eliminate the overhead of constructing and passing representations of types where it is not necessary.

Perhaps more importantly, the ability not to pass types allows abstraction and parametricity to be recovered. In most type systems abstraction may be achieved by hiding the identity of types either through parametric polymorphism [21] or through existential types [15]. However, when all types are passed and may be analyzed (as in $\lambda_m^{\mu}$), the identity of types cannot be hidden and consequently abstraction is impossible. In contrast, in $\lambda_R$ a type can be analyzed only when its representation is available at run time, so abstraction can be achieved simply by not supplying type representations.

For example, consider the type $\exists \alpha. \alpha$. When all types may be analyzed, this type implements a dynamic type; an expression of this type provides an object of some unknown type, and that unknown type’s identity can be determined at run time by analyzing $\alpha$. In $\lambda_R$, as in most other type systems, this implements an abstract type (in this particular example, a useless abstract type), because no representation of $\alpha$ is provided. Dynamic types are implemented in $\lambda_R$ by including a representation of the unknown type, as in $\exists \alpha. R(\alpha) \times \alpha$.

1.1 Expressiveness

In the interest of clarity of presentation, we express $\lambda_R$ as an extension of Harper and Morrisett’s $\lambda_m^{\mu}$ and focus on their differences. The principal difference is the restriction of type analysis to those types for which representations are provided. This does not diminish the expressiveness of our calculus; $\lambda_m^{\mu}$ may be translated in a straightforward syntax-directed manner into $\lambda_R$.

Moreover, the $\lambda_R$ calculus incorporates an additional improvement in expressiveness over $\lambda_m^{\mu}$ that is independent of explicit type passing: In $\lambda_m^{\mu}$, information gained by analyzing a type is not propagated to other variables having that type. Consequently, when analyzing a type $\alpha$ with the intention of processing an object with type $\alpha$, it is necessary to create a function with argument type $\alpha$ and then apply that function to the object of interest. In other words, the type system of $\lambda_m^{\mu}$ requires the use of beta-expansions that are not operationally necessary. In $\lambda_R$ we resolve this shortcoming by strengthening the typing rule for typecase so that it refines types in place.

1.2 Overview

The remainder of this paper is organized as follows: In Section 2 we review the $\lambda_m^{\mu}$ calculus. We then present, in Section 3, our $\lambda_R$ calculus and discuss its formal semantics, including representation terms, $R$-types, and the strengthened typecase rule. In Section 4, we discuss the simplification of polymorphic closure conversion by explicit type passing. We end with related work and conclusions in Sections 5 and 6. In the appendices we relate our typed semantics to an untyped one through type erasure (Appendix A), discuss the analysis of quasitified types (Appendix B), and provide the complete dynamic and static semantics.
2 Intensional Type Analysis

Suppose we wanted to store an array of booleans. Most computer architectures require that memory accesses are a word at a time, but it is a terrible waste of space to store booleans as integers. The solution is to “pack” 32 booleans into one word and use bit manipulations to retrieve the correct value. So, to subscript from a boolean array, we would use the following function:

\[
\text{val bitsub : array[\text{int}] \times \text{int}} \rightarrow \text{bool} = \text{fn (a,i) => sub(a,i div 32) \&\& (1<<(i mod 32))} \leftrightarrow 0
\]

This is fine when we know a given array contains boolean values, but we would like code that is polymorphic over the type of array to be able to use this mechanism. In essence, we need to define a new array constructor, PackedArray, which will produce an array of integers to hold booleans, and an ordinary array for other types. We also need an associated subscript operation, packedsub, so that when the array is of booleans, bitsub is called. This can be done with intensional type analysis, where in both cases an argument type is examined with a \text{typecase} form:

\[
\text{type PackedArray[\alpha] = Typecase } \alpha \text{ of bool} \rightarrow \text{array[\text{int}] | \_} \rightarrow \text{array[\alpha]}
\]

\[
\text{val packedsub : } \forall \alpha. \text{PackedArray[\alpha] \times \text{int}} \rightarrow \alpha = \text{Fn [\alpha] => typecase } \alpha \text{ of bool} \rightarrow \text{bitsub | \_} \rightarrow \text{sub}
\]

2.1 The \(\lambda^\text{ML}\) calculus

To formalize the tools of intensional type analysis, we begin by summarizing Harper and Morrisett’s \(\lambda^\text{ML}\) calculus [10]. The \(\lambda^\text{ML}\) calculus provides these tools in a form that is relatively simple, but already quite powerful.

The syntax of \(\lambda^\text{ML}\) appears below (modified slightly for presentation). The backbone of \(\lambda^\text{ML}\) is a predicative variant of Girard’s \(F_\omega\) [8, 7] in which the quantified type \(\forall \alpha: \kappa. \sigma\) ranges only over type constructors and “small” types (i.e., monotypes), which do not include the quantified types. The type analysis operators are \text{Typerec} and \text{typecase} at the constructor and term levels.

\[
\begin{align*}
\text{(kinds)} \quad & \kappa :: = \text{Type} | \kappa_1 \rightarrow \kappa_2 \\
\text{(con’s)} \quad & c :: = \alpha | \text{int} | c_1 \rightarrow c_2 | c_1 \times c_2 | \lambda \alpha: \kappa. c | c_1 \times c_2 \\
\text{Typerec} \quad & \text{c}_1(c_2, \ldots, c_n) \\
\text{(types)} \quad & \sigma :: = c_1 | c_1 \rightarrow c_2 | c_1 \times c_2 | \forall \alpha: \kappa. \sigma \\
\text{(terms)} \quad & e :: = x | i | \lambda \alpha: \sigma. e | \text{fix } f . \sigma. v | e_1 e_2 | (e_1, e_2) | \pi_1 e | \pi_2 e | \lambda \alpha: \kappa. v | e_1 | e_2 \\
\text{typecase}[\alpha, \sigma] | e | \text{e} \text{ of } int \Rightarrow e_{\text{int}} \\
& \beta \Rightarrow \gamma \Rightarrow e_{\text{int}} \\
& \beta \times \gamma \Rightarrow e_x
\end{align*}
\]

\[
\begin{align*}
\text{(values)} \quad & v :: = x | i | \lambda \alpha: \sigma. e | \text{fix } x: \sigma. v | (v_1, v_2) | \lambda \alpha: \kappa. v
\end{align*}
\]

Occasionally, for brevity, we will write \text{typecase} terms as \text{typecase}[\alpha, \sigma] | e | \text{e} \text{ of } \lambda \alpha: \kappa. v = \lambda \alpha: \kappa. v.

As an example of the use of type analysis in \(\lambda^\text{ML}\), consider the function \text{toString}, presented in Figure 1. With the addition of another base type, \text{string}, to the language, this function uses \text{typecase} to produce a string representation of a data object. For example, the call \text{toString[\text{int}]3} returns the string “3”. As we cannot provide any information about the implementation of functions, we just return the word “function” when one is encountered, as in the call:

\[
\text{toString[(\text{int} \rightarrow \text{int}) \times \text{int}] | (\lambda x: \text{int}. x + 1, 3)}
\]

which returns “(function, 3)”.

When the argument to \text{toString} is a product type, the function calls itself recursively. In this branch, the type variables \(\beta\) and \(\gamma\) are bound to the types of the first and second components of the tuple, so that the recursive call can be instantiated with the correct type.

\[
\text{fix toString} : (\forall \alpha. \text{Type. } \alpha \rightarrow \text{string}). \Lambda \alpha. \text{Type. typerec[\delta, \beta \rightarrow \text{string}] } \alpha \text{ of int \rightarrow int2string: string \Rightarrow lambda: string.obj} \\
& \beta \Rightarrow \gamma \Rightarrow lambda: (\beta \Rightarrow \gamma). \text{"function"} \\
& \lambda obj: (\beta \times \gamma). \text{"<\text{"(toString[\beta](\pi_1 obj))"\""} \\
& \text{"",""(toString[\gamma](\pi_2 obj))"\""} \\
\]

Figure 1: The function \text{toString}

The \text{typecase} form also has a type annotation to make it possible to type check it without type inference; the annotation \([\alpha, \sigma]\) indicates that given a type argument \(\tau\), the \text{typecase} computes a value with type \(\sigma[\tau/\alpha]\) (where this denotes the capture-avoiding substitution of \(\tau\) for \(\alpha\) in \(\sigma\)). In this example, each arm returns a function from \(\delta\) to \text{string}, where \(\delta\) is replaced by the appropriate
type, such as \(\text{int}\) in the \(\text{int}\) branch, and \(\beta \times \gamma\) in the product branch.

With this intuition, the typing rule for \text{typecase} is the natural one (but we will see that this is somewhat restrictive):

\[
\begin{align*}
\Gamma \vdash c : \text{Type} & \quad \Gamma, \alpha : \text{Type} \vdash \sigma : \text{Type} \\
\Gamma \vdash e_{\text{int}} : \sigma[\text{int}/\alpha] & \\
\Gamma, \beta : \text{Type}, \gamma : \text{Type} \vdash e_{\rightarrow} : \sigma[\beta \rightarrow \gamma/\alpha] \\
\Gamma, \beta : \text{Type}, \gamma : \text{Type} \vdash e_{\times} : \sigma[\beta \times \gamma/\alpha]
\end{align*}
\]

Often, to compute the result type \(\sigma\) of a \text{typecase} expression the constructor-level \text{Typerec} on the argument \(\alpha\) will be required. \text{Typerec} allows the creation of new types. It provides primitive recursion by calling its \text{fold} operation (or catamorphism) over inductively defined types. For this reason, \text{Typerec} is essentially a “fold” operation (or catamorphism) over inductively defined types. It provides primitive recursion by calling itself recursively on all of the components of the argument type. Also unlike \text{typecase}, where the branches explicitly bind arguments for the components of the type, the \(\rightarrow\) and \(\times\) branches of \text{Typerec} are constructor functions. For example, if the argument of a \text{Typerec} operation is \(c_1 \times c_2\), then that operation reduces to its \(\times\) branch (a constructor function of four arguments) applied to the components \(c_1\) and \(c_2\), and to the result of recursively computing the \text{Typerec} operation on those components.

\[
\begin{align*}
\text{Typerec}(c_1 \times c_2) & = \text{Typerec}(c_1, c_2) \\
\text{Typerec}(c_1, c_2) & = \text{Typerec}(c_1) \cdot \text{Typerec}(c_2)
\end{align*}
\]

The kinding rule for \text{Typerec} is again the natural one. To compute a constructor of kind \(\kappa\), present a type argument and three branches returning \(\kappa\) constructors:

\[
\begin{align*}
\Gamma \vdash c : \text{Type} & \quad \Gamma \vdash c_{\text{int}} : \kappa \\
\Gamma \vdash e_{\rightarrow} : \text{Type} \rightarrow \kappa & \quad \Gamma \vdash e_{\times} : \text{Type} \rightarrow \kappa \\
\Gamma \vdash e_{\times} : \text{Type} \rightarrow \kappa & \quad \Gamma \vdash e_{\times} : \text{Type} \rightarrow \kappa
\end{align*}
\]

The \(\lambda_R\) calculus

Figure 2 presents the syntax of \(\lambda_R\), which we describe in detail in the following section.

\[
\begin{align*}
\text{(kinds)} & \quad \kappa ::= \text{Type} | \kappa \rightarrow \kappa \\
\text{(con’s)} & \quad c ::= \text{int} | \alpha \times \kappa \cdot c_1 c_2 | c_1 \rightarrow c_2 \\
& \quad c_1 \times c_2 | \forall \alpha : \kappa \cdot c_1 c_2 | c_1 \rightarrow c_2 \ \text{Typerec} c(c_1, c_2, c_3, c_4, c_5, c_6)
\end{align*}
\]

3.1 Term Representations of Types

The key feature we add to the term language of \(\lambda_R\) is the representations of types as terms, which remain when the types themselves are ultimately erased. The base type, \(\text{int}\), has a corresponding representation constant \(\text{R_int}\). Likewise, inductive types have inductively defined representations; the type \(\text{int} \rightarrow \text{int}\) is represented by the term \(\text{R_{int}}(\text{R_{int}}, \text{R_{int}})\).

Accordingly, the argument to the term level \text{typecase} is the representation of a type, instead of a type. For example, if the argument \(e\) is of the form \(\text{R_{int}}(c_1, c_2)\), the arrow branch \(e_{\rightarrow}\) is taken. The type variables \(\beta\) and \(\gamma\) are still bound to the types that \(c_1\) and \(c_2\) represent, but, because we need not only the component types but also their representations, \(x\) and \(y\) are bound to \(c_1\) and \(c_2\). This is reflected in the following rule of the operational semantics:

\[
\begin{align*}
\text{typecase}[\delta, \beta, \gamma, x, y, e_{\rightarrow} \ldots] & = e_{\rightarrow}[\text{D}(\delta), \text{D}(\beta), \text{D}(\gamma), x, y]
\end{align*}
\]

The operation \(\text{D}(\cdot)\) in this rule converts a representation to the type that it denotes (Appendix, Figure 5). The rest of our dynamic semantics is formalized in Figure 4. It is presented as a call-by-value, small step operational semantics.

In order to assign a type to these representations, we have extended the type level of \(\lambda_R\) with the \(R\) construct, where the representation of a type \(\tau\) is given the type \(R(\tau)\), and extended the static semantics.
The static semantics we have defined ensures that these R-types are singular types; for each one there is exactly one value which inhabits it. This allows us to express constraints between types and their representations at a very fine level. For instance, in the tostring example, the representation argument must be the representation of the type of the object.

Furthermore, as we have added a new way to form types to the constructor language, we must add another term construct, \( R_B(\cdot) \), to form the representation of representation types. We also extend typecase with an extra branch to handle these terms and Typerec to handle R-types.

### 3.2 In-place Refinement of Types

The typing rules of \( \lambda^R \) quite often forces an inefficient use of typecase. In the tostring example in section 2, we were required to create closures in each of the branches of the typecase. It would be more efficient if we could lift the lambdas outside of the typecase and have each branch of the typecase return a string. We could then write this function as:

\[
\text{fix tostring} : (\forall \alpha. \text{Type}, R(\alpha) \rightarrow \alpha \rightarrow \text{string}).
\]

\[
\Delta : \text{Type}, \lambda x : \alpha. R(\alpha).
\]

\[
\text{typecase}\delta.x.x \alpha \text{of} R_{\text{int}} \Rightarrow \text{int2string}
\]

\[
R_{\text{string}} \Rightarrow \text{obj: string, obj}
\]

\[
R_{\Rightarrow}(x, y) \Rightarrow \beta \Rightarrow \gamma \Rightarrow
\]

\[
\text{obj; } \beta \times \gamma
\]

\[
\text{"<" (tostring } \beta \times x \text{ (} \tau_1 \text{ obj) })\}
\]

\[
\text{"", " (tostring } \gamma \times y \text{ (} \tau_2 \text{ obj) } \text{"} 
\]

The reason we could not do this in \( \lambda^{\text{MC}} \) is that we need the type of \( \text{obj} \) to change based upon which branch of the typecase is selected. All we know in the product branch is that \( \text{obj} \) is of type \( \alpha \), not a tuple. In order to project from it in the recursive calls, we need to update the type of \( \text{obj} \) to reflect the fact that we know that \( \alpha \) is \( \beta \times \gamma \) in the product branch.

With the right enhancement to the static semantics this is possible. We have held off the discussion of the \( \lambda^R \)’s typecase formation rule in order to emphasize this point. The basic idea is that in some cases typecase increases our knowledge of the argument type. We separate the formation rule into situations where typecase gives us new information, such as when the argument is of type \( R(\alpha) \), and when it does not, such as when the argument is of type \( R(\tau_1 \rightarrow \tau_2) \). In the inference rule for type checking a typecase term, when the argument is of type \( R(\alpha) \), we refine types containing \( \alpha \) to reflect the gain in information, as shown below. For simplicity, only some of the rule is included (the complete rule can be found in Figure 8):

\[
\begin{align*}
\Gamma, \alpha : \text{Type}, \Gamma' & \vdash e : R(\alpha) \\
\Gamma, \alpha : \text{Type}, \Gamma' & \vdash \text{e_int}[\text{int}/\alpha] : \text{c[int, } \Gamma / \alpha, \delta] \\
\Gamma, \beta : \text{Type}, \gamma : \text{Gamma}, \gamma' & \vdash \text{e}[\beta \rightarrow \gamma/\alpha] : \text{c[} \beta \rightarrow \gamma, \beta \rightarrow \gamma/\alpha, \delta]\end{align*}
\]

\[
\Gamma, \alpha : \text{Type}, \Gamma' \vdash \text{typecase}\delta.x.e(\text{e_int, } \beta \gamma x y, e, \ldots, ) : \text{c[} \alpha/\delta]
\]

For example, to typecheck the \( e_\ldots \) branch, we substitute \( \beta \times \gamma \) for \( \alpha \) everywhere, including the surrounding context.\(^2\) Consequently the types of the variables bound in the context will be refined by that substitution. Because \( \lambda^{\text{MC}} \) only makes this substitution in the return type of the branch, and not in the context, in order to propagate this information one must abstract over all variables of interest.

When we know more about the argument because of the singularity of the R-types, we can deduce statically which branch of the typecase will be taken. Therefore we do not need to typecheck the other branches at all, leading to a much simpler rule. For example, if we know the argument is of type \( R(\tau_1 \rightarrow \tau_2) \), we only need to examine the \( e_\ldots \) branch, as in the rule:

\[
\begin{align*}
\Gamma & \vdash e : R(\tau_1 \rightarrow \tau_2) \\
\Gamma & \vdash x R(\tau_1), y : R(\tau_2) \vdash e_\ldots[\tau_1, \tau_2/\beta, \gamma] : \text{c[} \tau_1 \rightarrow \tau_2/\delta]\end{align*}
\]

\[
\Gamma \vdash \text{typecase}\delta.x.e(\text{e_int, } \beta \gamma x y, e, \ldots, ) : \text{c[} \tau_1 \rightarrow \tau_2/\delta]
\]

### 3.3 Impredicativity

A final minor difference between \( \lambda_R \) and \( \lambda^{\text{MC}} \) is that we have chosen to make \( \lambda_R \) impredicative. We do this because our interest in \( \lambda_R \) is for compilation where it is difficult in some cases (such as typed closure conversion,\(^3\)

\(^2\)The substitution for \( \alpha \) is applied within the branches themselves in order to avoid creating a hole in the scope of \( \alpha \). In practice, a typechecker would implement this by a local type definition, rather than by substitution.
see Section 4) to avoid impredicativity. Also for the support of closure conversion we have added existential types.

So, unlike \( \lambda^u_h \), there is no distinction between types and constructors; everything is a constructor. Because of this increase in the number of constructors, \texttt{Typerec} must now include branches for universal and existential types. However, in order to retain strong normalization, this analysis is limited only to the outermost operator, and does not provide any information about the body of the type. We have also added \( R_6 \) and \( R_3 \) to the type level to represent these types, and have extended \texttt{typecase} to include branches for them. These constructs are essentially base representations and cannot provide any further information, so \texttt{typecase} as well is limited to the outermost operator. We discuss how to relax this in Appendix B.

### 3.4 Properties of the Formal Semantics

Formally, the static semantics of \( \lambda_R \) consists of a collection of rules for deriving judgments of the forms shown in Figure 3. In these judgments, \( \Gamma \) is a unified type and kind context, mapping either constructor variables \((\alpha, \beta, \ldots)\) to kinds, or term variables \((x, y, \ldots)\) to types. The full formal static and operational semantics of \( \lambda_R \) appear in Figures 4-9, and from them we can prove several useful properties about \( \lambda_R \).

First, we would like to prove the decidability of \( \lambda_R \) type-checking. The difficult part of the that is equivalence checking for constructors. Based upon the equivalence rules in Figure 7 we can define a notion of constructor reduction to a normal form in an obvious manner. This reduction relation can be proved to be strongly normalizing and confluent (in a manner similar to Morrisett [16]) from which it follows that constructor equivalence is decidable. Therefore we can state the following theorem:

**Theorem 3.1 (Decidability)** It is decidable whether or not \( \Gamma \vdash e : \tau \) is derivable in \( \lambda_R \).

Next, we would like to show that the static semantics guarantees safety: that is, if a term type checks, then the operational semantics will not get \textit{stuck}, where a term that is not a value, and for which no rule of our operational semantics applies, is \textit{stuck}:

**Theorem 3.2 (Type Safety)** If \( \emptyset \vdash e : \tau \) and \( e \mapsto^* e' \) then \( e' \) is not stuck.

The proof of this theorem is standard, relying on the usual progress, subject reduction and substitution lemmas.

### 4 Polymorphic Typed Closure Conversion

As a final example, we consider typed closure conversion in a \( \lambda_R \)-like framework. The key idea behind closure conversion is to shift from a substitution-based model of execution to an environment-based model via a source-to-source translation. In particular, all functions are replaced with explicit closures which are represented within the language as pairs consisting of a \( \lambda \)-abstraction (the code of the closure), and a tuple (the environment of the closure). The environment contains values for the free variables of the function. The code abstracts the environment as well as the arguments of the function and is thus closed. Hence, the code may be hoisted to the top-level, allocated at compile time, and shared among all substitution instances. Application is rewritten so that the code of a closure is first applied to its environment and then to its arguments.

In the monomorphic case no discrepancy arises between type-passing [14] and type-erasure [20] closure conversion. An existential type is used to hold the type of the closure’s environment abstract, so a closure for a \( \tau_1 \rightarrow \tau_2 \) function is given the type \( \exists \alpha.((\tau_1 \times \alpha) \rightarrow \tau_2) \times \alpha \).

However, with the introduction of polymorphism, significant differences arise between type-passing and type-erasure. The issue stems from the fact that functions may contain not only free value variables, but also free type variables, and closed code must abstract these as well. Closures must then provide somehow for applying such code to the appropriate type variables. In a type-erasure setting, application to type arguments has no run-time effect, so the partial application of code to the appropriate type variables may be performed when closures are created. Consequently, the possibility of free type variables does not figure into the type of a closure, and so closures have the same type \( \exists \alpha.((\tau_1 \times \alpha) \rightarrow \tau_2) \times \alpha \) as before.

However, in a type-passing semantics, the application to type arguments is a run-time operation and so such applications must be suspended until the closure is called. Thus, it is necessary for the closure to include a type environment as well as a value environment. The kind of the type environment must be hidden (as did the type of the value environment in the monomorphic case), and the closure’s type must enforce the requirement that the code be applied only to the proper type environment (see Minamide \textit{et al.} [14] for detailed explanations of why). The former requires the use of abstract kinds and the latter requires the use of translucent types [9]. This results in a closure having the considerably more complicated type (again, see Minamide \textit{et al.} [14] for a

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Gamma \vdash c : \kappa )</td>
<td>( c ) is a valid constructor of kind ( \kappa )</td>
</tr>
<tr>
<td>( \Gamma \vdash c_1 = c_2 : \kappa )</td>
<td>( c_1 ) and ( c_2 ) are equal constructors</td>
</tr>
<tr>
<td>( \Gamma \vdash e : \tau )</td>
<td>( e ) is a term of type ( \tau )</td>
</tr>
</tbody>
</table>

**Figure 3: Judgments of \( \lambda_R \)**
formalization of the necessary type theory):

$$\exists \text{tenv}. \text{Kind} \land \exists \text{tenv}. \exists \text{Type} \land \exists \text{tenv}. \land \exists \text{tenv}. \land \land \land \land$$

$$(\forall \gamma: \text{tenv} = \beta \land \text{tenv} \land \text{tenv} \land \text{tenv} \land \text{tenv} \land \text{tenv}) \land \land \land \land$$

In the above type, \(k\text{tenv}\) abstracts the kind of the type environment, \(\alpha\text{tenv}\) abstracts the type of the value of the value environment, and \(\beta\text{tenv}\) provides the type environment. The code type then takes a type environment \(\gamma\) of kind \(k\text{tenv}\) as an argument, but \(\gamma\) is constrained (using translucent types) to be the appropriate environment, \(\beta\text{tenv}\).

Since our framework is one of type-erasure, type environments may be resolved by partial application, resulting in the simpler type for closures. However, it is instructive to examine the details. Suppose the function to be closure converted is the function \(f\) with type \(\tau_1 \rightarrow \tau_2\) and suppose further that the function contains free occurrences of the type variable \(\alpha\) and its representation \(\text{tenv} R(\alpha)\).

First the function is rewritten in closed form as:

$$f' = \lambda \alpha. (\tau_1 \times R(\alpha)) \rightarrow \tau_2$$

Then (at run time) \(f'\) is instantiated with the type environment (that is, \(\alpha\)): \(f'' = (\tau_1 \times R(\alpha)) \rightarrow \tau_2 = f'[\alpha]\)

Finally, a closure is created:

$$f''' = \text{pack}(f'', \text{tenv}) \text{ as } \exists \beta. ((\tau_1 \times \beta) \rightarrow \tau_2) \times \beta$$

Consider what has become of the mechanisms for type-passing closure conversion: The type of \(f''\) requires that it be applied (for its second argument) only to the representation of \(\alpha\). So the translucency mechanism appears again, suggesting that translucency is inherent in type-passing closure conversion. However, this version of translucency has two advantages; the necessary type theory is simpler, and the translucency is completely hidden by the existential packaging in the eventual closure. On the other hand, abstract kinds do not appear in the process, suggesting them to be an artifact of true type-passing (but see Appendix B).

5 Related Work

Closely related to our work is the work of Minamide on lifting of type parameters for tag-free garbage collection [13]. Minamide was interested in lifting type parameters out of code so they could be preallocated at compile time. His lifting procedure required the maintenance of interrelated constraints between type parameters to retain type soundness, and he used a system similar to ours that makes explicit the passing of type parameters in order to simplify the expression of such constraints. The principal difference between Minamide’s system and ours is that Minamide did not consider intensional type analysis or first-class polymorphism. Minamide’s system also makes a distinction between type representations (which he calls evidence, following Jones [11]) and ordinary terms, while \(\lambda R\) type representations are fully first-class.

The issue of type parameter lifting is an important one for compilers based on \(\lambda R\). The construction of type representations at run time would likely lead to significant cost and should in practice be lifted out to compile time whenever possible. (Unfortunately, in the presence of polymorphic recursion, which \(\lambda R\) supports, it cannot always be possible.) Mechanisms for lifting such lifting have been developed by Minamide (in the work discussed above) and by Saha and Shao [23].

Dubois et al. [5] also pass explicit type representations to polymorphic functions when compiling ad-hoc polymorphism. However, their system differs from ours and Minamide’s in that no mechanism is provided for connecting representations to the types they denote, and consequently, information gained by analyzing type representations does not propagate into the type system.

Duggan [6] proposes another typed framework for intensional type analysis that is similar in some ways to \(\lambda^{LIC}\). Like \(\lambda^{LIC}\), Duggan’s system passes types implicitly and allows for the intensional analysis of types at the term level. Duggan’s system does not support intensional type analysis at the constructor level, as \(\lambda^{LIC}\) and \(\lambda R\) do, but it adds a facility for defining type classes (using union and recursive kinds) and allows type analysis to be restricted to members of such classes.

6 Conclusions and Future Directions

We have presented a type-theoretic framework that supports the passing and analysis of type information at run time, but that avoids the shortcomings associated with previous such frameworks (e.g., duplication of constructs, lack of abstraction, and complication of closure conversion). This new framework makes it feasible to use intensional type analysis in settings where the shortcomings previously made it impractical.

For example, Morrisett et al. [20] developed typing mechanisms for low-level intermediate and target languages that allow type information to be used all the way to the end of compilation. It would be desirable, in a system based on those mechanisms, to be able to exploit fully that type information using intensional type analysis. Unfortunately, the shortcomings of type-passing semantics made it incompatible with some of those low-level typing mechanisms. This unfortunate incompatibility has made it infeasible to use the mechanisms of Morrisett et al. in type-analyzing compilers such as TIL/ML [26, 19] and FLINT [25], and has made it infeasible to use intensional type analysis in the end-to-end typed compiler TALC [20]. This framework in this paper makes it possible to unify these two lines of work for the first time.

In pursuance of this aim, an important direction for future work is to extend the mechanisms of \(\lambda R\) into lower-
level typed intermediate languages such as typed assembly language [20]. Among the issues to be explored in such research is how to analyze the more complicated types used in typed assembly language, and how to perform type-directed dispatch without an atomic \texttt{case} construct. Another issue to explore is better mechanisms for analysis of quantified types (some initial ideas appear in Appendix B), and whether such mechanisms are merited in practice.

Another important question is whether a parametricity theorem like that of Reynolds [21] can be shown for $\lambda_R$. Polymorphism is clearly non-parametric in $\lambda^{\text{MC}}$, but the lowering of type analysis to explicit term-level representatives makes it plausible that some sort of parametricity could be shown for $\lambda_R$. In other words, we discussed at an intuitive level in Section 1 how the explicit passing of types restores the ability to abstract types that was discarded by $\lambda^{\text{MC}}$; it would be interesting to explore how that intuition may be formalized.

References


A Untyped Calculus

Although the formal static and operational semantics for $\lambda_R$ are for a typed language, we would like to emphasize the point that types are unnecessary for computation and can safely be erased. To do this we exhibit an untyped language, $\lambda_R^0$, a translation of $\lambda_R$ to this language through type erasure, and the following theorem, which states that execution in the untyped language mirrors execution in the typed language:

**Theorem A.1**

1. If $e_1 \mapsto^* e_2$ then $e_1^0 \mapsto^* e_2^0$.

2. If $\not\vdash e : \tau$ and $e_1^0 \mapsto^* u$ then there exists $e_2$ such that $e_1 \mapsto^* e_2$ and $e_2^0 = u$.

From this theorem and type safety for $\lambda_R$ it follows that our untyped semantics is safe.

**Corollary A.2**

If $\not\vdash e : \tau$ and $e^0 \mapsto^* u$ then $u$ is not stuck.

A.1 Syntax of Untyped Calculus

\[
\begin{align*}
\text{(terms) } u & \ ::= \ x \mid i \mid \lambda x. u \mid \text{fix } f. w \mid u_1 u_2 \\
& \quad \mid \pi_1 u \mid \pi_2 u \\
& \quad \mid R_{\text{int}} \mid R_{\text{int}} (u_1, u_2) \mid R_{\text{int}} (u) \mid R_v \mid R_3 \\
\text{typecase } u \mid R_{\text{int}} \Rightarrow u_{\text{int}} \\
& \quad R_{\text{int}} (x, y) \Rightarrow u_{\text{int}} \\
& \quad R_v (x, y) \Rightarrow u_x \\
& \quad R_3 (x) \Rightarrow u_R \\
& \quad R_v \Rightarrow u_v \\
& \quad R_3 \Rightarrow u_3 \\
\text{(values) } w & \ ::= \ x \mid i \mid \langle w_1, w_2 \rangle \mid \lambda x. u \mid \text{fix } f. w \\
& \quad \mid R_{\text{int}} \mid R_{\text{int}} (w_1, w_2) \mid R_{\text{int}} (w_1, w_2) \\
& \quad \mid R_v (w) \mid R_v \mid R_3
\end{align*}
\]

A.2 Type Erasure

\[
\begin{align*}
x^0 & = x \\
i^0 & = i \\
(e_1, e_2)^0 & = \langle e_1^o, e_2^o \rangle \\
(\pi_1 e)^0 & = \pi_1 e^o \\
(\lambda x. e)^0 & = \lambda x. e^o \\
(\lambda \alpha : \kappa. v)^0 & = v^o \\
(\text{fix } f. c.v)^0 & = \text{fix } f. v^o \\
(e_1 e_2)^0 & = e_1^o e_2^o \\
e[c]^0 & = e^o \\
\text{pack } e & \text{ as } \text{hiding } c^0 = e^0 \\
\text{unpack } (\alpha, x) & = e_1 \text{ in } e_2^o = (\lambda x. e_2^o) e_1^o \\
R_{\text{int}}^o & = R_{\text{int}} \\
R_{\text{int}} (e_1, e_2)^o & = R_{\text{int}} (e_1^o, e_2^o) \\
R_v (e_1, e_2)^o & = R_v (e_1^o, e_2^o) \\
R_{\text{int}} (e_1)^o & = R_{\text{int}} (e_1^o) \\
R_{\text{int}} (c)^o & = R_v \\
R_3 (c)^o & = R_3 \\
(\text{typecase } \delta. c \text{ of } e)^o & = \text{typecase } e^o \text{ of } e \\
R_{\text{int}} \Rightarrow e_{\text{int}} & = R_{\text{int}} \Rightarrow e_{\text{int}}^o \\
R_{\text{int}} (x, y) \text { as } e_{\text{int}} & = R_{\text{int}} (x, y) \Rightarrow e_{\text{int}}^o \\
(\beta \rightarrow \gamma) \Rightarrow e_{\text{int}} & = R_v (x, y) \Rightarrow e_{\text{int}}^o \\
R_v (x, y) \text{ as } R_{\text{int}} (x) \Rightarrow e_{\text{int}} & = R_{\text{int}} (x) \Rightarrow e_{\text{int}}^o \\
(\beta \times \gamma) \Rightarrow e_{\text{int}} & = R_v (x) \Rightarrow e_{\text{int}}^o \\
R_{\text{int}} (x) \text{ as } R_{\text{int}} (\beta) \Rightarrow e_{\text{int}} & = R_3 \Rightarrow e_{\text{int}}^o \\
R_v \Rightarrow e_{\text{int}} & = R_3 \Rightarrow e_{\text{int}}^o \\
R_3 \Rightarrow e_{\text{int}} & = R_3 \Rightarrow e_{\text{int}}^o
\end{align*}
\]

A.3 Operational Semantics of $\lambda_R^0$

\[
\begin{align*}
(\lambda x. c.w) w & \mapsto u[w/x] \\
(\text{fix } f. c.w) w' & \mapsto (w[\text{fix } f. c. w / f]) w' \\
\pi_1 (w_1, w_2) & \mapsto w_1 \\
\pi_2 (w_1, w_2) & \mapsto w_2 \\
\text{typecase } R_{\text{int}} (u_{\text{int}}, x, y. u.x, x, y. u_{\text{int}}) & \mapsto u_{\text{int}}
\end{align*}
\]
types at any other kind $\kappa$ would be built by an analogous constructor $R_{\forall \alpha}$ or $R_{\exists \alpha}$. Note that this requires an infinite collection of such constructors, two for each kind.

Analysis of types is also straightforward if quantification is restricted to $\text{Type}$, or to any finite set of kinds. In that case, the appropriate branches can be added to the Typerec and typecase operations, and each branch would return the appropriate representation function discussed above. General quantification could be handled, at the expense of additional complexity, by adding kind variables, term representations of kinds, and a Kindrec facility for analyzing such representations. With such additions, the quantifier branches would return a kind representation and the appropriate representation function.

However, with such a mechanism in place, the amount of useful analysis that can be performed is still quite limited. Type analyzing code may apply the representation function to an argument and analyze its output, but it could not analyze the function itself. This makes it impossible, for example, to print quantified types. The modal type theory of Despeyroux et al. [4] is intended for precisely this sort of application; it provides mechanisms for primitive recursion on higher-order syntax and might provide a solution to this problem.

Also, allowing any analysis of quantified types by Typerec sacrifices strong normalizations of type expressions. (An isomorphism between Type and Type $\to$ Type can be built, permitting the encoding of the untyped lambda calculus.) A modal type discipline holds some promise of a solution to this problem as well.

### B Analysis of Quantified Types

The analysis of quantified types in $\lambda_R$, as presented in this paper, is limited to the outermost operator; that is, a Typerec or typecase will determine that a type is a $\forall$ or $\exists$ type, but will not provide any information about the body of the type. The principal reasons for this decision are that support for full analysis of quantified types detracts from the elegance of the type theory, and that experience from the TIL/ML compiler suggests that such support may not be needed in practice. (In fact, the TIL/ML compiler uses a variant of $\lambda_{\forall,\exists}$ that handles quantified types in a manner relatively similar to $\lambda_R$.) In this section we briefly explore what would be required to extend $\lambda_R$ to support full analysis of quantified types.

Adding term-level representations for quantified types is straightforward. The representation for $\forall \alpha : \kappa . c$ must carry information expressing how to produce the appropriate $c$ given an appropriate $\alpha$ of kind $\kappa$. For example, a member of $R(\forall \alpha : \text{Type} . c)$ would be built by a constructor $R_{\forall \alpha}$ that takes a function with type $\forall \alpha : \text{Type} . R(\alpha) \to R(c)$. Representations of polymorphic
(λx.c) v → e[v/x]  (λλx.e) [c] → v[c/α]  π_1(v_1, v_2) → v_1  π_2(v_1, v_2) → v_2
(fix f.e) v' → (v[fix f.e.v/f]) v'  (fix f.e') [c] → (v[fix f.e.v/f])[c']

unpack (α, x) = (pack v as 3β.c1 hiding e_2) in e_2 → e_2[c_2, v/α, x]

typecase[δ.c] R_int (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_int

typecase[δ.c] (R_0 (e')) (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_v

typecase[δ.c] (R_3 (e')) (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_3

typecase[δ.c] (R_0 (v, e)) (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_int [D(v), D(e), v_1, v_2/β, γ, x, y]

typecase[δ.c] (R_0 (v, e)) (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_int [D(v), D(e), v_1, v_2/β, γ, x, y]

typecase[δ.c] (R_0 (v)) (e_int, βγxy.e_1, βγxy.e_2, e_v, e_z) → e_int [D(v), D(e), v/β, x]

\[
\begin{array}{cccc}
| e_1 \mapsto e'_1 & e \mapsto e' & e \mapsto e'[c] & e \mapsto e' [\pi_1 \mapsto \pi_1 e'] \\
| e_1 e_2 \mapsto e'_1 e'_2 & v e \mapsto v e' & e[c] \mapsto e'[c] & \langle e_1, e_2 \rangle \mapsto \langle e'_1, e'_2 \rangle \\
| e \mapsto e' & e_1 \mapsto e'_1 & \langle v, e \rangle \mapsto \langle v, e' \rangle & e \mapsto e' \\
\end{array}
\]

pack e as 3β.c1 hiding e_2 → pack e' as 3β.c1 hiding e_2

unpack (α, x) = e in e_2 → unpack (α, x) = e' in e_2

\[
\begin{array}{cccc}
| e \mapsto e' & e_1 \mapsto e'_1 & R_0 (v, e) \mapsto R_0 (v, e') & R_0 (e_1, e_2) \mapsto R_0 (e'_1, e'_2) \\
| R_0 (e_1, e_2) \mapsto R_0 (e'_1, e'_2) & R_0 (v, e) \mapsto R_0 (v, e') & R_0 (e) \mapsto R_0 (e') & e \mapsto e' \\
\end{array}
\]

Figure 4: Operational Semantics for \( \lambda_R \)

\[
\begin{align*}
D(R_{int}) &= \text{int} \\
D(R_x(e_1, e_2)) &= D(e_1) \times D(e_2) \\
D(R_0(e_1, e_2)) &= D(e_1) \rightarrow D(e_2) \\
D(R_0(e)) &= R(D(e)) \\
D(R_0(c)) &= c \\
D(R_3(c)) &= c \\
\end{align*}
\]

Figure 5: Translating representations to types
Figure 6: Constructor formation

Figure 7: Constructor equivalence (selected rules)
Γ ⊢ e : c

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Γ ⊢ i : int</td>
<td>(int)</td>
</tr>
<tr>
<td>Γ ⊢ x : c</td>
<td>(Γ(x) = c) (var)</td>
</tr>
<tr>
<td>Γ ⊢ e1 : c1, Γ ⊢ e2 : c2</td>
<td>(pair)</td>
</tr>
<tr>
<td>Γ ⊢ e : c1 × c2</td>
<td>(πi : e_i)</td>
</tr>
<tr>
<td>Γ ⊢ e1 : c1 → c, Γ ⊢ e2 : c2</td>
<td>(app)</td>
</tr>
<tr>
<td>Γ ⊢ e[Γ(e) = c]</td>
<td>(tapp)</td>
</tr>
<tr>
<td>Γ, α : k :: e : c</td>
<td>(fni)</td>
</tr>
<tr>
<td>Γ, α : k :: e : c</td>
<td>(λf : e_i : c \ → c)</td>
</tr>
<tr>
<td>Γ, α : k ⊢ e : c_i : k</td>
<td>(pack)</td>
</tr>
<tr>
<td>Γ ⊢ pack e as θα : k hiding e' : θα : k</td>
<td>(unpack)</td>
</tr>
<tr>
<td>Γ ⊢ R_{int} : R(int)</td>
<td>(rep_{int})</td>
</tr>
<tr>
<td>Γ ⊢ e : R(c)</td>
<td>(rep_r)</td>
</tr>
<tr>
<td>Γ ⊢ R_α(e) : R(R(c))</td>
<td>(rep_α)</td>
</tr>
<tr>
<td>Γ ⊢ e : R(vα : k : c) :: Type</td>
<td>(rep_v)</td>
</tr>
<tr>
<td>Γ ⊢ R_α(vα : k : c) : R(vα : k : c)</td>
<td>(rep_α)</td>
</tr>
<tr>
<td>Γ ⊢ e : c'</td>
<td>(equiv)</td>
</tr>
</tbody>
</table>

Figure 8: Term formation (except typecase)
Γ ⊢ e : e

Γ ⊢ e : R(int)  Γ ⊢ e_{int} : c[int/β]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[int/δ] (tcase_{int})

Γ ⊢ e : R(c_1 → c_2)  Γ, x:R(c_1), y:R(c_2) ⊢ e_{δ}[c_1, c_2/β, γ] : c[c_1 → c_2/δ]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[c_1 → c_2/δ] (tcase_{→})

Γ ⊢ e : R(c_1 × c_2)  Γ, x:R(c_1), y:R(c_2) ⊢ e_x[c_1, c_2/β, γ] : c[c_1 × c_2/δ]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[c_1 × c_2/δ] (tcase_{×})

Γ ⊢ e : R(R(c'))  Γ, x:R(c') ⊢ e_R[c'/β] : c[R(c')/δ]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[R(c')/δ] (tcase_{R})

Γ ⊢ e : R(∀α:k.e')  Γ ⊢ e_{v} : c[∀α:k.e'/δ]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[∀α:k.e'/δ] (tcase_{∀})

Γ ⊢ e : R(∃α:k.e')  Γ ⊢ e_{∃} : c[∃α:k.e'/δ]

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[∃α:k.e'/δ] (tcase_{∃})

Γ, α:Type, Γ' ⊢ e : R(α)

Γ', β:Type, γ:Type, (Γ'[β → γ/α]), x:R(β), y:R(γ) ⊢ e_{δ}[β → γ/α] : c[β → γ, β → γ/α, δ]

Γ, β:Type, γ:Type, (Γ'[β γ/α]), x:R(β), y:R(γ) ⊢ e_{x}[β × γ/α] : c[β × γ, β × γ/α, δ]

Γ, β:Type, (Γ'[R(β)/α]), x:R(β) ⊢ e_R[R(β)/α] : c[R(β), R(β)/α, δ]

Γ, α:Type, Γ ⊢ e_{v} : c[α/δ]  Γ, α:Type, Γ' ⊢ e_{v} : c[α/δ] (tcase_{α})

Γ ⊢ e : R(c')

Γ ⊢ e_{int} : c[int/δ]

Γ, β:Type, γ:Type, x:R(β), y:R(γ) ⊢ e_{δ}[β → γ/δ]

Γ, β:Type, γ:Type, x:R(β), y:R(γ) ⊢ e_{x}[β × γ/δ]

Γ, β:Type, x:R(β) ⊢ e_R : c[R(β)/δ]

Γ ⊢ e_{v} : c[c'/δ]  Γ ⊢ e_{3} : c[c'/δ] (tcase)

Γ ⊢ typecase[δ, c] e (e_{int}, βγxy.e_x, βγxy.e_x, βx.e_R, e_{v}, e_{3}) : c[c'/δ]

Figure 9: Term Formation (typecase)