A New Approach to Generic Functional Programming

July 1999

RALF HINZE
Institut für Informatik III, Universität Bonn
Römerstraße 164, 53117 Bonn, Germany
(*e-mail:* `ralf@informatik.uni-bonn.de`)

Abstract

This paper describes a new approach to generic functional programming, which allows us to define functions generically for all datatypes expressible in Haskell. A generic function is one that is defined by induction on the structure of types. Typical examples include pretty printers, parsers, and comparison functions. The advanced type system of Haskell presents a real challenge: datatypes may be parameterized not only by types but also by type constructors, type definitions may involve mutual recursion, and recursive calls of type constructors can be arbitrarily nested. We show that—despite this complexity—a generic function is uniquely defined by giving cases for primitive types and type constructors (such as disjoint unions and cartesian products). Given this information a generic function can be specialized to arbitrary Haskell datatypes. The key idea of the approach is to model types by terms of the simply typed λ-calculus augmented by a family of recursion operators. While conceptually simple, our approach places high demands on the type system: it requires polymorphic recursion, rank-n types, and a strong form of type constructor polymorphism. Finally, we point out connections to Haskell's class system and show that our approach in a sense generalizes type classes.

1 Introduction

Programming is sometimes a challenge and sometimes a nuisance. Recurring routine tasks such as changing the representation of data, converting the internal representation into a human readable form (pretty printing) or vice versa (parsing) fall into the latter category. This is the domain of generic programming, which aims at relieving the programmer from repeatedly writing functions of similar functionality for different user-defined datatypes. A generic functional program is essentially a collection of type-indexed values—typically functions—that are defined by induction on the structure of types. A generic function such as a pretty printer or a parser is written once and for all times; its specialization to different instances of datatypes happens without further effort from the user. This way generic programming greatly simplifies the construction and maintenance of software systems as it automatically adapts functions to changes in the representation of data.

Different approaches to generic functional programming differ in the way the structure of datatypes is modelled. The language extension PolyP (Jansson & Jeuring, 1997), for instance, is based on the initial algebra semantics of datatypes. An unfortunate consequence of this choice is that the class of datatypes is restricted to so-called regular datatypes. A regular datatype is a parameterized type whose definition neither contains function spaces nor recursive calls which involve a change of the type parameter(s). Thus, compared to the class of datatypes definable in Haskell 98 (Peyton Jones & Hughes, 1999) PolyP covers only a small, though important subset. This is a shame. Especially, because the importance of generic programming increases with the complexity of types. We will get to know several involved datatypes, where even the skilled Haskell programmer will experience considerable hardship when trying to implement, say, comparison functions for these types.

Admittedly, Haskell has a very rich type system, which presents a real challenge for implementors of generic programming extensions: datatypes may be parameterized not only by types but also by type constructors, type definitions may involve mutual recursion, and recursive calls of type constructors can be arbitrarily nested. If we aim at extending generic programming to the full type system of Haskell, we are well-advised to model datatypes as closely and as faithfully as possible. Now, since Haskell supports both type application and (an implicit form of) type abstraction, type terms correspond essentially to terms of the simply typed λ-calculus augmented by a family of recursion operators and by additional constants denoting primitive types (such as integers) and primitive type constructors (such as disjoint unions and cartesian products). The 'types' of type terms are described using so-called kinds (Jones, 1995b).

Though Haskell's type system is quite involved, defining generic values is easy. We will see that a generic value is uniquely defined by giving cases for primitive types and type constructors. Given this information a generic value can be specialized to arbitrary Haskell datatypes. Interestingly, the process of specialization can be seen as an interpretation of simply typed λ-terms.

The main contributions of this paper are the following.

- We explain how to define values indexed by types or by type constructors of first-order kind.
- We show how to specialize generic values to concrete instances of datatypes. We identify several programming language features that are required to make this translation work.
- We point out connections to Haskell's type classes (Jones, 1995a) and show that our approach in a sense generalizes type classes.

The rest of the paper is organized as follows. Sec. 2 introduces Haskell's type system by means of some examples. Sec. 3 sketches our approach to generic functional programming. Sec. 4 introduces the simply typed λ-calculus and makes explicit how datatypes are modelled by λ-terms. Sec. 5 shows how to define type-indexed values and explains how to specialize a type-indexed value to concrete instances of datatypes. Sec. 6 generalizes the approach to type constructors of first-order kind. Sec. 7 works towards encoding instances of generic values in Haskell and relates type-indexed definitions to type classes. Finally, Sec. 8 reviews related work and Sec. 9 concludes.

2 Haskell's type system

This section introduces the main features of Haskell's type system by means of some examples. It is not important to understand the examples in full detail. The purpose of this section is rather to demonstrate the expressiveness of Haskell's type system and to give evidence that real use has been made of these features.

As first, rather simple examples consider the datatypes of polymorphic lists and rose trees (Bird, 1998).

$$\begin{aligned}
\textbf{data}\ \mathit{List}\ a &= \mathit{Nil} \mid \mathit{Cons}\ a\ (\mathit{List}\ a)\\
\textbf{data}\ \mathit{Rose}\ a &= \mathit{Branch}\ a\ (\mathit{List}\ (\mathit{Rose}\ a))
\end{aligned}$$

The two equations introduce two type constructors, List and Rose, of kind $\star \to \star$. The kind system of Haskell specifies the 'type' of a type constructor (Jones, 1995b). The '$\star$' kind represents nullary constructors like Int. The kind $\kappa_1 \to \kappa_2$ represents type constructors that map type constructors of kind κ_1 to those of kind κ_2. The kind system is necessary since Haskell supports abstraction over type constructors of arbitrary kind. The following generalization of rose trees illustrates this feature.

$$\textbf{data}\ \mathit{GRose}\ f\ a \;=\; \mathit{GBranch}\ a\ (f\ (\mathit{GRose}\ f\ a))$$

A slight variant of this definition has been used in (Okasaki, 1998) to extend an implementation of priority queues with an efficient merge operation. The type constructor GRose has kind $(\star \to \star) \to (\star \to \star)$ and is related to Rose by $\mathit{GRose}\ \mathit{List} = \mathit{Rose}$. Note that GRose has a second-order kind. The order of a kind is given by $\mathit{order}(\star) = 0$ and $\mathit{order}(\kappa_1 \to \kappa_2) = \mathit{max}\{1 + \mathit{order}(\kappa_1), \mathit{order}(\kappa_2)\}$.

The following definition introduces a fixpoint operator on the level of types. This definition appears, for instance, in (Meijer & Hutton, 1995) where it is employed to give a generic definition of so-called cata- and anamorphisms.

$$\begin{aligned}
\textbf{data}\ \mathit{Fix}\ f &= \mathit{In}\ (f\ (\mathit{Fix}\ f))\\
\textbf{data}\ \mathit{BaseList}\ a\ b &= \mathit{Nil} \mid \mathit{Cons}\ a\ b
\end{aligned}$$

The kinds of these types are given by $\mathit{Fix} :: (\star \to \star) \to \star$ and $\mathit{BaseList} :: \star \to (\star \to \star)$. Using Fix and $\mathit{BaseList}$ the datatype of polymorphic lists can alternatively be defined by $\mathit{List}\ a = \mathit{Fix}\ (\mathit{BaseList}\ a)$. Note that binary type constructors like $\mathit{BaseList}$ are, in fact, curried.

The type terms on the right-hand side of datatype definitions may be arbitrarily nested. The following equations, which implement binary random-access lists (Okasaki, 1998), exemplify nested type terms involving recursion.

$$\begin{aligned}
\textbf{data}\ \mathit{Fork}\ a &= \mathit{Node}\ a\ a\\
\textbf{data}\ \mathit{Sequ}\ a &= \mathit{Empty}\\
&\mid \mathit{Zero}\ (\mathit{Sequ}\ (\mathit{Fork}\ a))\\
&\mid \mathit{One}\ a\ (\mathit{Sequ}\ (\mathit{Fork}\ a))
\end{aligned}$$

Since the type parameter of *Sequ* is changed in the recursive calls, *Sequ* is termed a nested or non-regular datatype (Bird & Meertens, 1998). Nested datatypes are practically important since they can capture data-structural invariants in a way that regular datatypes cannot. For instance, *Sequ* captures the invariant that binary random-access lists are sequences of perfect binary leaf trees stored in increasing order of height. We refer the interested reader to (Okasaki, 1998; Bird & Paterson, 1999b; Bird & Paterson, 1999a) for further examples of nested types.

The following equations employ both abstraction over type constructors and nested recursion.

$$
\begin{array}{l}
\textbf{data}\ \mathit{MapFork}\ m\ v \\
\quad =\ \ \mathit{TrieFork}\ (m\ (m\ v)) \\
\textbf{data}\ \mathit{MapSequ}\ m\ v \\
\quad =\ \ \mathit{TrieSequ}\ v \\
\qquad\qquad\quad (\mathit{MapSequ}\ (\mathit{MapFork}\ m)\ v) \\
\qquad\qquad\quad (m\ (\mathit{MapSequ}\ (\mathit{MapFork}\ m)\ v))
\end{array}
$$

The types $\mathit{MapFork}, \mathit{MapSequ} :: (\star \to \star) \to (\star \to \star)$ represent the so-called generalized tries (Hinze, 1999a) for *Fork* and *Sequ*. Note that *MapFork* is the type-level counterpart of the function $\mathit{twice}\ f\ x = f\ (f\ x)$, which applies a given function twice to a given value. More examples of nested types of higher-order kinds can be found in (Hinze, 1998; Okasaki, 1999; Hinze, 1999c).

Now, before going on the reader is invited to consider programming, say, comparison functions for the datatypes above. This is a non-trivial task especially for the nested and for the second-order kinded types. In the sequel we will show that defining a *generic* comparison function that works for *all* datatypes is—perhaps surprisingly—a much simpler task.

3 Generic programming in a nutshell

This section sketches the main ideas of our approach to generic functional programming primarily from the programmer's perspective. A formal treatment of the approach is deferred to Sec. 5 and Sec. 6.

Before explaining how to define type-indexed values let us first take a closer look at datatype definitions. Haskell's **data** construct combines several features in a single coherent form: type abstraction, n-ary disjoint sums, n-ary cartesian products, and type recursion. The following rewritings of *List* and *GRose* make the structure of the datatype definitions more explicit.

$$
\begin{array}{lcl}
\mathit{List}\ a & = & 1 + a \times \mathit{List}\ a \\
\mathit{GRose}\ f\ a & = & a \times f\ (\mathit{GRose}\ f\ a)
\end{array}
$$

Here, 1 denotes the unit type, and '+' and '×' are more conventional notation for binary disjoint sums and binary cartesian products. For simplicity, we assume that

n-ary sums are reduced to binary sums and n-ary products to binary products. We treat 1, '+', and '×' as if they were given by the following datatype declarations.

$$
\begin{array}{lcl}
\textbf{data}\ 1 & = & () \\
\textbf{data}\ a_1 + a_2 & = & \mathit{Inl}\ a_1 \mid \mathit{Inr}\ a_2 \\
\textbf{data}\ a_1 \times a_2 & = & (a_1, a_2)
\end{array}
$$

Now, to define a type-indexed value it suffices to specify cases for the primitive types (say, 1 and *Int*) and for the primitive type constructors (say, '+' and '×'). As an example, the following equations define the generic function $enc\langle a\rangle$, which encodes elements of type a as bit strings implementing a simple form of data compression. The type argument of *enc* is written in angle brackets to distinguish it from the value argument.

$$
\begin{array}{lcl}
\textbf{type}\ \mathit{Bin} & = & [\mathit{Bit}] \\
\textbf{data}\ \mathit{Bit} & = & \mathtt{0} \mid \mathtt{1} \\
enc\langle a :: \star\rangle & :: & a \to \mathit{Bin} \\
enc\langle 1\rangle\ () & = & [\,] \\
enc\langle \mathit{Int}\rangle\ n & = & \mathit{encInt}\ n \\
enc\langle a_1 + a_2\rangle\ (\mathit{Inl}\ x_1) & = & \mathtt{0} : enc\langle a_1\rangle\ x_1 \\
enc\langle a_1 + a_2\rangle\ (\mathit{Inr}\ x_2) & = & \mathtt{1} : enc\langle a_2\rangle\ x_2 \\
enc\langle a_1 \times a_2\rangle\ (x_1, x_2) & = & enc\langle a_1\rangle\ x_1 \mathbin{+\!\!+} enc\langle a_2\rangle\ x_2
\end{array}
$$

The type signature of *enc* makes explicit that the type of $enc\langle a\rangle$ depends on the type parameter a. The definition consists of one equation for 1, *Int*, and '×' and of two equations for '+'. Each equation is more or less inevitable. To encode the single element of 1 no bits are required. Integers are encoded using the primitive function *encInt*, whose existence we assume. To encode an element of a disjoint union we emit one bit for the constructor followed by the encoding of its argument. Finally, the encoding of a pair is given by the concatenation of the component's encodings.[1]

This simple definition contains all ingredients needed to derive specializations for compressing elements of arbitrary datatypes. For instance, $enc\langle \mathit{Sequ}\ \mathit{Int}\rangle$ of type $\mathit{Sequ}\ \mathit{Int} \to \mathit{Bin}$ compresses random-access lists with integer elements and $enc\langle \mathit{GRose}\ \mathit{List}\ \mathit{Int}\rangle$ compresses generalized rose trees with integer labels.

Remark 1
Note that generic values cannot directly be implemented in Haskell or Standard ML. The reason is simply that $enc\langle t\rangle$ is a *value* which depends on the *type* t. Value-type dependencies cannot be expressed in current functional languages.[2] Even if we circumvented this problem by using encodings into a universal datatype *(Yang, 1999)*

[1] The definition of *enc* exhibits $\Theta(n^2)$ worst-case behaviour, but this is easy to remedy.

[2] Note that dependent types as implemented in Cayenne (Augustsson, 1999) allow to define types which depend on values. Here, we require values which depend on types. One could argue that value-type dependencies are present in the second-order λ-calculus (Girard, 1972) since a polymorphic function depends on the type argument supplied. The dependence is, however, quite loose since a polymorphic function uses 'the same algorithm' at each type. Note that a function of type $\forall a.a \to \mathit{Bit}$ is necessarily a constant function—this is a simple consequence of the parametricity theorem (Wadler, 1989).

or by using dynamic types and a **typecase** *(Abadi* et al., *1995)*, the result would be rather inefficient because *enc* would repeatedly interpret its type argument. By specializing $enc\langle t\rangle$ for a given t we remove this interpretative layer.

In order to specialize $enc\langle t\rangle$ we cannot simply unfold or partially evaluate the definition of *enc*. To see why consider specializing $enc\langle Sequ\ Int\rangle$: to define $enc\langle Sequ\ Int\rangle$ we require $enc\langle Sequ\ (Fork^n\ Int)\rangle$ for each $n \geqslant 1$. If we simply unfold the definition, we will in general not obtain a finite representation of $enc\langle t\rangle$.

The key idea of the specialization is to mimic the structure of types at the value level. For instance, $enc\langle Sequ\ Int\rangle$ should be compositionally defined in terms of specializations for the constituent types, say, *encSequ* and *encInt*. Since *Sequ* is a function on types, *encSequ* is consequently a function on encoders. Then the encoder for the type application *Sequ Int* is given by the application of *encSequ* to *encInt*. In a nutshell, type abstraction is mapped to value abstraction, type application to value application, and type recursion to value recursion. To exemplify, for *GRose*, *Fork*, and *Sequ* we can automatically derive the following specializations (for clarity, we use the original constructor names).

$$
\begin{aligned}
&encGRose\ encf\ enca\ (GBranch\ x\ ts)\\
&\quad = \quad enca\ x \mathbin{+\!\!+} encf\ (encGRose\ encf\ enca)\ ts\\
&encFork\ enca\ (Node\ x_1\ x_2)\\
&\quad = \quad enca\ x_1 \mathbin{+\!\!+} enca\ x_2\\
&encSequ\ enca\ Empty\\
&\quad = \quad [\mathtt{0}]\\
&encSequ\ enca\ (Zero\ s)\\
&\quad = \quad \mathtt{1} : \mathtt{0} : encSequ\ (encFork\ enca)\ s\\
&encSequ\ enca\ (One\ x\ s)\\
&\quad = \quad \mathtt{1} : \mathtt{1} : enca\ x \mathbin{+\!\!+} encSequ\ (encFork\ enca)\ s
\end{aligned}
$$

Note that we have intentionally omitted the types of the functions. They are motivated and explained in Sec. 5.3.

The function *enc* is indexed by a type of kind $\star$. The question naturally arises how to define values that are indexed by type constructors of arbitrary kind. This paper deals with type arguments of first-order kind only. The generalization to higher-order kinds is subject of future research. In the remainder of this section we confine ourselves furthermore to type indices of kind $\star \to \star$. The archetypical example of a generic value indexed by a type constructor of this kind is $size\langle f\rangle :: \forall x.f\ x \to Int$, which counts the number of elements of type x in given value of type $f\ x$. In order to define $size\langle f\rangle$ generically for all f we must first explicate the structure of type constructors of kind $\star \to \star$. To this end we lift 1, *Int*, '+', and '×' into a higher realm.

$$
\begin{aligned}
\underline{1}\ a &= 1\\
\underline{Int}\ a &= Int\\
(f_1 \mathbin{\underline{+}} f_2)\ a &= f_1\ a + f_2\ a\\
(f_1 \mathbin{\underline{\times}} f_2)\ a &= f_1\ a \times f_2\ a
\end{aligned}
$$

Note that '+' has kind $\star \to (\star \to \star)$ whereas the lifted version has kind $(\star \to \star) \to ((\star \to \star) \to (\star \to \star))$. Now, every type constructor of kind $\star \to \star$ can be expressed in terms of $\underline{1}$, $\underline{\mathit{Int}}$, '$\underline{+}$', '$\underline{\times}$', and the identity type given by $\mathit{Id}\ a = a$—this claim will be justified in Sec. 6.1. Thus, $\mathit{size}\langle f\rangle$ is uniquely defined by the following equations.

$$\begin{array}{lcl}
\mathit{size}\langle f :: \star \to \star\rangle & :: & \forall x . f\ x \to \mathit{Int} \\
\mathit{size}\langle \mathit{Id}\rangle\ x & = & 1 \\
\mathit{size}\langle \underline{1}\rangle\ () & = & 0 \\
\mathit{size}\langle \underline{\mathit{Int}}\rangle\ n & = & 0 \\
\mathit{size}\langle f_1 \mathbin{\underline{+}} f_2\rangle\ (\mathit{Inl}\ x_1) & = & \mathit{size}\langle f_1\rangle\ x_1 \\
\mathit{size}\langle f_1 \mathbin{\underline{+}} f_2\rangle\ (\mathit{Inr}\ x_2) & = & \mathit{size}\langle f_2\rangle\ x_2 \\
\mathit{size}\langle f_1 \mathbin{\underline{\times}} f_2\rangle\ (x_1, x_2) & = & \mathit{size}\langle f_1\rangle\ x_1 + \mathit{size}\langle f_2\rangle\ x_2
\end{array}$$

Again, each equation is more or less inevitable. A value of type $\mathit{Id}\ x = x$ contains exactly one element of type x; values of type $\underline{1}\ x = 1$ or $\underline{\mathit{Int}}\ x = \mathit{Int}$ contain no elements. To determine the size of an element of type $(f_1 \mathbin{\underline{+}} f_2)\ x$ we must either calculate the size of a structure of type $f_1\ x$ or that of a structure of type $f_2\ x$ depending on which component of the disjoint sum the argument comes from. Finally, the size of a structure of type $(f_1 \mathbin{\underline{\times}} f_2)\ x$ is given by the sum of the size of the two components.

Specializing *size* to concrete instances of datatypes works essentially as before. For instance, for *List* and *GRose* we obtain the following specializations.

$$\begin{array}{l}
\mathit{sizeList}\ \mathit{sizea}\ \mathit{Nil} = 0 \\
\mathit{sizeList}\ \mathit{sizea}\ (\mathit{Cons}\ x\ \mathit{xs}) \\
\quad = \quad \mathit{sizea}\ x + \mathit{sizeList}\ \mathit{sizea}\ \mathit{xs} \\
\mathit{sizeGRose}\ \mathit{sizef}\ \mathit{sizea}\ (\mathit{GBranch}\ x\ \mathit{ts}) \\
\quad = \quad \mathit{sizea}\ x + \mathit{sizef}\ (\mathit{sizeGRose}\ \mathit{sizef}\ \mathit{sizea})\ \mathit{ts}
\end{array}$$

Again, the definitions rigidly follow the structure of types: since *List* is a function on types, *sizeList* transforms size functions into size functions. More precisely, if *sizea* determines the size of a structure of type $a\ x$, then *sizeList sizea* determines the size of a structure of type $\mathit{List}\ (a\ x)$—the full typings will be given in Sec. 6.3. Now, to obtain the length function for lists we simply pass the size function for *Id*, ie $\mathit{sizeId}\ x = 1$, to *sizeList*. Likewise, to determine the size of a structure of type, say, $\mathit{GRose}\ \mathit{List}\ x$ we call *sizeGRose sizeList sizeId*.

4 Kinds and types

This section introduces kind and type terms. Kind terms are formed according to the following grammar.

$$K \quad ::= \quad \star \mid (K \to K)$$

We agree upon that '$\to$' associates to the right. Given a fixed set of type variables X and a fixed set of primitive type constructors $C = \{1, \mathit{Int}, (+), (\times)\}$ type terms are formed according to the following grammar.

$$T \quad ::= \quad X \mid C \mid (T\ T) \mid (\Lambda X :: K \to T) \mid (\mu T)$$

$$\frac{(x :: \kappa) \in \Gamma}{\Gamma \vdash x :: \kappa} \qquad \frac{\Gamma \vdash t_1 :: \kappa_2 \to \kappa_1 \qquad \Gamma \vdash t_2 :: \kappa_2}{\Gamma \vdash (t_1\ t_2) :: \kappa_1}$$

$$\frac{\Gamma, x :: \kappa_1 \vdash t :: \kappa_2}{\Gamma \vdash (\Lambda x :: \kappa_1 \to t) :: (\kappa_1 \to \kappa_2)} \qquad \frac{\Gamma \vdash t :: \kappa \to \kappa}{\Gamma \vdash (\mu t) :: \kappa}$$

Fig. 1. Kind rules.

Here, $t_1\ t_2$ denotes type application, $\Lambda x :: \kappa \to t$ denotes type abstraction, and μt denotes the fixpoint of t. We agree upon that type application associates to the left and that type abstraction extends as far to the right as possible. We abbreviate $\Lambda x_1 :: \kappa_1 \to \cdots \Lambda x_m :: \kappa_m \to t$ by $\Lambda x_1 :: \kappa_1\ \ldots\ x_m :: \kappa_m \to t$ and write $(+)\ t_1\ t_2$ as $t_1 + t_2$ and similarly for '$\times$'. Note that in a Λ-abstraction the bound type variable is annotated with its kind. For reasons of readability we will usually omit the kind annotation. Furthermore, note that the choice of C is more or less arbitrary. For instance, C could additionally include the function space constructor '$\to$'. We have only omitted '$\to$' because most of the generic functions cannot sensibly be defined for the function space.

The kinds of the primitives are given by

$$\begin{aligned} 1, \mathit{Int} &:: \star \\ (+), (\times) &:: \star \to \star \to \star\ . \end{aligned}$$

The kinds of type terms are determined by the rules depicted in Fig. 1. The notation $\Gamma \vdash t :: \kappa$ means that the statement $t :: \kappa$ is derivable from the set of kind assumptions Γ. If $t :: \kappa_1 \to \cdots \to \kappa_n \to \star$ is derivable, we say that t has arity n. Note that 'μ' is polymorphic with respect to the kinds, ie it represents a family of fixpoint operators.

In essence, Haskell types are represented by terms of the simply typed λ-calculus with kinds playing the rôle of types. The translation of the datatype declarations given in Sec. 2 into type terms is fairly straightforward. Here are some examples.

$$\begin{aligned} \mathit{List} &= \Lambda a \to \mu(\Lambda \ell \to 1 + a \times \ell) \\ \mathit{Fork} &= \Lambda a \to a \times a \\ \mathit{Sequ} &= \mu(\Lambda s\ a \to 1 + s\ (\mathit{Fork}\ a) + a \times s\ (\mathit{Fork}\ a)) \end{aligned}$$

First-order kinded, regular types, such as *List* and *Rose*, can be modelled using a fixpoint operator of kind $\star \to \star$ while nested types, such as *Sequ*, require an operator of kind $(\star \to \star) \to (\star \to \star)$.

Remark 2

Type recursion is expressed using a fixpoint operator. Whilst this is sufficient for the datatype definitions introduced in Sec. 2, it does not generalize easily to handle mutually recursive types. A viable alternative is to consider systems of recursion equations:

$$x_1 = t_1; \ldots; x_n = t_n\ ,$$

where the x_i are type variables and the t_i are type terms. The approach to generic programming works equally well if we use recursion equations instead of a fixpoint operator. The development, however, becomes more verbose and provides no additional insights.

In the following sections we require the notion of Böhm tree which we introduce next. Böhm trees can be considered as a kind of 'infinite normal form' for type terms and are obtained by unwinding type terms ad infinitum.

Definition 1
The *convertibility relation* on type terms, denoted '$\leftrightarrow$', is given by the following axioms

$$\begin{array}{rcll} (\Lambda x \to t)\ u & \leftrightarrow & t\,[x := u] & & (\beta) \\ \Lambda x \to t\ x & \leftrightarrow & t\,, & x \text{ not free in } t & (\eta) \\ \mu t & \leftrightarrow & t\ (\mu t) & & (\mu) \end{array}$$

plus the usual 'logical' rules for reflexivity, symmetry, transitivity, and congruence.

Definition 2
A *head normal form* (hnf) is a type term of the form $\Lambda x_1\ \ldots\ x_m \to z\ u_1\ \ldots\ u_n$ with $m, n \geqslant 0$, $z \in C \cup X$, and z has arity n. A type term t has hnf u if $t \leftrightarrow u$ and u is an hnf.

The definition of hnf is a bit unusual in that we additionally require hnfs to be η-expanded (in order to guarantee that Böhm trees are well-defined). The term $(+)$, for instance, has hnf $\Lambda x_1\ x_2 \to x_1 + x_2$. Not every type term has an hnf, consider, for instance $\mu(\Lambda a \to a)$. Type terms which have no hnf are in a sense pointless and can be excluded by a simple syntactic restriction, which we will adopt in the sequel: t in μt must have the form $\Lambda x_1\ \ldots\ x_m \to c\ u_1\ \ldots\ u_n$ with $c \in C$ of arity n.

Definition 3
The *Böhm tree* of the type term t, denoted $\mathrm{BT}(t)$, is a labelled tree defined as follows: if t has hnf $\Lambda x_1\ \ldots\ x_m \to z\ u_1\ \ldots\ u_n$, then

$$\mathrm{BT}(t) \;=\; \Lambda x_1\ \ldots\ x_m \to \begin{array}{c} z \\ \diagup \qquad \diagdown \\ \mathrm{BT}(u_1) \quad \cdots \quad \mathrm{BT}(u_n) \end{array} .$$

For a more formal treatment of Böhm trees we refer the reader to (Barendregt, 1984). Now, from a generic programming point of view we can identify type terms that have the same Böhm trees, ie $t = u$ if $\mathrm{BT}(t) = \mathrm{BT}(u)$. We have, for instance, $\mathit{Rose} = \mathit{GRose}\ (\Lambda a \to \mathit{Fix}\ (\mathit{BaseList}\ a))$. Since generic values are defined by induction on the *structure* of types, the structure is all that matters. By contrast, type inference in Haskell is essentially based on name rather than on structural equivalence.

5 Type-indexed values

The theme is developed in three steps: (1) we characterize the set of normal forms of types of kind $\star$, (2) we give a prototype for generic values indexed by types of this kind, and (3) we show how to promote a generic value thus defined to types of arbitrary kind.

5.1 Normal forms of types of kind $\star$

Types of kind $\star$ have a very simple normal form. Consider the Böhm tree of a type of kind $\star$. Clearly, the root of the tree cannot be labelled with a type abstraction. Instead, it must be labelled with a primitive type constructor, say, c. Moreover, if c has arity n, the root must have n direct successors. Thus, the normal form of type terms of kind $\star$ is described by the following grammar.

$$T_\star \quad ::= \quad 1 \mid \mathit{Int} \mid (T_\star + T_\star) \mid (T_\star \times T_\star)$$

It is understood that $T_\star$ contains *type terms* of kind $\star$. The set of all *type trees* (finite and infinite) that can be formed according to this grammar is denoted $T_\star^\infty$. We have $T_\star^\infty = \{\mathrm{BT}(t) \mid \emptyset \vdash t :: \star\}$.

5.2 Defining $\star$-indexed values

The characterization of normal forms motivates the following prototype for type-indexed values.

$$\begin{array}{lcl}
\mathit{poly}\langle a :: \star\rangle & :: & F\ a \\
\mathit{poly}\langle 1\rangle & = & \mathit{poly}_1 \\
\mathit{poly}\langle \mathit{Int}\rangle & = & \mathit{poly}_{\mathit{Int}} \\
\mathit{poly}\langle a_1 + a_2\rangle & = & \mathit{poly}_+\ (\mathit{poly}\langle a_1\rangle)\ (\mathit{poly}\langle a_2\rangle) \\
\mathit{poly}\langle a_1 \times a_2\rangle & = & \mathit{poly}_\times\ (\mathit{poly}\langle a_1\rangle)\ (\mathit{poly}\langle a_2\rangle)
\end{array}$$

Here, *poly* is the name of the type-indexed value; a, a_1, and a_2 are type variables of kind $\star$; F, poly_1, $\mathit{poly}_{\mathit{Int}}$, poly_+, and $\mathit{poly}_\times$ are the ingredients that have to be supplied by the generic programmer. The type of $\mathit{poly}\langle a\rangle$ is given by $F\ a$, where F is a type constructor of kind $\star \to \star$. Unlike the type index F may also contain function types, $t_1 \to t_2$, and universally quantified types, $\forall x.t$. The definition of *poly* is inductive on the structure of $T_\star$: we have one equation for each primitive type constructor. Now, a standard result from the theory of infinite trees (Courcelle, 1983) guarantees that *poly* possesses a unique (least) extension in $T_\star^\infty$. In that sense, *poly* is uniquely defined by its action on primitive type constructors, ie by poly_1, $\mathit{poly}_{\mathit{Int}}$, poly_+, and $\mathit{poly}_\times$.

The poly_c values must have the following types:

$$\begin{array}{lcl}
\mathit{poly}_1 & :: & F\ 1 \\
\mathit{poly}_{\mathit{Int}} & :: & F\ \mathit{Int} \\
\mathit{poly}_+ & :: & \forall x_1, x_2.F\ x_1 \to F\ x_2 \to F\ (x_1 + x_2) \\
\mathit{poly}_\times & :: & \forall x_1, x_2.F\ x_1 \to F\ x_2 \to F\ (x_1 \times x_2)
\end{array}$$

In the latter two cases x_1 and x_2 must be universally qualified since poly_+ and $\mathit{poly}_\times$ must work for all possible argument types.

Before we proceed let us take a look at further examples of type-indexed values. The first example, *dec*, is essentially the inverse of *enc*: it takes a bit string, decodes a prefix of that string, and returns the decoded value coupled with the unused suffix.

$$
\begin{array}{lcl}
dec\langle a :: \star\rangle & :: & Bin \to (a, Bin) \\
dec\langle 1\rangle\ bs & = & ((), bs) \\
dec\langle Int\rangle\ bs & = & decInt\ bs \\
dec\langle a_1 + a_2\rangle\ [\,] & = & error\ \texttt{"dec"} \\
dec\langle a_1 + a_2\rangle\ (0 : bs) & = & app_1\ Inl\ (dec\langle a_1\rangle\ bs) \\
dec\langle a_1 + a_2\rangle\ (1 : bs) & = & app_1\ Inr\ (dec\langle a_2\rangle\ bs) \\
dec\langle a_1 \times a_2\rangle\ bs & = & \mathbf{let}\ (x_1, bs_1) = dec\langle a_1\rangle\ bs \\
 & & \phantom{\mathbf{let}}\ (x_2, bs_2) = dec\langle a_2\rangle\ bs_1 \\
 & & \mathbf{in}\ ((x_1, x_2), bs_2) \\
app_1\ \varphi\ (x_1, x_2) & = & (\varphi\ x_1, x_2)
\end{array}
$$

The functions *enc* and *dec* can be seen as very simple printers and parsers. *Pretty* printers and parsers, which produce and process a human readable format, can be implemented if we give the generic programmer additionally access to the constructor names. In (Hinze, 1999b) we use such an extension to define Haskell's *show* and *read* functions in a generic way.

Comparison functions are typical examples of type-indexed values. The following program realizes Haskell's *compare* function, which determines the precise ordering of two elements.

$$
\begin{array}{lcl}
\mathbf{data}\ Ordering & = & LT \mid EQ \mid GT \\
cmp\langle a :: \star\rangle & :: & a \to a \to Ordering \\
cmp\langle 1\rangle\ ()\ () & = & EQ
\end{array}
$$

$$
\begin{array}{lcl}
cmp\langle Int\rangle\ m\ n & = & cmpInt\ m\ n \\
cmp\langle a_1 + a_2\rangle\ (Inl\ x_1)\ (Inl\ y_1) & = & cmp\langle a_1\rangle\ x_1\ y_1 \\
cmp\langle a_1 + a_2\rangle\ (Inl\ x_1)\ (Inr\ y_2) & = & LT \\
cmp\langle a_1 + a_2\rangle\ (Inr\ x_2)\ (Inl\ y_1) & = & GT \\
cmp\langle a_1 + a_2\rangle\ (Inr\ x_2)\ (Inr\ y_2) & = & cmp\langle a_2\rangle\ x_2\ y_2 \\
\multicolumn{3}{l}{cmp\langle a_1 \times a_2\rangle\ (x_1, x_2)\ (y_1, y_2)} \\
\multicolumn{3}{l}{\quad = \ cmp\langle a_1\rangle\ x_1\ y_1\ \text{`}lexord\text{`}\ cmp\langle a_2\rangle\ x_2\ y_2}
\end{array}
$$

The helper function *lexord* used in the last equation implements the lexicographic product of two orderings.

$$
\begin{array}{lcl}
lexord & :: & Ordering \to Ordering \to Ordering \\
lexord\ LT\ ord & = & LT \\
lexord\ EQ\ ord & = & ord \\
lexord\ GT\ ord & = & GT
\end{array}
$$

5.3 Specializing $\star$-indexed values

The purpose of a generic value is to be specialized. We have already discussed the key idea of promoting a generic value to types of arbitrary kinds: type abstraction is interpreted by value abstraction, type application by value application, and type recursion by value recursion. The promoted version of $poly\langle\cdot\rangle$, which we denote $poly\langle\!\langle\cdot\rangle\!\rangle$, is consequently given by the following equations.

$$
\begin{array}{lcl}
poly\langle\!\langle t :: \kappa\rangle\!\rangle\ \varrho & :: & F\langle\!\langle t :: \kappa\rangle\!\rangle \\
poly\langle\!\langle c\rangle\!\rangle\ \varrho & = & poly_c \\
poly\langle\!\langle x\rangle\!\rangle\ \varrho & = & \varrho\ x \\
poly\langle\!\langle t_1\ t_2\rangle\!\rangle\ \varrho & = & (poly\langle\!\langle t_1\rangle\!\rangle\ \varrho)\ (poly\langle\!\langle t_2\rangle\!\rangle\ \varrho) \\
poly\langle\!\langle \Lambda x \to t\rangle\!\rangle\ \varrho & = & \lambda v \to poly\langle\!\langle t\rangle\!\rangle\ (\varrho(x := v)) \\
poly\langle\!\langle \mu t\rangle\!\rangle\ \varrho & = & fix\ (poly\langle\!\langle t\rangle\!\rangle\ \varrho)
\end{array}
$$

Here, ϱ ranges over environments, which map type variables to value variables, and $\varrho(x := v)$ is syntax for extending the environment ϱ by the binding $x := v$. The function *fix* is the polymorphic fixpoint operator on the value level. Note that $poly\langle\!\langle t\rangle\!\rangle$ is inductive on the structure of type terms. We can, in fact, view its definition as an interpretation of the simply typed λ-calculus. Furthermore, note that $poly\langle\cdot\rangle$ is related to $poly\langle\!\langle\cdot\rangle\!\rangle$ by

$$poly\langle t :: \star\rangle \quad = \quad poly\langle\!\langle t :: \star\rangle\!\rangle\ \epsilon\ ,$$

where ϵ denotes the empty environment. Thus, in order to specialize $poly\langle t\rangle$ we simply specialize $poly\langle\!\langle t\rangle\!\rangle$.

It remains to specify $F\langle\!\langle t :: \kappa\rangle\!\rangle$, which is defined by induction on the structure of kinds.

$$
\begin{array}{lcl}
F\langle\!\langle u :: \star\rangle\!\rangle & = & F\ u \\
F\langle\!\langle u :: \kappa_1 \to \kappa_2\rangle\!\rangle & = & \forall x.F\langle\!\langle x :: \kappa_1\rangle\!\rangle \to F\langle\!\langle u\ x :: \kappa_2\rangle\!\rangle
\end{array}
$$

If u is a type constructor of kind $\kappa_1 \to \kappa_2$, then $poly\langle\!\langle u\rangle\!\rangle$ is a function which maps values of type $F\langle\!\langle x :: \kappa_1\rangle\!\rangle$ to values of type $F\langle\!\langle u\ x :: \kappa_2\rangle\!\rangle$, for all x. Again, it is important that x is universally quantified since u may be applied to different types. The nesting of universal quantifiers is dictated by the kind: if κ has order n, then $F\langle\!\langle u :: \kappa\rangle\!\rangle$ is a rank-n type (McCracken, 1984)—assuming that $F\ a$ has rank 0. For instance, for F given by $F\ a = a \to Bin$ we have

$$
\begin{array}{cl}
 & F\langle\!\langle GRose :: (\star \to \star) \to (\star \to \star)\rangle\!\rangle \\
= & \forall f.F\langle\!\langle f :: \star \to \star\rangle\!\rangle \to F\langle\!\langle GRose\ f :: \star \to \star\rangle\!\rangle \\
= & \forall f.(\forall b.F\langle\!\langle b :: \star\rangle\!\rangle \to F\langle\!\langle f\ b :: \star\rangle\!\rangle) \\
 & \qquad \to (\forall a.F\langle\!\langle a :: \star\rangle\!\rangle \to F\langle\!\langle GRose\ f\ a :: \star\rangle\!\rangle) \\
= & \forall f.(\forall b.F\ b \to F\ (f\ b)) \to (\forall a.F\ a \to F\ (GRose\ f\ a)) \\
= & \forall f.(\forall b.(b \to Bin) \to (f\ b \to Bin)) \\
 & \qquad \to (\forall a.(a \to Bin) \to (GRose\ f\ a \to Bin))\ .
\end{array}
$$

Since *GRose* has an order-2 kind, $F\langle\!\langle GRose\rangle\!\rangle$ is a rank-2 type.

Let us consider some examples. Fig. 2 lists the specializations of *enc* to the datatypes introduced in Sec. 2. For clarity, the definitions use the original constructor names and functions are written in an equational style. The specializations illustrate several interesting points. For instance, the function *encSequ* makes use

$$
\begin{array}{lcl}
\mathit{encList} & :: & \forall a.(a \to \mathit{Bin}) \to (\mathit{List}\ a \to \mathit{Bin}) \\
\mathit{encList}\ \mathit{enca} & = & \mathit{encl} \\
\quad \mathbf{where}\ \mathit{encl}\ \mathit{Nil} & = & [\mathtt{0}] \\
\qquad\qquad \mathit{encl}\ (\mathit{Cons}\ x\ \mathit{xs}) & = & \mathtt{1} : \mathit{enca}\ x +\!\!+\ \mathit{encl}\ \mathit{xs} \\
\mathit{encRose} & :: & \forall a.(a \to \mathit{Bin}) \to (\mathit{Rose}\ a \to \mathit{Bin}) \\
\mathit{encRose}\ \mathit{enca} & = & \mathit{encr} \\
\quad \mathbf{where}\ \mathit{encr}\ (\mathit{Branch}\ x\ \mathit{ts}) & = & \mathit{enca}\ x +\!\!+\ \mathit{encList}\ \mathit{encr}\ \mathit{ts} \\
\mathit{encGRose} & :: & \forall f.(\forall b.(b \to \mathit{Bin}) \to (f\ b \to \mathit{Bin})) \\
 & & \qquad \to (\forall a.(a \to \mathit{Bin}) \to (\mathit{GRose}\ f\ a \to \mathit{Bin})) \\
\mathit{encGRose}\ \mathit{encf}\ \mathit{enca} & = & \mathit{encg} \\
\quad \mathbf{where}\ \mathit{encg}\ (\mathit{GBranch}\ x\ \mathit{ts}) & = & \mathit{enca}\ x +\!\!+\ \mathit{encf}\ \mathit{encg}\ \mathit{ts} \\
\mathit{encFix} & :: & \forall f.(\forall a.(a \to \mathit{Bin}) \to (f\ a \to \mathit{Bin})) \\
 & & \qquad \to (\mathit{Fix}\ f \to \mathit{Bin}) \\
\mathit{encFix}\ \mathit{encf} & = & \mathit{encr} \\
\quad \mathbf{where}\ \mathit{encr}\ (\mathit{In}\ x) & = & \mathit{encf}\ \mathit{encr}\ x \\
\mathit{encBaseList} & :: & \forall a.(a \to \mathit{Bin}) \to (\forall b.(b \to \mathit{Bin}) \\
 & & \qquad\qquad\qquad\qquad \to (\mathit{BaseList}\ a\ b \to \mathit{Bin})) \\
\mathit{encBaseList}\ \mathit{enca}\ \mathit{encb}\ \mathit{Nil} & = & [\mathtt{0}] \\
\mathit{encBaseList}\ \mathit{enca}\ \mathit{encb}\ (\mathit{Cons}\ x\ y) & & \\
 & = & \mathtt{1} : \mathit{enca}\ x +\!\!+\ \mathit{encb}\ y \\
\mathit{encFork} & :: & \forall a.(a \to \mathit{Bin}) \to (\mathit{Fork}\ a \to \mathit{Bin}) \\
\mathit{encFork}\ \mathit{enca}\ (\mathit{Node}\ x_1\ x_2) & = & \mathit{enca}\ x_1 +\!\!+\ \mathit{enca}\ x_2 \\
\mathit{encSequ} & :: & \forall a.(a \to \mathit{Bin}) \to (\mathit{Sequ}\ a \to \mathit{Bin}) \\
\mathit{encSequ}\ \mathit{enca}\ \mathit{Empty} & = & [\mathtt{0}] \\
\mathit{encSequ}\ \mathit{enca}\ (\mathit{Zero}\ s) & = & \mathtt{1} : \mathtt{0} : \mathit{encSequ}\ (\mathit{encFork}\ \mathit{enca})\ s \\
\mathit{encSequ}\ \mathit{enca}\ (\mathit{One}\ x\ s) & = & \mathtt{1} : \mathtt{1} : \mathit{enca}\ x +\!\!+\ \mathit{encSequ}\ (\mathit{encFork}\ \mathit{enca})\ s \\
\mathit{encMapFork} & :: & \forall m.(\forall w.(w \to \mathit{Bin}) \to (m\ w \to \mathit{Bin})) \\
 & & \qquad \to (\forall v.(v \to \mathit{Bin}) \to (\mathit{MapFork}\ m\ v \to \mathit{Bin})) \\
\mathit{encMapFork}\ \mathit{encm}\ \mathit{encv}\ (\mathit{TrieFork}\ \mathit{tf}) & & \\
 & = & \mathit{encm}\ (\mathit{encm}\ \mathit{encv})\ \mathit{tf} \\
\mathit{encMapSequ} & :: & \forall m.(\forall w.(w \to \mathit{Bin}) \to (m\ w \to \mathit{Bin})) \\
 & & \qquad \to (\forall v.(v \to \mathit{Bin}) \to (\mathit{MapSequ}\ m\ v \to \mathit{Bin})) \\
\mathit{encMapSequ}\ \mathit{encm}\ \mathit{encv}\ (\mathit{TrieSequ}\ \mathit{te}\ \mathit{tz}\ \mathit{to}) & & \\
 & = & \mathit{encv}\ \mathit{te} \\
 & +\!\!+ & \mathit{encMapSequ}\ (\mathit{encMapFork}\ \mathit{encm})\ \mathit{encv}\ \mathit{tz} \\
 & +\!\!+ & \mathit{encm}\ (\mathit{encMapSequ}\ (\mathit{encMapFork}\ \mathit{encm})\ \mathit{encv})\ \mathit{to}
\end{array}
$$

Fig. 2. Specializing *enc* to the types of Sec. 2.

of polymorphic recursion (Mycroft, 1984): the recursive call has type $\forall a.(\mathit{Fork}\ a \to \mathit{Bin}) \to (\mathit{Sequ}\ (\mathit{Fork}\ a) \to \mathit{Bin})$, which is a substitution instance of the declared type. In general, polymorphic recursion is required whenever the type recursion is nested. Several functions have rank-2 type signatures; *encMapFork* shows in a nutshell why this is necessary: the argument *encm* is applied at two different instances: the inner call has type $\forall a.(a \to \mathit{Bin}) \to (m\ a \to \mathit{Bin})$ while the outer call has type $\forall a.(m\ a \to \mathit{Bin}) \to (m\ (m\ a) \to \mathit{Bin})$. The function *encMapSequ* even combines polymorphic recursion and the specialized use of a polymorphic argument.

6 Values indexed by first-order kinded types

In the previous section we have considered values indexed by types of kind $\star$. For values that are indexed by first-order kinded types, such as *size*, some additional machinery is needed. We will develop the main ideas for type indices of kind $\star \to \star$. The straightforward extension to first-order kinds is discussed in Sec. 6.4. We proceed as in the previous section: (1) we characterize the set of normal forms of types of kind $\star \to \star$, (2) we give a prototype for generic values indexed by types of this kind, and (3) we show how to promote a generic value to types of arbitrary kind.

6.1 Normal forms of types of kind $\star \to \star$

Essential for the development is the notion of lifting which we introduce now. Lifting maps a type $t :: \kappa$ to a type $\uparrow t :: \uparrow\kappa$ where $\uparrow\kappa$ is defined as follows.

$$\begin{aligned} \uparrow\star &= \star \to \star \\ \uparrow\kappa_1 \to \kappa_2 &= (\uparrow\kappa_1) \to (\uparrow\kappa_2) \end{aligned}$$

The lifted version $\uparrow t$ of type t is given by (we assume that for each type variable x of kind κ there is a lifted variable named $\underline{x}$ of kind $\uparrow\kappa$)

$$\begin{aligned} \uparrow c &= \underline{c} \\ \uparrow x &= \underline{x} \\ \uparrow t_1\ t_2 &= (\uparrow t_1)\ (\uparrow t_2) \\ \uparrow \Lambda x \to t &= \Lambda \underline{x} \to \uparrow t \\ \uparrow \mu t &= \mu(\uparrow t)\ . \end{aligned}$$

The lifted versions, $\underline{c}$, of the primitive type constructors have already been defined in Sec. 3. The lifted version of *List*, for instance, reads

$$\underline{List} \quad = \quad \Lambda\underline{a} \to \mu(\Lambda\underline{\ell} \to \underline{1} \mathbin{\underline{+}} \underline{a} \mathbin{\underline{\times}} \underline{\ell})\ .$$

Expanding the lifted primitives we obtain

$$\underline{List} \quad = \quad \Lambda\underline{a} \to \mu(\Lambda\underline{\ell}\ z \to 1 + \underline{a}\ z \times \underline{\ell}\ z)\ .$$

The lifted and the unlifted version of a type are closely related: if t has kind $\star \to \star$, then

$$t \quad = \quad (\uparrow t)\ Id\ . \tag{1}$$

This relation will be employed later when we define $poly\langle\cdot\rangle$ in terms of $poly\langle\!\langle\cdot\rangle\!\rangle$.

Using the notion of lifting we can easily characterize the set of normal forms of types of kind $\star \to \star$. Assume that we are given a type t of kind $\star \to \star$. Applying η-expansion we have $t = \Lambda a \to t\ a$. The body of the abstraction has kind $\star$ and we know from the previous section the shape of its normal form. The free variable, a, is treated as an additional constant of kind $\star$. Now, to make the passing of a explicit we abstract a out. The abstraction process replaces the primitive type constructors by their lifted versions and a by $Id = \Lambda a \to a$. This motivates the following characterization.

$$\begin{array}{l} T_{\star\to\star} \quad ::= \\ \quad Id \mid \underline{1} \mid \underline{Int} \mid (T_{\star\to\star} \mathbin{\underline{+}} T_{\star\to\star}) \mid (T_{\star\to\star} \mathbin{\underline{\times}} T_{\star\to\star}) \end{array}$$

6.2 Defining $\star \to \star$-indexed values

The characterization of normal forms suggests the following prototype for values indexed by types of kind $\star \to \star$.

$$\begin{array}{lcl} poly\langle f :: \star \to \star\rangle & :: & H\ f \\ poly\langle Id\rangle & = & poly_{Id} \\ poly\langle \underline{1}\rangle & = & poly_{\underline{1}} \\ poly\langle \underline{Int}\rangle & = & poly_{\underline{Int}} \\ poly\langle f_1 \mathbin{\underline{+}} f_2\rangle & = & poly_{\underline{+}}\ (poly\langle f_1\rangle)\ (poly\langle f_2\rangle) \\ poly\langle f_1 \mathbin{\underline{\times}} f_2\rangle & = & poly_{\underline{\times}}\ (poly\langle f_1\rangle)\ (poly\langle f_2\rangle) \end{array}$$

The type of $poly\langle f\rangle$ is given by $H\ f$, where H is a type constructor of kind $(\star \to \star) \to \star$. For instance, $size\langle f\rangle$ as defined in Sec. 3 has type $H\ f = \forall x. f\ x \to Int$. Thus, $size\langle f\rangle$ is not only a generic, but also a polymorphic function. Note that this combination is not cogent: the generic function $sum\langle f\rangle$, which sums a structure of integers, has the monomorphic type $f\ Int \to Int$.

In the rest of this section we consider further examples of generic values. The most paradigmatic example is probably the generic definition of map, which applies a given function to each element of type a in a given structure of type $f\ a$.

$$\begin{array}{lcl} map\langle f :: \star \to \star\rangle & :: & \forall a, b. (a \to b) \to (f\ a \to f\ b) \\ map\langle Id\rangle\ \varphi\ x & = & \varphi\ x \\ map\langle K\ t\rangle\ \varphi\ x & = & x \\ map\langle f_1 \mathbin{\underline{+}} f_2\rangle\ \varphi\ (Inl\ x_1) & = & Inl\ (map\langle f_1\rangle\ \varphi\ x_1) \\ map\langle f_1 \mathbin{\underline{+}} f_2\rangle\ \varphi\ (Inr\ x_2) & = & Inr\ (map\langle f_2\rangle\ \varphi\ x_2) \\ map\langle f_1 \mathbin{\underline{\times}} f_2\rangle\ \varphi\ (x_1, x_2) & = & (map\langle f_1\rangle\ \varphi\ x_1, map\langle f_2\rangle\ \varphi\ x_2) \end{array}$$

Note that the 'type pattern' $K\ t = \Lambda a \to t$ covers both $\underline{1}$ and $\underline{Int}$. Furthermore, note that Haskell provides a class *Functor* for mapping functions. Alas, the user must program instances of *Functor* by hand—which is for the most part tedious but sometimes quite involving.

The functions *size* is an instance of a more general concept termed reduction (Meertens, 1996). A reduction is a function of type $f\ a \to a$, which collapses a structure of values of type a into a single value of type a. To define a reduction we require two ingredients: a value $e :: a$ and a binary operation $op :: a \to a \to a$. Usually but not necessarily e is the neutral element of op.

$$\begin{array}{lcl}
reduce\langle f :: \star \to \star\rangle :: \forall a.a \to (a \to a \to a) \to (f\ a \to a) \\
reduce\langle f\rangle\ e\ op & = & red\langle f\rangle \\
\quad \mathbf{where} \\
\quad red\langle f :: \star \to \star\rangle & :: & f\ a \to a \\
\quad red\langle Id\rangle\ x & = & x \\
\quad red\langle K\ t\rangle\ x & = & e \\
\quad red\langle f_1 \underline{+} f_2\rangle\ (Inl\ x_1) & = & red\langle f_1\rangle\ x_1 \\
\quad red\langle f_1 \underline{+} f_2\rangle\ (Inr\ x_2) & = & red\langle f_2\rangle\ x_2 \\
\quad red\langle f_1 \underline{\times} f_2\rangle\ (x_1, x_2) & = & red\langle f_1\rangle\ x_1\ `op`\ red\langle f_2\rangle\ x_2
\end{array}$$

A number of useful functions can be defined in terms of $reduce\langle f\rangle$ and $map\langle f\rangle$, see, for instance, (Meertens, 1996; Jansson & Jeuring, 1998; Hinze, 1999d).

6.3 Specializing $\star \to \star$-indexed values

Promoting $poly\langle\cdot\rangle$ to types of arbitrary kind proceeds exactly as before except that we are now working in a higher realm, ie we work with lifted kinds and types. To begin with, the type of the promoted version is given by $H\langle\!\langle\cdot\rangle\!\rangle$, which is inductively defined as follows.

$$\begin{array}{lcl}
H\langle\!\langle u :: \uparrow\star\rangle\!\rangle & = & H\ u \\
H\langle\!\langle u :: \uparrow\kappa_1 \to \kappa_2\rangle\!\rangle & = & \forall x.H\langle\!\langle x :: \uparrow\kappa_1\rangle\!\rangle \to H\langle\!\langle u\ x :: \uparrow\kappa_2\rangle\!\rangle
\end{array}$$

The promoted version of $poly\langle\cdot\rangle$ reads

$$\begin{array}{lcl}
poly\langle\!\langle t :: \kappa\rangle\!\rangle\ \varrho & :: & H\langle\!\langle t :: \kappa\rangle\!\rangle \\
poly\langle\!\langle \underline{c}\rangle\!\rangle\ \varrho & = & poly_{\underline{c}} \\
poly\langle\!\langle \underline{x}\rangle\!\rangle\ \varrho & = & \varrho\ \underline{x} \\
poly\langle\!\langle t_1\ t_2\rangle\!\rangle\ \varrho & = & (poly\langle\!\langle t_1\rangle\!\rangle\ \varrho)\ (poly\langle\!\langle t_2\rangle\!\rangle\ \varrho) \\
poly\langle\!\langle \Lambda\underline{x} \to t\rangle\!\rangle\ \varrho & = & \lambda v \to poly\langle\!\langle t\rangle\!\rangle\ (\varrho(\underline{x} := v))\ . \\
poly\langle\!\langle \mu t\rangle\!\rangle\ \varrho & = & fix\ (poly\langle\!\langle t\rangle\!\rangle\ \varrho)
\end{array}$$

Note that $poly\langle\!\langle\cdot\rangle\!\rangle$ depends only on $poly_{\underline{c}}$ but not on $poly_{Id}$. The latter value is used when defining $poly\langle\cdot\rangle$ in terms of $poly\langle\!\langle\cdot\rangle\!\rangle$.

$$poly\langle t :: \star \to \star\rangle \quad = \quad poly\langle\!\langle \uparrow t :: \uparrow\star \to \star\rangle\!\rangle\ \epsilon\ poly_{Id}$$

Thus, in order to specialize $poly\langle t\rangle$ we specialize $poly\langle\!\langle\uparrow t\rangle\!\rangle$, which is defined by induction on the structure of lifted types. The resulting function has type $\forall x.H\ x \to H\ ((\uparrow t)\ x)$. Supplying $poly_{Id}$ as the first argument we obtain a value of type $H\ ((\uparrow t)\ Id)$, which is equal to $H\ t$.

One practical problem remains: the type of $poly\langle\!\langle\cdot\rangle\!\rangle$ uses lifted types. Consider, for instance, the type signature of $size\langle\!\langle \underline{List}\rangle\!\rangle$.

$$size\langle\!\langle \underline{List}\rangle\!\rangle \quad :: \quad \forall \underline{a}.(\forall x.\underline{a}\ x \to Int) \to (\forall x.\underline{List}\ \underline{a}\ x \to Int)$$

If we intend to present the specialized program to a Haskell compiler, we must find a way of expressing $\underline{List}$ in terms of $List$. For type constructors of kind $\star \rightarrow \star$ the relationship is quite simple: we have, for instance, $\underline{List}\ \underline{a} = \Lambda z \rightarrow List\ (\underline{a}\ z)$, ie $\underline{List}\ \underline{a}$ equals the type composition $List \cdot \underline{a}$. Given this relation we can rewrite the type signature above.

$$size\langle\!\langle \underline{List} \rangle\!\rangle \quad :: \quad \forall \underline{a}.(\forall x.\underline{a}\ x \rightarrow Int) \rightarrow (\forall x.List\ (\underline{a}\ x) \rightarrow Int)$$

For the general case we must delve a bit into the theory of combinators. To begin with we require type-level counterparts of the combinators K and S.

$$\begin{aligned} K\ t &= \Lambda x \rightarrow t \\ S\ t_1\ t_2 &= \Lambda x \rightarrow (t_1\ x)\ (t_2\ x) \end{aligned}$$

Next, we relate types of kind $\uparrow\kappa$ to types of kind $\star \rightarrow \kappa$.

Definition 4
The relation $(\sim_\kappa) \subseteq (\uparrow\kappa) \times (\star \rightarrow \kappa)$ is defined by induction on the structure of kinds.

$$\begin{aligned} t \sim_\star t' &\iff t = t' \\ t \sim_{\kappa_1 \rightarrow \kappa_2} t' &\iff \forall x, x'.(x \sim_{\kappa_1} x' \Rightarrow t\ x \sim_{\kappa_2} S\ t'\ x') \end{aligned}$$

In the base case '$\sim_\star$' relates types of the same kind, so we require them to be equal. 'Type functions' are related if related arguments are mapped to related results.

Proposition 1
Let t be a type of kind κ, then

$$\uparrow t \quad \sim_\kappa \quad K\ t\ .$$

For types of kind $\star$ we have $\uparrow t = K\ t$, ie the lifted version simply ignores its additional argument—which comes as little surprise. For types of kind $\star \rightarrow \star$ Prop. 1 implies $(\uparrow t)\ x = S\ (K\ t)\ x = t \cdot x$. Setting $x = Id$ we obtain $(\uparrow t)\ Id = t$, which shows that Eq. (1) is a direct consequence of Prop. 1. Now, let us apply Prop. 1 to rewrite the type signature of, say, $size\langle\!\langle \underline{Fix} \rangle\!\rangle$:

$$\begin{aligned} size\langle\!\langle \underline{Fix} \rangle\!\rangle \quad :: \quad & \forall \underline{f}.(\forall \underline{a}.(\forall x.\underline{a}\ x \rightarrow Int) \\ & \qquad\qquad \rightarrow (\forall x.\underline{f}\ \underline{a}\ x \rightarrow Int)) \\ & \qquad \rightarrow (\forall x.\underline{Fix}\ \underline{f}\ x \rightarrow Int) \end{aligned}$$

For the second-order type Fix Prop. 1 reads

$$\underline{Fix}\ \underline{f} = \Lambda x \rightarrow Fix\ (f\ x) \quad \Leftarrow \quad \underline{f}\ \underline{a} = \Lambda x \rightarrow (f\ x)\ (\underline{a}\ x)\ .$$

Given this relationship we can replace $\underline{Fix}\ \underline{f}$ and $\underline{f}\ \underline{a}$ in $size\langle\!\langle \underline{Fix} \rangle\!\rangle$'s type signature.

$$\begin{aligned} size\langle\!\langle \underline{Fix} \rangle\!\rangle \quad :: \quad & \forall f.(\forall \underline{a}.(\forall x.\underline{a}\ x \rightarrow Int) \\ & \qquad\qquad \rightarrow (\forall x.(f\ x)\ (\underline{a}\ x) \rightarrow Int)) \\ & \qquad \rightarrow (\forall x.Fix\ (f\ x) \rightarrow Int) \end{aligned}$$

Note that the rewrite involves the change of a bound variable: $\underline{f}$ of kind $(\star \rightarrow \star) \rightarrow (\star \rightarrow \star)$ is replaced by f of kind $\star \rightarrow (\star \rightarrow \star)$. This change is, however, perfectly fine since $\underline{f}$ is only instantiated to lifted types (by construction) and we know by Prop. 1 that lifted types of this kind can be expressed in terms of the original types. In other words, if we consistently change the types of the $\mathit{poly}\langle\!\langle\cdot\rangle\!\rangle$ functions, the resulting program is well-typed.

It is high time to consider some examples. Fig. 3 lists the specializations of *size* to the datatypes of Sec. 2. The code looks pretty similar to the code generated for *enc*, see Fig. 2. This is not surprising since the 'code generation' is largely independent of the kind of the type index. Only the types are more involved: since *size* has already a rank-1 type, $\mathit{size}\langle\!\langle \underline{F} \rangle\!\rangle$ is assigned a rank-$(n+1)$ type if F has an order-n kind.

In Sec. 5.3 we have already discussed two of the features the type system of the host language must support: rank-n types and polymorphic recursion. Here, we require an additional feature: a strong form of type constructor polymorphism. Consider the recursive call $\mathit{sizeSequ}\ (\mathit{sizeFork}\ \mathit{sizea})$ in the definition of $\mathit{sizeSequ}$: $\mathit{sizeSequ}$ requires an argument of type $\forall x.a\ x \to \mathit{Int}$ while $\mathit{sizeFork}\ \mathit{sizea}$ has type $\forall x.\mathit{Fork}\ (a'\ x) \to \mathit{Int}$. To determine whether the call is well-typed we must solve the equation $a\ x = \mathit{Fork}\ (a'\ x)$. Setting $a = \Lambda z \to \mathit{Fork}\ (a'\ z)$ the call and the definition of $\mathit{sizeSequ}$ can be type-checked. Clearly, some kind of higher-order unification is required here. Unfortunately, we know of no practical language that supports this feature. Haskell, for instance, uses a kinded first-order unification (Jones, 1995b), which reduces $a\ x = \mathit{Fork}\ (a'\ x)$ to $a = \mathit{Fork}$ and $x = a'\ x$. The latter equation is, however, not solvable. The lack of type constructor polymorphism is also noted in (Bird & Paterson, 1999a), which suggests that generalizing Haskell's type system might be worthwhile.

6.4 Generalizing to first-order kinds

So far we have considered generic values indexed by type constructors of kind $\star \to \star$. The generalization to type indices of first-order kind is straightforward. Assume that the type index has kind $\star^n \to \star$ (which abbreviates $\underbrace{\star \to \cdots \to \star}_{n\ \text{times}} \to \star$). We now redefine lifting and put $\uparrow\!\star = \star^n \to \star$ (the case $n = 1$ specializes to the preceding treatment). The lifted version of a primitive type constructor c of arity m is given by

$$
\begin{aligned}
&\underline{c}\, f_1\ \ldots\ f_m \\
&\quad = \quad \Lambda a_1\ \ldots\ a_n \to c\ (f_1\ a_1\ \ldots\ a_n)\ \ldots\ (f_m\ a_1\ \ldots\ a_n)\ .
\end{aligned}
$$

To define a value indexed by a type of kind $\star^n \to \star$ the programmer must provide cases for each of the lifted primitive type constructors and additionally for n 'projection types': $\pi_i = \Lambda a_1\ \ldots\ a_n \to a_i$ with $1 \leqslant i \leqslant n$. Specializing a generic value works exactly as before except that the 'initial call' now uses $\mathit{poly}_{\pi_1}, \ldots, \mathit{poly}_{\pi_n}$.

$$
\begin{aligned}
&\mathit{poly}\langle t :: \star^n \to \star\rangle \\
&\quad = \quad \mathit{poly}\langle\!\langle \uparrow t :: \uparrow\!\star^n \to \star\rangle\!\rangle\ \epsilon\ \mathit{poly}_{\pi_1}\ \ldots\ \mathit{poly}_{\pi_n}
\end{aligned}
$$

$$
\begin{array}{lcl}
\mathit{sizeList} & :: & \forall a.(\forall x.a\ x \to \mathit{Int}) \to (\forall x.\mathit{List}\ (a\ x) \to \mathit{Int}) \\
\mathit{sizeList}\ \mathit{sizea} & = & \mathit{sizel} \\
\quad \mathbf{where}\ \mathit{sizel}\ \mathit{Nil} & = & 0 \\
\qquad\qquad \mathit{sizel}\ (\mathit{Cons}\ x\ \mathit{xs}) & = & \mathit{sizea}\ x + \mathit{sizel}\ \mathit{xs} \\
\mathit{sizeRose} & :: & \forall a.(\forall x.a\ x \to \mathit{Int}) \to (\forall x.\mathit{Rose}\ (a\ x) \to \mathit{Int}) \\
\mathit{sizeRose}\ \mathit{sizea} & = & \mathit{sizer} \\
\quad \mathbf{where}\ \mathit{sizer}\ (\mathit{Branch}\ x\ \mathit{ts}) & = & \mathit{sizea}\ x + \mathit{sizeList}\ \mathit{sizer}\ \mathit{ts} \\
\mathit{sizeGRose} & :: & \forall f.(\forall b.(\forall x.b\ x \to \mathit{Int}) \to (\forall x.(f\ x)\ (b\ x) \to \mathit{Int})) \\
 & & \quad \to (\forall a.(\forall x.a\ x \to \mathit{Int}) \\
 & & \qquad\qquad \to (\forall x.\mathit{GRose}\ (f\ x)\ (a\ x) \to \mathit{Int})) \\
\mathit{sizeGRose}\ \mathit{sizef}\ \mathit{sizea} & = & \mathit{sizeg} \\
\quad \mathbf{where}\ \mathit{sizeg}\ (\mathit{GBranch}\ x\ \mathit{ts}) & = & \mathit{sizea}\ x + \mathit{sizef}\ \mathit{sizeg}\ \mathit{ts} \\
\mathit{sizeFix} & :: & \forall f.(\forall b.((\forall x.b\ x \to \mathit{Int}) \to (\forall x.(f\ x)\ (b\ x) \to \mathit{Int}))) \\
 & & \quad \to (\forall x.\mathit{Fix}\ (f\ x) \to \mathit{Int}) \\
\mathit{sizeFix}\ \mathit{sizef} & = & \mathit{sizer} \\
\quad \mathbf{where}\ \mathit{sizer}\ (\mathit{In}\ x) & = & \mathit{sizef}\ \mathit{sizer}\ x \\
\mathit{sizeBaseList} & :: & \forall a.(\forall x.a\ x \to \mathit{Int}) \\
 & & \quad \to (\forall b.(\forall x.b\ x \to \mathit{Int}) \\
 & & \qquad\qquad \to (\forall x.\mathit{BaseList}\ (a\ x)\ (b\ x) \to \mathit{Int})) \\
\mathit{sizeBaseList}\ \mathit{sizea}\ \mathit{sizeb}\ \mathit{Nil} & = & 0 \\
\mathit{sizeBaseList}\ \mathit{sizea}\ \mathit{sizeb}\ (\mathit{Cons}\ x\ y) & & \\
 & = & \mathit{sizea}\ x + \mathit{sizeb}\ y \\
\mathit{sizeFork} & :: & \forall a.(\forall x.a\ x \to \mathit{Int}) \to (\forall x.\mathit{Fork}\ (a\ x) \to \mathit{Int}) \\
\mathit{sizeFork}\ \mathit{sizea}\ (\mathit{Node}\ x_1\ x_2) & = & \mathit{sizea}\ x_1 + \mathit{sizea}\ x_2 \\
\mathit{sizeSequ} & :: & \forall a.(\forall x.a\ x \to \mathit{Int}) \to (\forall x.\mathit{Sequ}\ (a\ x) \to \mathit{Int}) \\
\mathit{sizeSequ}\ \mathit{sizea}\ \mathit{Empty} & = & 0 \\
\mathit{sizeSequ}\ \mathit{sizea}\ (\mathit{Zero}\ s) & = & \mathit{sizeSequ}\ (\mathit{sizeFork}\ \mathit{sizea})\ s \\
\mathit{sizeSequ}\ \mathit{sizea}\ (\mathit{One}\ x\ s) & = & \mathit{sizea}\ x + \mathit{sizeSequ}\ (\mathit{sizeFork}\ \mathit{sizea})\ s \\
\mathit{sizeMapFork} & :: & \forall m.(\forall w.(\forall x.w\ x \to \mathit{Int}) \to (\forall x.(m\ x)\ (w\ x) \to \mathit{Int})) \\
 & & \quad \to (\forall v.(\forall x.v\ x \to \mathit{Int}) \\
 & & \qquad\qquad \to (\forall x.\mathit{MapFork}\ (m\ x)\ (v\ x) \to \mathit{Int})) \\
\mathit{sizeMapFork}\ \mathit{sizem}\ \mathit{sizev}\ (\mathit{TrieFork}\ \mathit{tf}) & & \\
 & = & \mathit{sizem}\ (\mathit{sizem}\ \mathit{sizev})\ \mathit{tf} \\
\mathit{sizeMapSequ} & :: & \forall m.(\forall w.(\forall x.w\ x \to \mathit{Int}) \to (\forall x.(m\ x)\ (w\ x) \to \mathit{Int})) \\
 & & \quad \to (\forall v.(\forall x.v\ x \to \mathit{Int}) \\
 & & \qquad\qquad \to (\forall x.\mathit{MapSequ}\ (m\ x)\ (v\ x) \to \mathit{Int})) \\
\mathit{sizeMapSequ}\ \mathit{sizem}\ \mathit{sizev}\ (\mathit{TrieSequ}\ \mathit{te}\ \mathit{tz}\ \mathit{to}) & & \\
 & = & \mathit{sizev}\ \mathit{te} \\
 & + & \mathit{sizeMapSequ}\ (\mathit{sizeMapFork}\ \mathit{sizem})\ \mathit{sizev}\ \mathit{tz} \\
 & + & \mathit{sizem}\ (\mathit{sizeMapSequ}\ (\mathit{sizeMapFork}\ \mathit{sizem})\ \mathit{sizev})\ \mathit{to}
\end{array}
$$

Fig. 3. Specializing *size* to the types of Sec. 2.

7 Towards Haskell

The specialization of generic values as described in Sec. 5.3 and 6.3 places high demands on the type system: it requires polymorphic recursion, rank-n types, and a strong form of type constructor polymorphism. Haskell 98 (Peyton Jones & Hughes, 1999), for instance, only supports polymorphic recursion. In this section we show that the requirement for rank-n types can be alleviated since they can be encoded using so-called dictionaries. This encoding also provides an interesting link to Haskell's type classes.

Recent extensions of Haskell as implemented in GHC (Peyton Jones, 1998) and in Hugs 98 (Jones, 1998) provide rank-2 type signatures and local universal quantification in datatypes. Using the latter feature we can circumvent rank-n types. The idea is very simple: instead of passing a polymorphic value directly as an argument we pass a dictionary that contains the value as the single component. Of course, we require not only a single dictionary but a kind-indexed family of dictionaries. The following definitions introduce suitable dictionaries for the *enc* function.

$$\begin{aligned}
&\textbf{type}\ \mathit{Enc}\langle\star\rangle\ a &&= a \to \mathit{Bin}\\
&\textbf{type}\ \mathit{Enc}\langle\kappa_1 \to \kappa_2\rangle\ a &&= \forall x.\mathit{EncD}\langle\kappa_1\rangle\ x \to \mathit{Enc}\langle\kappa_2\rangle\ (a\ x)\\
&\textbf{data}\ \mathit{EncD}\langle\kappa\rangle\ a &&= \mathit{EncD}_\kappa\{\,\mathit{enc}_\kappa :: \mathit{Enc}\langle\kappa\rangle\ a\,\}
\end{aligned}$$

The promoted function $\mathit{enc}\langle\!\langle t :: \kappa\rangle\!\rangle$ now has type $\mathit{Enc}\langle\kappa\rangle\ t$. Fig. 4 displays the specializations of *enc* for some types. The dictionary translation is interesting in at least two respects. First, it suggests an easy way of dealing with mutually recursive generic definitions, see (Hinze, 1999b) for examples. We simply use dictionaries with multiple entries, one for each recursive function. Second, it relates generic definitions to Haskell's type classes (Hall *et al.*, 1996), which use a similar implementation. Recall that an overloaded function is translated into a non-overloaded function that takes as an additional argument the dictionary containing the operations of the class. For instance, the class definition

$$\textbf{class}\ \mathit{Eq}\ a\ \textbf{where}\ (==), (/=) :: a \to a \to \mathit{Bool}$$

gives rise to the following dictionary type.

$$\textbf{data}\ \mathit{EqD}\ a \quad = \quad \mathit{EqD}\{\,(==), (/=) :: a \to a \to \mathit{Bool}\,\}$$

Each instance declaration of the form **instance** $\mathit{Eq}\ T$ defines an element of $\mathit{EqD}\ T$ and each declaration of the form **instance** $(\mathit{Eq}\ a) \Rightarrow \mathit{Eq}\ (T\ a)$ defines a function of type $\forall a.\mathit{EqD}\ a \to \mathit{EqD}\ (T\ a)$. Thus, generic definitions and type classes use the same mechanism on the implementation level. Using the **deriving** construct instance declarations can even be automatically generated for user-defined, first-order kinded datatypes. However, instance declarations are too limited to handle types of higher-order kinds. For instance, **deriving** *Eq* for *MapFork* or *MapSequ* fails with a compile-time error in Haskell 98. In (Peyton Jones *et al.*, 1997) an extension of the class system is described which allows for more general instance contexts. If we use this extension, which has been implemented in GHC and in Hugs 98, to derive *Eq* for *MapSequ*, both compilers loop! The problem is that the class mechanism tries to manage with a single dictionary type though in general a family of dictionaries is required. Thus, generic definitions generalize type classes in some respects: they allow us to define values generically for *all* types—while class instances must be programmed by hand, except for some predefined classes like *Eq*—and the specialization is not restricted to first-order kinded types. Note that instance declarations can be mimicked in our framework by so-called ad-hoc

$$
\begin{array}{lcl}
\mathbf{data}\ \mathit{EncD}_{\star}\ t & = & \mathit{EncD}_{\star}\{\mathit{enc}_{\star} :: t \to \mathit{Bin}\} \\
\mathbf{data}\ \mathit{EncD}_{\star\to\star}\ t & = & \mathit{EncD}_{\star\to\star}\{\mathit{enc}_{\star\to\star} :: \forall x.\mathit{EncD}_{\star}\ x \to (t\ x \to \mathit{Bin})\} \\
\mathit{encGRose} & :: & \forall f, a.\mathit{EncD}_{\star\to\star}\ f \to \mathit{EncD}_{\star}\ a \to (\mathit{GRose}\ f\ a \to \mathit{Bin}) \\
\mathit{encGRose}\ \mathit{df}\ \mathit{da} & = & \mathit{encg} \\
\quad \mathbf{where}\ \mathit{encg}\ (\mathit{GBranch}\ x\ \mathit{ts}) & & \\
 & = & \mathit{enc}_{\star}\ \mathit{da}\ x \mathbin{+\!\!+} \mathit{enc}_{\star\to\star}\ \mathit{df}\ (\mathit{EncD}_{\star}\{\mathit{enc}_{\star} = \mathit{encg}\})\ \mathit{ts} \\
\mathit{encFork} & :: & \forall a.\mathit{EncD}_{\star}\ a \to (\mathit{Fork}\ a \to \mathit{Bin}) \\
\mathit{encFork}\ \mathit{da}\ (\mathit{Node}\ x_1\ x_2) & = & \mathit{enc}_{\star}\ \mathit{da}\ x_1 \mathbin{+\!\!+} \mathit{enc}_{\star}\ \mathit{da}\ x_2 \\
\mathit{encSequ} & :: & \forall a.\mathit{EncD}_{\star}\ a \to (\mathit{Sequ}\ a \to \mathit{Bin}) \\
\mathit{encSequ}\ \mathit{da}\ \mathit{Empty} & = & [\mathtt{0}] \\
\mathit{encSequ}\ \mathit{da}\ (\mathit{Zero}\ s) & = & \mathtt{1} : \mathtt{0} : \mathit{encSequ}\ (\mathit{EncD}_{\star}\{\mathit{enc}_{\star} = \mathit{encFork}\ \mathit{da}\})\ s \\
\mathit{encSequ}\ \mathit{da}\ (\mathit{One}\ x\ s) & = & \mathtt{1} : \mathtt{1} : \mathit{enc}_{\star}\ \mathit{da}\ x \\
 & \mathbin{+\!\!+} & \mathit{encSequ}\ (\mathit{EncD}_{\star}\{\mathit{enc}_{\star} = \mathit{encFork}\ \mathit{da}\})\ s
\end{array}
$$

Fig. 4. Dictionary implementation of *enc*.

definitions, see (Hinze, 1999b). For instance, an ad-hoc compression scheme for lists, which yields better compression rates than the generic scheme, can be defined by

$$
\begin{array}{lcl}
\mathit{enc}\langle \mathit{List}\ a\rangle\ \mathit{xs} & = & \mathit{encInt}\ (\mathit{length}\ \mathit{xs}) \\
 & \mathbin{+\!\!+} & \mathit{concat}\ (\mathit{map}\ (\mathit{enc}\langle a\rangle)\ \mathit{xs})\ .
\end{array}
$$

This equation extends the definition of *enc* given in Sec. 3 and specifies an exception to the general scheme.

8 Related work

Generic programming The concept of generic functional programming trades under a variety of names: F. Ruehr refers to this concept as *structural polymorphism* (Ruehr, 1992; Ruehr, 1998), T. Sheard calls generic functions *type parametric* (Sheard, 1993), C.B. Jay and J.R.B. Cocket use the term *shape polymorphism* (Jay & Cocket, 1994), and J. Jeuring invented the word *polytypism* (Jeuring & Jansson, 1996).

The mainstream of generic programming is based on the initial algebra semantics of datatypes, see, for instance (Hagino, 1987), and puts emphasis on general recursion operators like *map* and catamorphisms (folds). In (Sheard, 1991) several variations of these operators are informally defined and algorithms are given that specialize these functions for given datatypes. The programming language *Charity* (Cockett & Fukushima, 1992) automatically provides *map* and catamorphisms for each user-defined datatype. Since general recursion is not available, Charity is strongly normalizing. *Functorial ML* (Jay *et al.*, 1998) has a similar functionality, but a different background. It is based on the theory of *shape polymorphism*, in which values are separated into shape and contents. The polytypic programming language extension *PolyP* (Jansson & Jeuring, 1997)—already mentioned in the introduction—offers a special construct for defining generic functions. The generic definitions are similar to ours (modulo notation) except that the generic programmer must additionally consider cases for type composition and for type recursion (see (Hinze, 1999e) for a more detailed comparison).

All the approaches are restricted to first-order kinded, regular datatypes (or even subsets of this class). One notable exception is the work of F. Ruehr (Ruehr, 1992), who presents a higher-order language based on a type system related to ours (only type recursion is missing). Genericity is achieved through the use of type patterns which are interpreted at run-time.

This paper can be regarded as a successor to (Hinze, 1999e), where a similar approach restricted to first-order kinded types is presented. A companion paper (Hinze, 1999b) contains a concrete proposal for a generic programming extension of Haskell, which is based on the theoretical framework developed here.

Type systems Several variations of the type system given in Sec. 4 have been described in the literature, see (Barendregt, 1992) for a good survey article. For instance, if we drop the fixpoint operator from the type language, we obtain the system $\lambda\underline{\omega}$, which forms one corner of Barendregt's λ-cube. Dropping type application and abstraction yields the system $\lambda\mu$, which supports structural equivalence of types. This system is, however, restricted to types of kind $\star$ and cannot handle parametric types. To the best of the author's knowledge the combination of $\lambda\underline{\omega}$ and $\lambda\mu$ (with a polymorphic fixpoint operator) is original.

Type inference algorithms for languages with generic constructs have been developed by F. Ruehr (Ruehr, 1992), C.B. Jay *et.al.* (Jay *et al.*, 1998), and P. Jansson and J. Jeuring (Jansson & Jeuring, 1997). Note that our system does not permit type reconstruction in general. Consider the expression *size xs* where *xs* has type *List* (*Rose Int*). Should *size* count the number of integers or the number of rose trees in the list?

9 Conclusion

We have presented a new approach to generic functional programming, which is both simpler—from the generic programmer's point of view—and considerably more general than previous work—the complete type system of Haskell is covered (previous approaches were limited to first-order kinded, regular types). The basic idea is to model types by terms of the simply typed λ-calculus augmented by a family of recursion operators. Specializing a generic value can be seen as an interpretation of simply typed λ-terms. The generated code places high demands on the type system of the underlying language: polymorphic recursion, rank-n types, and a strong form of type constructor polymorphism are required. We have seen that rank-n types can be circumvented using a dictionary translation, which provides an interesting link to Haskell's type classes. In particular, generic definitions generalize Haskell's **deriving** construct.

Acknowledgements

I am grateful to Ian Bayley, Johan Jeuring, Lambert Meertens, and Fritz Ruehr for suggesting dozens of improvements regarding contents and presentation.

References

Abadi, M., Cardelli, L., Pierce, B., & Rémy, D. (1995). Dynamic typing in polymorphic languages. *Journal of functional programming*, **5**(1), 111–130.

Augustsson, Lennart. (1999). Cayenne – a language with dependent types. *SIGPLAN notices*, **34**(1), 239–250.

Barendregt, H. P. (1984). *The lambda calculus — Its syntax and semantics*. revised edn. North-Holland, Amsterdam New York Oxford.

Barendregt, H.P. (1992). Lambda calculi with types. *Pages 118–309 of:* Abramsky, S., Gabbay, Dov M., & Maibaum, T.S.E. (eds), *Handbook of logic in computer science, volume 2, background: Computational structures*. Clarendon Press, Oxford.

Bird, Richard. (1998). *Introduction to functional programming using Haskell*. 2nd edn. London: Prentice Hall Europe.

Bird, Richard, & Meertens, Lambert. (1998). Nested datatypes. *Pages 52–67 of:* Jeuring, J. (ed), *Fourth international conference on mathematics of program construction, MPC'98, Marstrand, Sweden*. Lecture Notes in Computer Science, vol. 1422. Springer-Verlag.

Bird, Richard, & Paterson, Ross. (1999a). De Bruijn notation as a nested datatype. *Journal of functional programming*, **9**(1), 77–91.

Bird, Richard, & Paterson, Ross. (1999b). Generalised folds for nested datatypes. *Formal aspects of computing*. To appear.

Cockett, Robin, & Fukushima. 1992 (June). *About Charity*. Yellow Series Report 92/480/18. Dept. of Computer Science, Univ. of Calgary.

Courcelle, Bruno. (1983). Fundamental properties of infinite trees. *Theoretical computer science*, **25**(2), 95–169.

Girard, Jean-Yves. (1972). *Interprétation fonctionelle et élimination des coupures dans l'arithétique d'ordre supérieur*. Ph.D. thesis, Université Paris VII.

Hagino, T. (1987). *Category theoretic approach to data types*. Ph.D. thesis, University of Edinburgh.

Hall, Cordelia V., Hammond, Kevin, Peyton Jones, Simon L., & Wadler, Philip L. (1996). Type classes in Haskell. *Acm transactions on programming languages and systems*, **18**(2), 109–138.

Hinze, Ralf. 1998 (December). *Numerical representations as higher-order nested datatypes*. Tech. rept. IAI-TR-98-12. Institut für Informatik III, Universität Bonn.

Hinze, Ralf. (1999a). Generalizing generalized tries. *Journal of functional programming*. Accepted for publication.

Hinze, Ralf. (1999b). *A generic programming extension for Haskell*. In submission.

Hinze, Ralf. 1999c (April). *Manufacturing datatypes*. Tech. rept. IAI-TR-99-5. Institut für Informatik III, Universität Bonn.

Hinze, Ralf. (1999d). Polytypic functions over nested datatypes. *Discrete mathematics and theoretical computer science*. Accepted for publication.

Hinze, Ralf. 1999e (February). *Polytypic programming with ease*. Tech. rept. IAI-TR-99-2. Institut für Informatik III, Universität Bonn.

Jansson, Patrik, & Jeuring, Johan. (1997). PolyP—a polytypic programming language extension. *Pages 470–482 of: Conference record 24th ACM SIGPLAN-SIGACT symposium on principles of programming languages, POPL'97, Paris, France.* ACM-Press.

Jansson, Patrik, & Jeuring, Johan. (1998). PolyLib—A library of polytypic functions. Backhouse, Roland, & Sheard, Tim (eds), *Informal proceedings workshop on generic programming, WGP'98, Marstrand, Sweden.* Department of Computing Science, Chalmers University of Technology and Göteborg University.

Jay, C.B., & Cocket, J.R.B. (1994). Shapely types and shape polymorphism. *Pages 302–316 of:* Sanella, D. (ed), *Programming languages and systems — ESOP'94: 5th european symposium on programming, Edinburgh, UK, proceedings.* Lecture Notes in Computer Science, vol. 788. Berlin: Springer-Verlag.

Jay, C.B., Bellè, G., & Moggi, E. (1998). Functorial ML. *Journal of functional programming*, **8**(6), 573–619.

Jeuring, Johan, & Jansson, Patrik. (1996). Polytypic programming. *Pages 68–114 of:* Launchbury, J., Meijer, E., & Sheard, T. (eds), *Tutorial text 2nd international school on advanced functional programming, Olympia, WA, USA.* Lecture Notes in Computer Science, vol. 1129. Springer-Verlag.

Jones, Mark. (1998). *A technical summary of the new features in Hugs 1.3c.* unpublished.

Jones, Mark P. (1995a). A system of constructor classes: overloading and implicit higher-order polymorphism. *Journal of functional programming*, **5**(1), 1–35.

Jones, Mark P. (1995b). Functional programming with overloading and higher-order polymorphism. *Pages 97–136 of: First international spring school on advanced functional programming techniques.* Lecture Notes in Computer Science, vol. 925. Springer-Verlag.

McCracken, Nancy Jean. (1984). The typechecking of programs with implicit type structure. *Pages 301–315 of:* Kahn, Gilles, MacQueen, David B., & Plotkin, Gordon D. (eds), *Semantics of data types: International symposium, Sophia-Antipolis, France.* Lecture Notes in Computer Science, vol. 173. Springer-Verlag.

Meertens, Lambert. (1996). Calculate polytypically! *Pages 1–16 of:* Kuchen, H., & Swierstra, S.D. (eds), *Proceedings 8th international symposium on programming languages: Implementations, logics, and programs, PLILP'96, Aachen, Germany.* Lecture Notes in Computer Science, vol. 1140. Springer-Verlag.

Meijer, Erik, & Hutton, Graham. (1995). Bananas in space: Extending fold and unfold to exponential types. *Pages 324–333 of: Conference record 7th ACM SIGPLAN/SIGARCH and IFIP WG 2.8 international conference on functional programming languages and computer architecture, FPCA'95, La Jolla, San Diego, CA, USA.* ACM-Press.

Mycroft, Alan. (1984). Polymorphic type schemes and recursive definitions. *Pages 217–228 of:* Paul, M., & Robinet, B. (eds), *Proceedings of the international symposium on programming, 6th colloquium, Toulouse, France.* Lecture Notes in Computer Science, vol. 167.

Okasaki, Chris. (1998). *Purely functional data structures.* Cambridge University Press.

Okasaki, Chris. (1999). From fast exponentiation to square matrices: An adventure in types. *Proceedings of the 1999 ACM SIGPLAN international conference on functional programming, Paris, France.* To appear.

Peyton Jones, Simon. (1998). *Explicit quantification in Haskell.* Available from `http://research.microsoft.com/Users/simonpj/Haskell/quantification.html`.

Peyton Jones, Simon, & Hughes, John (eds). 1999 (February). *Haskell 98 — A non-strict, purely functional language.*

Peyton Jones, Simon, Jones, Mark, & Meijer, Erik. (1997). Type classes: Exploring the design space. *Proceedings of the Haskell workshop.*

Ruehr, Fritz. (1998). Structural polymorphism. Backhouse, Roland, & Sheard, Tim (eds), *Informal proceedings workshop on generic programming, WGP'98, Marstrand, Sweden, 18 june 1998.* Dept. of Computing Science, Chalmers Univ. of Techn. and Göteborg Univ.

Ruehr, Karl Fritz. (1992). *Analytical and structural polymorphism expressed using patterns over types.* Ph.D. thesis, University of Michigan.

Sheard, Tim. (1991). Automatic generation and use of abstract structure operators. *Acm transactions on programming languages and systems*, **13**(4), 531–557.

Sheard, Tim. (1993). *Type parametric programming.* Tech. rept. Oregon Graduate Institute of Science and Technology, Portland, OR, USA.

Wadler, Philip. (1989). Theorems for free! *Pages 347–359 of: The fourth international conference on functional programming languages and computer architecture (fpca'89), London, UK.* Addison-Wesley Publishing Company.

Yang, Zhe. (1999). Encoding types in ML-like languages. *SIGPLAN notices*, **34**(1), 289–300.