Liberating Effects with Rows and Handlers

Daniel Hillerström
The University of Edinburgh, UK
daniel.hillerstrom@ed.ac.uk

Sam Lindley
The University of Edinburgh, UK
sam.lindley@ed.ac.uk

Abstract

Algebraic effects and effect handlers provide a modular abstraction for effectful programming. They support user-defined effects, as in Haskell, in conjunction with direct-style effectful programming, as in M.L. They also present a structured interface to programming with delimited continuations.

In order to be modular, it is necessary for an effect type system to support extensible effects. Row polymorphism is a natural abstraction for modelling extensibility at the level of types. In this paper we argue that the abstraction required to implement extensible effects and their handlers is exactly row polymorphism.

We use the Links functional web programming language as a platform to substantiate this claim. Links is a natural starting point as it uses row polymorphism for polymorphic variants, records, and its built-in effect types. It also has infrastructure for manipulating continuations. Through a small extension to Links we smoothly add support for effect handlers, making essential use of rows in the frontend and first-class continuations in the backend.

We demonstrate the usability of our implementation by modelling the mathematical game of Nim as an abstract computation. We interpret this abstract computation in a variety of ways, illustrating how rows and handlers support modularity and smooth composition of effectful computations.

We present a core calculus of row-polymorphic effects and handlers based on a variant of A-normal form used in the intermediate representation of Links. We give an operational semantics for the calculus and a novel generalisation of the CEK machine that implements the operational semantics, and prove that the two coincide.

Categories and Subject Descriptors D.3.3 [Language Constructs and Features]; F.3.2 [Semantics of Programming Languages]

Keywords algebraic effects; effect handlers; effect typing; abstract machine semantics; operational semantics; delimited control

1. Introduction

Algebraic effects [29] and effect handlers [30] are a more modular alternative to monads for managing user-defined computational effects [6, 15, 17]. Effect handlers generalise exception handlers, providing a mechanism for interpreting arbitrary algebraic effects, and they present a structured interface to programming with delimited continuations.

As a simple example consider a choice effect given by a single effectful operation:

\[ \text{Choose} : \text{Bool} \]

We can write an abstract computation \( M \) that invokes \text{Choose}, independently of specifying the meaning of \text{Choose}. We can then handle \( M \) in multiple ways. For instance, we can define a handler \text{allResults} that interprets the operation as nondeterministic choice, returning a list of every possible outcome of \( M \). We can also define a different handler \text{coin} that interprets \text{Choose} as random choice, returning a single value that depends on all of the random choices made in \( M \). Where effect handlers really come into their own as a programming abstraction is when we start composing them: we can handle some effects while forwarding all others, using row types to statically track the forwarded effects. We make extensive use of handler composition in Section 2.

Many existing implementations of effect handlers are Haskell libraries. Notable examples include the effect handlers library of Kammar et al. [15], the extensible effects library of Kiselyov et al. [17] and Kiselyov and Ishii [16], and implementations based on variants of Swierstra’s data types à la carte technique [34], such as the work of Wu et al. [36] on scoped effect handlers. Another notable effect handlers library is the Idris effects library [6]. Each of these libraries uses its own sophisticated encoding of an abstraction which amounts to a restricted form of row polymorphism. In this paper we present the first, to our knowledge, implementation of effect handlers using actual Remy-style row polymorphism [33].

Links [8] is a functional programming language for building web applications. The defining feature of Links is that it provides a single source language that targets all three tiers of a web application: client, server, and database. Links source code is translated into an intermediate representation (IR) based on A-normal form [11]. For the client, the IR is compiled to JavaScript. For the server, the IR is interpreted using a variant of the CEK machine [10]. For the database, the IR is translated into an SQL query, taking advantage of the effect type system and the subformula property to guarantee query generation [22].

Links is a strict language with Hindley-Milner type inference. Links has a row type system for polymorphic variants, records, and its built-in effect types for concurrency and database integration [8, 22]. It also has support for manipulating first-class continuations, a feature that is central to implementing effect handlers.

The row-polymorphic effect type system and continuation support make Links a natural choice for experimenting with row-based algebraic effects and effect handlers. We have implemented an effect handlers extension to Links. Currently, it is supported only on the server-side. The frontend to our implementation makes essential use of row polymorphism, while the backend is implemented as a novel generalisation of the CEK machine.

Our main contributions are as follows.

• An implementation of effect handlers using Remy-style row polymorphism [33].
operation Move: curried. Unlike in JavaScript, functions are statically typed and the function capture the essence of our implementation. Section 5 discusses implementation details. Section 6 discusses related work. Finally, Section 7 concludes and discusses future work.

2. Programming with Handlers in Links
To demonstrate that rows and handlers provide an elegant and modular abstact for effectful programming, we use a simplified version of the mathematical game Nim [5] as a running example.

Starting from an abstract representation of the game, we iteratively extend it with cheat detection and high score tracking capabilities through smooth composition of handlers, without needing to change the initial representation.

2.1 The Game of Nim and Effect Rows
The game of Nim is played between two players: Alice and Bob. The game begins with a heap of \( n \) sticks. The players alternate to take one, two, or three sticks from the heap. Alice makes the first move. The player who takes the last stick wins the game.

We abstract over the notion of making a move by defining it as an abstract effectful operation \( \text{Move}: \text{(Player,Int)} \rightarrow \text{Int} \), where \( \text{Player} \) is variant type with two constructors \( \text{Alice} \) and \( \text{Bob} \). The first parameter to \( \text{Move} \) is the active player, the second parameter is the current number of sticks on the heap. We will discuss the meaning of the braces (\{\}) prefix on the arrow shortly.

In Links, abstract operations like \( \text{Move} \), are invoked using the \( \text{do} \) primitive, for instance

\[
\text{do Move}(\text{Alice},3)
\]

invokes the \( \text{Move} \) operation with values \( \text{Alice} \) and \( 3 \). Operation names, data constructors, and type aliases all begin with a capital letter. Records, variants, and effect signatures all have row types. All typing is structural in Links, thus it is unnecessary to declare a row occupant, such as an operation, before use. However, we consider it good practice to wrap the invocation of operations as any effects they might have are ultimately captured by the interpreter.

In the type signature, the function arrow (\( \rightarrow \)) is prefixed by a row enclosed in curly braces. This row is the effect signature, or effect row, of the function. The presence of \( \text{Move} \) in the effect row indicates that the function may perform the \( \text{Move} \) operation. Furthermore, the effect row is equipped with an effect variable \( e \), which can be instantiated with additional operations. This means that \( \text{move} \) may be invoked in the scope of additional effects. We say an effect row is closed if it has no effect variable, and open if it does. In general an effect row consists of an unordered collection of operation specifications and an optional effect variable. An operation specification either specifies that an operation is admissible (or present) and has a particular type signature, or that it is absent, or that it is polymorphic in its presence. We discuss the use of absence in Section 2.5.

The effect row on the type signature of the \( \text{Move} \) operation itself is empty, denoted by a pair of braces (\{\}). This is always the case for abstract operations as any effects they might have are ultimately conferred by their handlers.

The Nim game is modelled as two mutually recursive functions \( \text{aliceTurn} \) and \( \text{bobTurn} \). Here we show \( \text{aliceTurn} \):

\[
\text{sig aliceTurn} : \ (\text{Int} \rightarrow \text{Comp}(:\text{Row},a)) = () \ { |e| \rightarrow a;}
\]

The keyword \textit{typename} \textit{Comp} is used to define type aliases. The \textit{Comp} type captures our notion of abstract computation, it is an alias for a thunk with an open effect row and return type \( a \). The notation \( :\text{Row} \) denotes that the type variable \( a \) has kind \text{Row}, which is the kind of effect variables.

The \textit{game} function begins a game with a given number of sticks. Alice starts:

\[
\text{sig game} : (\text{Int}) \rightarrow \text{Comp}((\text{Move}:\text{(Player,Int)} \rightarrow \text{Int}|e), \text{Player}) \ \text{fun game}(n)() \ {\text{aliceTurn}(n)}
\]

2.2 Strategies and Handlers
In general, algebraic effects come with equations [29], but as with most other implementations of effect handlers, we do not consider equations. Thus, on their own, abstract operations have no meaning; handlers give them a semantics. We can use handlers to encode particular strategies for Alice and Bob by interpreting the operation \( \text{Move} \). We start by considering the \textit{perfect} strategy, defined by \( \text{ps}(n) = n \bmod 4 \), where \( n \) is the number of sticks left.
in the game. If player $p$ adopts the perfect strategy, then $p$ is guaranteed to win if on $p$'s turn $n$ is not divisible by four. We define a handler $pp$ (short for perfect-vs-perfect), that assigns the perfect strategy to both players

```
1  sig pp : (Comp({Move:(Player,Int) {}-> Int|e}, a)) -> Comp({Move- |e}, a)
2  fun pp(m) {} {
3      handle(m) {
4          case Return(x) -> x
5          case Move(p,n,k) -> k(maximum(1, n 'mod' 4)) }}
```

We describe the handler line by line.

**Lines 1 and 2** give the type of $pp$: it takes a computation, which may invoke the $Move$ operation, and yields another computation where the operation is absent (denoted by $Move$). The computation returns a value of type $a$. We may omit this type signature altogether as the type system is capable of inferring the appropriate types.

**Lines 3 and 4** begin the definition. The curried function $pp$ wraps the actual handler, that is applied to the argument $m$ using the $handle$ construct, which specifies how to interpret abstract operations through a sequence of clauses.

**Line 5** is a return clause. It defines how to handle the final return value of the input computation. In this case, this value is simply returned as is.

**Line 6** is an operation clause. It expresses how to handle $Move$. In general, an operation clause takes the form $Op(p_1, \ldots, p_n, k) \rightarrow M$, where $p_1, \ldots, p_n$ are patterns that bind the operation parameters and $k$ is a pattern that binds the continuation of the computation in $M$. In this case $p$ and $n$ are bound to the active player and number of sticks in the heap, respectively. The continuation is invoked with the perfect strategy, irrespective of the player. The input type of the continuation is given by the output type of $Move$. The return type of the continuation is the return type of the handler.

The handler $pp$ returns a computation. For convenience we define an auxiliary function to $run$ (or force) a thunk:

```
1  sig run : (Comp({}, a)) {}-> a
2  fun run(m) {} { m }  
```

We can now compute the winner of a game in which both players play the perfect strategy:

```
links> run(pp(game(7)));
Alice : Player
links> run(pp(game(12)));
Bob : Player
```

**Syntactic Sugar** The input computation in $pp$ is immediately supplied to $handle$. This abstract-over-handle idiom arises frequently, so Links provides syntactic sugar for it. We can give a more succinct definition of $pp$ using the $handle$ keyword:

```
handler pp {
    case Return(x) -> x
    case Move(_,n,k) -> k(maximum(1, n 'mod' 4)) }  
```

We may choose to name the input computation by adding a parameter to the $handler$ keyword in square brackets, for instance, $handler[m]$ $pp$ $\{ \ldots \}$. Naming the computation can be useful if we want to invoke the wrapper function recursively with the computation from inside the handler.

**2.3 Game Trees and Multi-shot Continuations**

The handler $pp$ computes the winner of a particular game. It only considers one scenario in which both players play the same strategy, but we can use handlers to compute other data about a game.

![Figure 1: Game Tree Generated by run(gametree(game(3)))](image-url)

For instance, we can give an interpretation that computes the $game$ tree. Figure 1 shows an example game tree. Each node represents the active player, and each edge corresponds to a possible move for that player. We define a game tree inductively:

```
typename GTree = [ [Take:Player, ( [Int, GTree] ) ] [Winner:Player] ];
```

The syntax $[ |...| ]$ denotes a (polymorphic) variant type in Links in which components of the variant type are delimited by the pipe symbol ($|$). A $Take$ node includes the active player and a list of possible moves, where each move is paired with the subsequent game tree. A $Winner$ leaf denotes the winner of a game.

We define a handler $gametree$ that generates game trees:

```
1  sig gametree : (Comp({Move:(Player,Int) {}-> Int|e}, Player)) -> Comp({Move- |e}, GTree)
2  handler gametree {} {
3      case Return(x) -> Winner(x)
4      case Move(p,n,k) ->
5          var subgames = map(k, validMoves(n));
6          var subtrees = zip([1,2,3], subgames);
7          Take(p, subtrees) 
```

The effect signatures of $pp$ and $gametree$ are identical, though their interpretations of $Move$ differ. The return clause wraps the winning move in a leaf node. The operation clause for $Move$ refines the move as a node in the game tree. The $var$ keyword denotes a let binding. The crucial part is the invocation of $map$ which applies the continuation multiple times, once for each valid move, enumerating every possible subtree. The function $validMoves$ is a simple filter:

```
fun validMoves(n) {
    filter(fun(m) { m <= n }, [1,2,3])
```

Figure 1 shows the game tree generated by the handler when $n = 3$.

**2.4 Cheating and Forwarding**

Thus far we have considered a single operation $Move$, but in general we may allow arbitrary algebraic operations. We could define a monolithic handler that interprets every operation that may occur in a computation. However, a more modular alternative is to define a series of fine-grained, specialised handlers that each handle a particular operation, and then compose them together to fully interpret a computation. Fortunately, handlers compose seamlessly. Composed handlers can cooperate to interpret an abstract computation. Each handler operates on a subset of the abstract operations, leaving the remainder for other handlers. Consequently, we obtain a considerable amount of flexibility as it becomes possible to reinterpret computations by swapping in and out individual handlers.
We defer a full discussion of the role that row polymorphism plays during composition of handlers until Section 2.5. In this section we omit type signatures for handlers and instead focus on the dynamic semantics of handler composition by augmenting the game model with a cheat detection mechanism. A cheating strategy might remove all remaining sticks from the heap, winning in a single move. We introduce an additional operation Cheat to signal that a cheater has been detected. The operation is parameterised by the player, who was caught cheating:

\[
\text{sig cheat : (Player) \{Cheat:(Player) \{} \rightarrow \text{Zero}\}, a\}} \rightarrow \_ \\
\text{fun cheat(p) \{ switch (do Cheat(p)) \{ \} \}}
\]

The Cheat operation can never return a value as its return type is the empty type Zero. Thus invoking Cheat amounts to raising an exception. Concretely, an operation clause for Cheat can never invoke the continuation. The switch statement construct pattern matches on the expression e, through a possibly empty list of clauses. We define an exception handler that interprets Cheat by outputting an error message and exiting the program:

\[
\text{handler report \{(} \\
\text{case Return(x) \rightarrow x} \\
\text{case Cheat(Alice,\_\_) \rightarrow error("Alice cheated!")} \\
\text{case Cheat(Bob,\_\_) \rightarrow error("Bob cheated!")} \\
\text{\}}}
\]

We implement the heart of the cheat detection machinery as a handler:

\[
\text{handler checker \{(} \\
\text{case Return(x) \rightarrow x} \\
\text{case Move(p,n,k) \rightarrow m = move(p,n);} \\
\text{if (m \text{'elem' validMoves(n)) k(m)} \\
\text{else cheat(p) \}}}
\]

In order to detect cheating, the handler analyses the active player's move. If it is legal then the game continues. Otherwise, the Cheat operation is invoked to signal that cheating has occurred. We may compose pp with report and check to give an interpretation of a game in which no player may cheat. To make handler composition syntactically lightweight we define a pipeline operator -<- for composing handlers and another operator <- for applying a computation to a pipeline of handlers. Infix binary operators are defined using the op keyword in Links:

\[
\text{op f \text{-<} g \{ f(m) \rightarrow g(f(m)) \}} \\
\text{op f \text{-<} m \{ f(m) \}}
\]

The operators are meant to indicate that unhandled operations are forwarded from right to left in a pipeline. In order to run a pipeline of handlers, we may apply the function run:

\[
\text{links\> run \text{-<} pp \text{-<} report \text{-<} checker \text{-<} game(7);} \\
\text{Alice : Player}
\]

The Cheat operation is never invoked as both players play the same legal strategy. Let us define another handler that assigns the perfect strategy to Alice and a cheating strategy to Bob:

\[
\text{handler pc \{(} \\
\text{case Return(x) \rightarrow x} \\
\text{case Move(Alice,n,k) \rightarrow k(maximum(1, n \text{'mod' 4}))} \\
\text{case Move(Bob,n,k) \rightarrow k(n) \}}
\]

Now the cheat detection handler catches Bob:

\[
\text{links\> run \text{-<} pc \text{-<} report \text{-<} checker \text{-<} game(7);} \\
\text{Bob : Player}
\]

The order of composition is important as pc and checker both handle moves. Bob gets away with cheating if we swap the two handlers:

\[
\text{links\> run \text{-<} pp \text{-<} report \text{-<} checker \text{-<} pc \text{-<} game(?)}
\]

Here we also use pp, because the type system does not know that checker is not performing any Move operations.

### 2.5 Composition and Row Polymorphism

In this section we discuss the typing of composed handlers. First, consider the type signature for the report handler:

\[
\text{sig report : } \langle \text{Comp}(\langle \text{Move:(Player)} \rightarrow \text{Int}\rangle, a) \rightarrow \text{Comp}(\langle \text{Cheat:(Player)} \rightarrow \text{Zero}\rangle, a) \rangle
\]

In general, a handler accepts a computation as input and produces another computation as output. Moreover, handlers have open input and output effect rows, which both share the same effect variable. As a consequence both rows mention the same operation names. However, some of these operation names may be marked as absent or polymorphic in their presence. In the output effect row of report, the syntax Cheat(p) denotes that the operation is presence polymorphic. The type variable p can be instantiated to be either present with a particular type (:A) or absent (.). Presence polymorphism is useful for seamless composition of handlers. We illustrate why by type checking the composition:

\[
\text{var f = run \text{-<} (pp \text{-<} report)}
\]

Recall the type signature for pp:

\[
\text{sig pp : } \langle \text{Comp}(\langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle, a) \rightarrow \text{Comp}(\langle \text{Move:(Player)} \rightarrow \text{Zero}\rangle, a) \rangle
\]

The output effects of report must be compatible with the input effects of pp, therefore the composition gives rise to the following unification constraint:

\[
\langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle \sim \langle \text{Cheat:(Player)} \rightarrow \text{Zero}\rangle
\]

which is solved by introducing a fresh effect variable e3, instantiating e1 to \langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle, and e2 to \langle \text{Cheat:(Player)} \rightarrow \text{Zero}\rangle. Thus the unified type is:

\[
\langle \text{Move:(Player)} \rightarrow \text{Int}\rangle \rightarrow \text{Cheat:(Player)} \rightarrow \text{Zero}
\]

Note that with rows the order of operations is unimportant. The output of pp must be compatible with the input to run, giving rise to the constraint:

\[
\langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle \sim \langle \text{Cheat:(Player)} \rightarrow \text{Zero}\rangle
\]

which is solved by instantiating p to - and e3 to {}. Thus f has type:

\[
\text{Comp}(\langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle, a) \rightarrow a
\]

Now we consider the type signature for the checker handler:

\[
\text{sig checker : } \langle \text{Comp}(\langle \text{Cheat:(Player)} \rightarrow \text{Zero}\rangle, a) \rightarrow \text{Comp}(\langle \text{Move:(Player,Int)} \rightarrow \text{Int}\rangle, a) \rangle
\]

The reason Cheat:(Player) \rightarrow Zero appears in the input effect is because Cheat is not handled, so if a Cheat operation is forwarded it must have the correct type.

**Remy’s IML** What we actually require for soundness is that if the Cheat effect is present then it must have type (Player) \rightarrow Zero, as that is the type it has in the output. In a slightly more refined system along the lines of Remy’s IML [33], we could specify this as follows:
The handler uniformly interprets returns either Alice dishonestly. The computation using this handler Bob will be equally likely to play honestly or 0 floating point value between 0 and 1 (both inclusive).

We use Links' built-in random number generator, which returns a Thus Bob wins only when he cheats.

Alternatively, we can replace Bob's oracular powers with a fair coin and let him perform a coin flip to decide which strategy to pick. We use Links' built-in random number generator, which returns a floating point value between 0.0 and 1.0 (both inclusive):

We can give a nondeterministic interpretation of Choose that infuses Bob with oracular powers that enable him to explore both alternatives. We define it as a handler allResults

The handler wraps the result of the input computation into a singleton list. In the Choose-clause the handler accumulates the results of both the alternatives by invoking the continuation twice.

Now, we can put everything together:

Thus Bob wins only when he cheats.

We represent high scores as an association list and refer to a value of this type as the game state:

typename GState = [(Player,Int)];

We define an initial state s0 = [(Alice,0),(Bob,0)]. We now need a mechanism to update the game state when a game finishes. Recall that game(n) returns a computation whose type is:

The computation returns the winner of the game. We may exploit the fact that the return clauses of handlers are invoked in the order of composition. Therefore, we define a simple post-processing handler that contains only a Return case to update the scoreboard:

The function updateScore updates the score of the given player p by 1, given game state, in which the number of wins for the given player p has been incremented by one. The handler then reads and updates the game state. The composition scoreUpdater(game(n)) causes the effect row to grow accordingly:

In a similar fashion, we define a handler that prints the scoreboard:

The function printBoard is impure as it prints an ASCII representation of the given game state to standard out. To make matters more interesting, we add replay functionality, which we implement by invoking a handler recursively on its input computation:

The replay handler reevaluates the computation n precisely n times. Note, that the handler's effect signature is an empty, open row which means that it forwards every operation that might occur to subsequent handlers. Now we can wire everything together:

Figure 2 shows a possible output. In the same manner, we can effortlessly merge the cheating infrastructure into the pipeline without changing the underlying computation.
cheating strategies.
plays the perfect strategy and Bob chooses between the perfect and cheating strategies.

Dually, the variant type record type \langle \ldots \rangle denotes tagged sums constrained by presence \rho. More generally, the kind \Comp of computation types is given by a value type A and an effect E, which specifies the operations that the computation may perform.

\begin{align*}
\text{Values } & \quad V, W ::= x \mid \lambda x^A.M \mid \lambda \alpha^K.M \\
\text{Computations } & \quad M, N ::= VW \mid VA \\
& \quad \quad | \langle \ell \mapsto M \rangle = V \text{ in } N \\
& \quad \quad | \text{ case } V \{ x \mapsto M; y \mapsto N \} | \text{ absurd}^\ell V \\
& \quad \quad | \text{ return } V \\
& \quad \quad | \text{ let } x \leftarrow M \text{ in } N \\
& \quad \quad | (\text{do } \ell V)^\rho \\
& \quad \quad | \text{ handle } M \text{ with } H
\end{align*}

Handlers \quad H ::= \{ \text{return } x \mapsto M \} \\
& \quad \quad | \{ \ell x k \mapsto M \} \cup H

\begin{itemize}
\item \textbf{Computation Types} \ A computation type \text{A}!E is given by a value type \text{A} and an effect \text{E}, which specifies the operations that the computation may perform.
\item \textbf{Row Types} \ Effect types, records and variants are defined in terms of rows. A row type embodies a collection of distinct labels, each of which is annotated with a presence type. A presence type indicates whether a label is present with some type \text{A} (\text{Pre}(\text{A})), absent (\text{Abs}) or polymorphic in its presence \theta.
\item \\
\end{itemize}

Row types are either \textit{closed} or \textit{open}. A closed row type ends in \text{\ell}, whilst an open row type ends with a \textit{row variable} \text{\rho}. Furthermore, a closed row term can have only the labels explicitly mentioned in its type. Conversely, the row variable in an open row can be instantiated with additional labels. We identify rows up to reordering of labels. For instance, we consider the following two rows equivalent:

\begin{align*}
\ell_1 : P_1; \cdots ; \ell_n : P_n \equiv \ell_a : P_a; \cdots ; \ell_1 : P_1
\end{align*}

The unit and empty type are definable in terms of row types. We define the unit type as the empty, closed record, that is, \{\}. Similarly, we define the empty type as the empty, closed variant \[\].

Usually, we usually omit the \cdot for closed rows.

\begin{itemize}
\item \textbf{Handler Types} \ A handler type \text{C} \Rightarrow \text{D} is given by an input computation type \text{C} and an output computation type \text{D}.
\item \textbf{Kinds} \ We have six kinds: Type, Comp, Effect, Row, Presence, Handler, which classify value types, computation types, effect types, row types, presence types, and handler types, respectively. Row kinds are annotated with a set of labels \text{L}. The kind of a complete row is \text{Row}_{\text{L}}. More generally, the kind \text{Row}_{\text{L}} denotes a partial row that cannot mention the labels in \text{L}.
\item \textbf{Type Variables} \ We let \alpha, \rho \text{ and } \theta \text{ range over type variables}. By convention we use \alpha \text{ for value type variables or for type variables of unspecified kind, } \rho \text{ for type variables of row kind, and } \theta \text{ for type variables of presence kind.}
\item \textbf{Type and Kind Environments} \ Type environments map term variables to their kinds and kind environments map type variables to their kinds.
\end{itemize}

3.2 Names

3.2 Terms

The terms are given in Figure 4. We let \text{x, y, z, k} range over term variables. By convention, we use \text{k} to denote continuation names.

The syntax partitions terms into values, computations and handlers. Value terms comprise variables (\text{x}), lambda abstraction (\text{\lambda x^A.M}), type abstraction (\text{\lambda \alpha^K.M}), and the introduction forms for records and variants. Records are introduced using the empty record \text{\{\}} and record extension \text{\langle \ell = V; W \rangle}, whilst variants are introduced using injection \text{\langle \ell V \rangle}^\rho, which injects a field with label \text{\ell}.
and value \( V \) into a row whose type is \( R \). We include the row type annotation in order to support bottom-up type reconstruction.

All elimination forms are computation terms. Abstraction and type abstraction are eliminated using application (\( V W \)) and type application (\( V A \)) respectively. The record eliminator (let \( (\ell = x. y) \mapsto V \) in \( N \)) splits a record \( V \) into \( x \), the value associated with \( \ell \), and \( y \), the rest of the record. Non-empty variants are eliminated using the case construct (case \( V \) \{ \( \ell \) \( x \mapsto M ; y \mapsto N \) \}), which evaluates the computation \( M \) if the tag of \( V \) matches \( \ell \). Otherwise it falls through to \( y \) and evaluates \( N \). The elimination form for empty variants is (absurd \( V \)). A trivial computation (return \( V \)) returns value \( V \). The expression (let \( x \mapsto M \) in \( N \)) evaluates \( M \) and binds the result value to \( x \) in \( N \).

The construct \((\text{do } \ell \ V)^{\ell}\) invokes an operation \( \ell \) with value argument \( V \). The handle construct (handle \( M \) with \( H \)) runs a computation \( M \) with handler definition \( H \). A handler definition \( H \) consists of a return clause return \( x \mapsto M \) and a possibly empty set of operation clauses \( \{ \ell_i \mapsto M_i \} \). The return clause defines how to handle the final return value of the handled computation, which is bound to \( x \) in \( M \). The \( i \)-th operation clause binds the operation parameter to \( \ell_i \) and the continuation \( k_i \) in \( M_i \).

We write \( \text{Id}(M) \) for handle \( M \) with \( \text{return } x \mapsto \lambda \). We write \( H(\text{return}) \) for the return clause of \( H \) and \( \ell(\ell) \) for the set of either zero or one operation clauses in \( H \) that handle the operation \( \ell \). We write \( \text{dom}(H) \) for the set of operations handled by \( H \). As our calculus is Church-style, we annotate various term forms with type or kind environment \( \Delta \) and value type \( \alpha \) and type environment \( \Gamma \). We write \( \text{FTV}(\Gamma) \) to be the set of free type variables in \( \Gamma \).

The kinding and typing rules are mostly straightforward. The interesting typing rules are T-HANDLE and the two handler rules. The T-HANDLE rule states that handle \( M \) with \( H \) produces a computation of type \( D \) given that the computation \( M \) has type \( C \), and that \( H \) is a handler that transforms a computation of type \( C \) into another computation of type \( D \).

The T-HANDLER rule is crucial. The effect rows on the computation type \( C \) and the output computation type \( D \) must share the same suffix \( R \). This means that the effect row of \( D \) must explicitly mention each of the operations \( \ell_i \), whether that be to say that an \( \ell_i \) is present with a given type signature, absent, or polymorphic in its presence. The row \( R \) describes the operations that are forwarded. It may include a row-variable, in which case an arbitrary number of effects may be forwarded by the handler. The typing of the return clause is straightforward. In the typing of each operation clause, the continuation returns the output computation type \( D \). Thus, we are here defining deep handlers [15] in which the handler is implicitly wrapped around the continuation, such that any subsequent operations are handled uniformly by the same handler. The Links implementation also supports shallow handlers [15], in which the continuation is instead annotated with the input effect and one has to explicitly reinvoke the handler after applying the continuation inside an operation clause.

### 3.4 Operational Semantics

We give a small-step operational semantics for \( \lambda^\text{et} \). Figure 7 displays the operational rules. The reduction relation \( \rightarrow \) is defined on computation terms. The statement \( M \rightarrow M' \) reads: term \( M \) reduces to term \( M' \) in a single step. Most of the rules are standard. We use evaluation contexts to focus on the active expression. The interesting rules are the handler rules.

We write \( BL(\ell) \) for the set of operation labels bound by \( \ell \).

\[
BL(\ell) = \emptyset \quad BL(\text{let } x \leftarrow \ell \text{ in } N) = BL(\ell) \quad BL(\text{handle } \ell \text{ with } H) = BL(\ell) \cup \text{dom}(H)
\]

The rule S-HANDLE-RET invokes the return clause of a handler. The rule S-HANDLE-OP handles an operation by invoking the appropriate operation clause. The constraint \( \ell \not\in BL(\ell) \) ensures that no inner handler inside the evaluation context is able to handle the operation: thus a handler is able to reach past any other inner handlers that do not handle \( \ell \). In our abstract machine semantics we realise this behaviour using explicit forwarding operations, but more efficient implementations are perfectly feasible.

**Remark** The attentive reader may have noticed that the two rules S-LET and S-HANDLE-RET are strikingly similar. Indeed, we could omit let bindings altogether once any let binding

\[
\text{let } x \leftarrow M \text{ in } N
\]

may be simulated by a trivial handler:

\[
\text{handle } M \text{ with } \{ \text{return } x \rightarrow N \}
\]

Nevertheless, we elect to keep let bindings in \( \lambda^\text{et} \). As we discuss in Section 4, one reason for doing so is in order to facilitate more efficient implementations.

We write \( R^+ \) for the transitive closure of relation \( R \). Subject reduction and type soundness for \( \lambda^\text{et} \) are standard.
Theorem 3.1 (Subject Reduction). If \( \Delta; \Gamma \vdash M : A!E \) and \( M \rightarrow M' \), then \( \Delta; \Gamma \vdash M' : A!E \).

There are two ways in which a computation can terminate. It can either successfully return a value, or it can get stuck on an unhandled operation.

Definition 3.2. We say that computation term \( N \) is normal with respect to effect \( E \), if \( N \) is either of the form \( \text{return } V \), or \( E[\text{do } \ell \ W] \), where \( \ell \in E \) and \( \ell \notin BL(E) \).

If \( N \) is normal with respect to the empty effect \( \{\} \), then \( N \) has the form \( \text{return } V \).

Theorem 3.3 (Type Soundness). If \( \vdash M : A!E \), then there exists \( \vdash N : A!E \), such that \( M \rightarrow^* N \), and \( N \) is normal with respect to effect \( E \).

4. Abstract Machine Semantics

In this section we present an abstract machine semantics for \( \lambda_{\text{eff}} \) that is closely related to the actual implementation of effect handlers in Links. We prove that the abstract machine simulates the operational semantics in the sense that each reduction in the small step semantics corresponds exactly to a finite sequence of one or more steps of the abstract machine.

The Links interpreter is based on a CEK-style abstract machine [10] and operates directly on ANF terms [11]. The standard
CEK machine operates on configurations which are triples of the form \( C \mid E \mid K \).

- The control \( C \) is the expression currently being evaluated.
- The environment \( E \) binds the free variables.
- The continuation \( K \) instructs the machine what to do once it is done evaluating the current term in the \( C \) component.

In order to accommodate handlers we generalise the CEK machine. The syntax of abstract machine states is given in Figure 8. Just like in the standard CEK machine, a standard configuration \( C = (M \mid \gamma \mid \kappa) \) of our abstract machine is a triple of a computation term \( M \), an environment \( \gamma \) mapping free variables to values, and a continuation \( \kappa \). However, our continuations differ from the standard machine. On the one hand, they are somewhat simplified, due to our strict separation between computations and values. On the other hand, they have considerably more structure in order to accommodate effects and handlers. In order to account for forwarding of unhandled operations, configurations occasionally gain an additional continuation argument.

Values consist of function closures, type function closures, records, variants, and captured continuations. A continuation \( \kappa \) consists of a stack of continuation frames \( [d_{\ell}, \ldots, d_1] \). We choose to annotate captured continuations with their input type in order to make the results of Section 4.1 easier to state. Intuitively, each continuation frame \( \delta = (\sigma, \chi) \) represents the pure continuation \( \sigma \), corresponding to a sequence of let bindings, inside a particular handler closure \( \chi \). A pure continuation is a stack of pure continuation frames. A pure continuation frame \( (\gamma, x, N) \) closes a let-binding \( \text{let } x = N \text{ in } \gamma \). A handler closure \( (\gamma, H) \) closes a handler definition \( H \) over environment \( \gamma \). We write \([\emptyset]\) for an empty stack, \( x :: s \) for the result of pushing \( x \) on top of stack \( s \), and \( s ++ s' \) for the concatenation of stack \( s \) on top of \( s' \). We use pattern matching to deconstruct stacks.

The abstract machine semantics is given in Figure 9. The transition function is given by \( \Rightarrow \). This depends on an interpretation function \( [\cdot] \) for values. The machine is initialised \( (M \mid \gamma \mid \kappa) \) by placing a term in a configuration alongside the empty environment and identity continuation \( \kappa_0 \). The rules \( (M-\text{APP}), (M-\text{TYAPP}), (M-\text{SPLIT}), \) and \( (M-\text{CASE}) \) eliminate the evaluation of values. Note that \( (M-\text{APP}) \) handles application of both closures and of captured continuations. The rules \( (M-\text{LET}) \) and \( (M-\text{HANDLE}) \) extend the current continuation with let bindings and handlers respectively. The rule \( (M-\text{RETCASE}) \) binds a returned value if there is a pure continuation in the current continuation frame. The rule \( (M-\text{RETHANDLER}) \) invokes the return clause of a handler if there is no pure continuation in the current continuation frame, but there is a handler. The rule \( (M-\text{RETPOP}) \) returns a final value if the continuation is empty. The rule \( (M-\text{OP}) \) switches to a special four place configuration in order to handle an operation. The fourth component of the configuration is an auxiliary forwarding continuation, which keeps track of the continuation frames through which the operation has been forwarded. It is initialised to be empty. The rule \( (M-\text{OP-HANDLER}) \) uses the current handler to handle an operation if the label matches one of the operation clauses of the current handler. The captured continuation is assigned the forwarding continuation with the current continuation frame appended to the bottom of it. The rule \( (M-\text{OP-FORWARD}) \) appends the current continuation frame onto the bottom of the forwarding continuation. Notice that if the main continuation is empty then the machine gets stuck. This occurs when an operation is unhandled, and the forwarding continuation describes the succession of handlers that have failed to handle the operation along with any pure continuations that were encountered along the way.

Assuming the input is a well-typed closed computation term \( \vdash M : A \langle E \rangle \), the machine will either return a value of type \( A \), or it will get stuck failing to handle an operation appearing in \( E \). We now make the correspondence between the operational semantics and the abstract machine more precise.

**Remark** Taking advantage of the observation that let bindings can be simulated by handlers, it is straightforward to define a variant of our abstract machine with flat continuations in which each continuation frame is just a handler closure without a pure continuation. In practice, we believe it is helpful to distinguish between pure and impure parts of the continuation. Apart from anything else, it supports more efficient implementations by reducing the number of continuation frames that a forwarded effect must bubble through.

### 4.1 Correctness

The \( (M-\text{INIT}) \) rule immediately provides a canonical way to map a computation term onto the abstract machine. A more interesting question is how to map an arbitrary configuration to a computation term. Figure 10 describes such a mapping \( [\cdot] \) from configurations to terms via a collection of mutually recursive functions defined on configurations, continuations, computation terms, handler definitions, value terms, and values. We write \( \text{dom}(\gamma) \) for the domain of \( \gamma \), and \( \gamma \backslash \{x_1, \ldots, x_n\} \) for the restriction of environment \( \gamma \) to \( \text{dom}(\gamma) \backslash \{x_1, \ldots, x_n\} \).

The \( [\cdot] \) function enables us to classify the abstract machine reduction rules according to how they relate to the operational semantics. The rules \( (M-\text{INIT}) \) and \( (M-\text{RETCASE}) \) are concerned only with initial input and final output, neither of which is a feature of the operational semantics, so we can ignore them. The rules \( (M-\text{APPCONT}), (M-\text{LET}), (M-\text{HANDLE}), (M-\text{OP}), \) and \( (M-\text{OP-FORWARD}) \) are administrative in the sense that \( [\cdot] \) is invariant under these rules. This leaves the \( \beta \)-rules \( (M-\text{APP}), (M-\text{TYAPP}), (M-\text{SPLIT}), (M-\text{CASE}), (M-\text{RETCASE}), (M-\text{RETHANDLER}), \) and \( (M-\text{OP-HANDLER}) \). Each of these corresponds directly with performing a reduction in the operational semantics. We write \( \beta \rightarrow_\beta \) for administrative steps, \( \beta \rightarrow_\beta \) for \( \beta \)-steps, and \( \beta \rightarrow_\beta \) for a sequence of steps of the form \( \rightarrow_\beta \cdots \rightarrow_\beta \).

The following lemma describes how we can simulate each reduction in the operational semantics by a sequence of administrative steps followed by one \( \beta \)-step in the abstract machine. The idea is to represent a computation term \( M \) by the equivalence class of configurations \( C \) such that \( (C) = \text{Id}(M) \). The \( \text{Id} \) wrapper captures the top-level identity handler.

**Lemma 4.1.** If \( M \leadsto N \), then for any \( C \) such that \( (C) = \text{Id}(M) \) there exists \( C' \) such that \( C \equiv C' \) and \( (C') = \text{Id}(N) \).

**Proof.** By induction on the derivation of \( M \leadsto N \). If \( (C) = \text{Id}(M) \), then the underlying structure of the term in the configuration \( C \) must be the same as \( M \), as \( [\cdot] \) is homomorphic on computation terms. Some value subterms of \( M \) may appear in the environment,
Identity continuation
\[ \kappa_0 = [[[\[], (\emptyset, \{return x \mapsto x\})]]] \]

Transition function
\[
\begin{align*}
M - INIT & \quad M \rightarrow (M \mid \emptyset \mid \kappa_0) \\
M - APP & \quad \langle V \mid W \mid \gamma \mid \kappa \rangle \rightarrow (M \mid \gamma \cdot [x \mapsto [W]\gamma] \mid \kappa), \quad \text{if } [V]\gamma = (\gamma', \lambda x^\delta M) \\
M - APPCONT & \quad \langle V \mid W \mid \gamma \mid \kappa \rangle \rightarrow \langle V \cdot \gamma \mid \kappa' \rangle, \quad \text{if } [V]\gamma = (\kappa')^2 \\
M - TYAPP & \quad \langle M \cdot A \mid \gamma \mid \kappa \rangle \rightarrow (M[A/\alpha] \mid \gamma^\prime \mid \kappa), \quad \text{if } [V]\gamma = (\gamma', \Lambda\alpha^\delta M) \\
M - SPLIT & \quad \langle \text{let } (\ell = x; y) = V \ln N \mid \gamma \mid \kappa \rangle \rightarrow \langle M \mid \gamma \cdot [x \mapsto y; y \mapsto w] \mid \kappa \rangle, \quad \text{if } [V]\gamma = (\ell = v; w) \\
M - CASE & \quad \langle \text{case } V \{\ell \mapsto M \mid y \mapsto N \mid \gamma \mid \kappa \rangle \rightarrow \langle M \mid \gamma \cdot [x \mapsto y] \mid \kappa \rangle, \quad \text{if } [V]\gamma = (\ell = v \text{ and } \ell \neq \ell') \\
M - LET & \quad \langle \text{let } x \gets M \ln N \mid \gamma \mid (\sigma, \chi) :: \kappa \rangle \rightarrow (M \mid \gamma \mid ([\sigma, x, N] :: \sigma, \chi) :: \kappa) \\
M - HANDLE & \quad \langle \text{handle } M \ln H \mid \gamma \mid \kappa \rangle \rightarrow (M \mid \gamma \mid ([], (\gamma, H)) :: \kappa) \\
M - RETCONT & \quad \langle \text{return } V \mid \gamma \mid (\gamma', x, N) :: \sigma, \chi) :: \kappa \rangle \rightarrow \langle N \mid \gamma' \cdot [x \mapsto [V]\gamma] :: \sigma, \chi :: \kappa \rangle \\
M - RETHANDLER & \quad \langle \text{return } V \mid \gamma \mid ([], (\gamma', H)) :: \kappa \rangle \rightarrow (M \mid \gamma' \cdot [x \mapsto [V]\gamma] :: \kappa), \quad \text{if } H(\text{return}) = \{\text{return } x \mapsto M\} \\
M - RETTOP & \quad \langle \text{return } V \mid \gamma \mid [] \rangle \rightarrow [V]\gamma \\
M - E & \quad \langle \text{do } \ell \ln V \rangle^E \mid \gamma \mid \kappa \rangle \rightarrow \langle \text{do } \ell \ln V \cdot \gamma \mid \kappa \mid [] \rangle^E \\
M - E - HANDLE & \quad \langle \text{do } \ell \ln V \rangle^E \mid \gamma \mid (\sigma, \chi', H) :: \kappa :: \kappa' \rangle \rightarrow (M \mid \gamma' \cdot [x \mapsto [V]\gamma] \mid \kappa', \ell' :: [\sigma, (\gamma', H)] :: \kappa \rangle, \quad \text{if } \text{ell} : A \cdot B \in E \text{ and } H(\ell) = \{\text{ell} \cdot k \mapsto M\} \\
M - E - FORWARD & \quad \langle \text{do } \ell \ln V \rangle^E \mid \gamma \mid (\sigma, \chi', H) :: \kappa :: \kappa' \rangle \rightarrow \langle \text{do } \ell \ln V \cdot \gamma \mid \kappa \mid \kappa' :: [\sigma, (\gamma', H)] \rangle^E, \quad \text{if } H(\ell) = 0 \\
\end{align*}
\]

Value interpretation
\[
\begin{align*}
[x]\gamma &= \gamma(x) & [\lambda x^\delta M]\gamma &= (\gamma', \lambda x^\delta M) & [\Lambda\alpha^\delta M]\gamma &= (\gamma, \Lambda\alpha^\delta M) \\
[\emptyset]\gamma &= \emptyset & [\ell = V; W]\gamma &= \ell = [V]\gamma; [W]\gamma & [\ell V]^E\gamma &= (\ell \ ln V)^E \gamma \\
\end{align*}
\]

Figure 9: Abstract Machine Semantics

Configurations
\[
\langle M \mid \gamma \mid \kappa \rangle = \{\kappa\}(\{M\}\gamma) \quad \langle M \mid \gamma \mid \kappa \mid \kappa' \rangle = \{\kappa' \mapsto \kappa\}(\{M\}\gamma) = \{\kappa\}(\{\kappa\}(\{M\}\gamma))
\]

Continuations
\[
\langle [] \rangle M = M \\
\langle (\gamma, x, N) :: \sigma, x \rangle :: \kappa M = \langle (\sigma, x) :: \kappa \rangle(\text{let } x \gets M \ln (\{N\}(\gamma \\{\emptyset\})) \\
\langle [], (\gamma, H) \rangle :: \kappa M = \{\kappa\}(\text{handle } M \ln (H\gamma))
\]

Computation terms
\[
\begin{align*}
\langle V \mid W \rangle \gamma &= \langle V \rangle \gamma \langle W \rangle \gamma \\
\langle V \cdot A \rangle \gamma &= \langle V \rangle \gamma A \\
\langle \text{let } (\ell = x; y) = V \ln N \rangle \gamma &= \langle \text{let } (\ell = x; y) = \langle V \rangle \gamma \ln (N \gamma \{x, y\}) \rangle \\
\langle \text{case } V \{\ell \mapsto M \mid y \mapsto N \rangle \gamma &= \langle \text{case } V \gamma \{\ell \mapsto \{M \gamma \{x\}\}; y \mapsto \{N \gamma \{y\}\}\} \\
\langle \text{return } V \rangle \gamma &= \langle \text{return } V \rangle \gamma \\
\langle \text{let } x \gets M \ln N \rangle \gamma &= \langle \text{let } x \gets \{M \rangle \gamma \ln (N \gamma \{x\}) \\
\langle \text{do } \ell \ln V \rangle \gamma &= \langle \text{do } \ell \ln V \rangle \gamma \\
\langle \text{handle } M \ln H \rangle \gamma &= \langle \text{handle } M \rangle \gamma \ln (H \gamma)
\end{align*}
\]

Handler definitions
\[
\langle \{\text{return } x \mapsto M\}\rangle \gamma = \{\text{return } x \mapsto \{M\}(\gamma \\{x\})\} \\
\langle \{\ell \cdot k \mapsto M \} \ODOT H \gamma &= \{\ell \cdot k \mapsto \{M\}(\gamma \\{x, k\}) \ODOT (H \gamma) \gamma
\]

Value terms and values
\[
\begin{align*}
{x}\gamma &= \emptyset, \quad \text{if } x = x \in \text{dom}(\gamma) \\
{x}\gamma &= x, \quad \text{if } x \notin \text{dom}(\gamma) \\
[\lambda x^\delta M]\gamma &= \lambda x^\delta.\{M\}(\gamma \{x\}) & [\lambda x^\delta M]\gamma &= \lambda x^\delta.\{M\}(\gamma \{x\}) \\
[\Lambda\alpha^\delta M]\gamma &= \Lambda\alpha^\delta.\{M\}\gamma & [\Lambda\alpha^\delta M]\gamma &= \Lambda\alpha^\delta.\{M\}\gamma \\
[\emptyset]\gamma &= \emptyset & [\emptyset]\gamma &= [] \\
[\ell = V; W]\gamma &= \ell = [V]\gamma; [W]\gamma & [\ell = V; W]\gamma &= \ell = [V]\gamma; [W]\gamma \\
[\ell V]^E\gamma &= (\ell \ ln V)^E \gamma & [\ell V]^E\gamma &= (\ell \ ln V)^E \gamma \\
[\kappa]\gamma &= \lambda x^\delta.\{\kappa\}(\text{return } x)
\end{align*}
\]

Figure 10: Mapping from Abstract Machine Configurations to Terms
and part of the evaluation context of \(M\) may appear in the continuation. Administrative reductions update \(C\) by growing the continuation, whilst maintaining the invariant that \(|C| = \text{Id}(M)|\). This process corresponds directly to traversing an evaluation context. It is straightforward to see that only a finite number of administrative reductions can occur consecutively as they either reduce the size of \(M\) or leave it unchanged and reduce the size of \(\kappa\) in the case of forwarding. Eventually the control part of the continuation will contain a redex corresponding to the active redex in \(M\) and it will then transition via a \(\beta\)-rule to a configuration \(C'\), such that \(|C'| = \text{Id}(N)|\).

The correspondence here is rather strong: there is a one-to-one mapping between \(\rightsquigarrow\) and \(\Rightarrow\). The inverse of the lemma is straightforward as the semantics is deterministic. Notice that Lemma 4.1 does not require that \(M\) be well-typed. We have chosen here not to perform type-erasure, but it is straightforward to adapt the results to semantics in which all polymorphism is erased and all type annotations are erased.

**Theorem 4.2** (Simulation). If \( M : A|^c E, \) and \( M \rightsquigarrow N, \) such that \( N\) is normal with respect to \(E\), then \( M \Rightarrow \gamma C, \) such that \(|C| = N|\).

**Proof.** By repeated application of Lemma 4.1.

5. Implementation

Our implementation of handlers is based on a mild syntactic extension to Links: the syntax is extended with \(\text{do}\) for invoking operations and \(\text{handle}(\cdot)\) \{\ldots\} for handling abstract computations.

**Syntactic Sugar** We provide syntactic sugar to make it more convenient to program with handlers. The function \(D\) is a source-to-source translation that elaborates the sugar. Figure 11 shows the cases for handlers only: \(D\) is a homomorphism on the other syntax constructors. Crucially, handlers are desugared into a function, that returns a thunk. This is important to ensure that handlers compose smoothly. For the same reason a parameterised handler desugars into a curried function, where the parameters precede the computation argument \(n\). The parameters are passed around by enclosing each operation clause by a function. Thus, the initial parameter values are applied directly to the \text{handle} expression.

**Backend** The Links interpreter is based on a CEK machine for ANF expressions. We have generalised this machine to support handlers based on the abstract machine of Section 4. The interpreter maintains a stack of handlers with first-in-last-out semantics, which makes it straightforward to implement effect forwarding. The invocation of an operation causes the interpreter to unwind the stack to find a suitable handler for the operation.

**Row Polymorphism** We have extended Links with support for user-defined operations, making use of the existing row type system. The current row type system is based on that of Rémy [33], adapted to support effect typing in a similar manner to the work of Leroy and Pessaux [19] and Blume et al. [4] on typing exceptions. Fields in a record can be absent, present at a particular type, or polymorphic in their presence. An earlier version of Links [22] was based on a slightly more refined variant of Rémy’s system, \(\Pi ML'\) [33], in which the type of a label is independent of its presence. This system was abandoned because it seemed somewhat counterintuitive for the purposes of record typing, but (as discussed in Section 2.5) it may offer advantages for effect typing.

**Subtyping and Row Typing** Subtyping is in fact a poor man’s row polymorphism. — Andreas Rossberg

Subtyping (or subeffecting) and row typing address similar concerns. However, they are not the same. In one dimension row polymorphism is more expressive than subtyping and in another dimension subtyping is more expressive than row polymorphism. Row polymorphism allows part of an effect to be named and reused in several places, which is essential for typing polymorphic functions such as \(\text{map}\). Row polymorphism can also be used whenever one might otherwise use subtyping in a first order manner. In terms of effects, this amounts to always keeping functions polymorphic in their effects such that the effect variable can be instantiated in order to simulate an upcast. On the other hand, subtyping applies at higher-order whereas row typing does not.

Links does not support subsumption (implicitly inferred subtyping), but does support subtyping via explicit upcasts. Anecdotally, such casts seem to be rarely needed in practice and can often be replaced by first-class polymorphism (another Links feature) instead.

**Shallow Handlers** We have not covered them in much detail here, but our implementation also supports shallow handlers [15]. These are indicated by using the \text{shallowhandler} keyword in place of \text{handler}. Whereas a deep handler performs a fold over a computation, a shallow handler merely performs a case-split. This means that one must explicitly invoke the handler each time the continuation is implied inside an operation clause. An advantage is that it makes it easy to switch to a different handler midway through a computation. A disadvantage is that shallow handlers are not much use without an external notion of recursion, and they are less easy to optimise than deep handlers [35]. The changes to the typing rules and operational semantics to accommodate shallow handlers are standard [15]. The modifications to the abstract machine are modest; the \text{M-OP-HANDLE} rule is adapted to drop the current handler and append the pure continuation on to its successor:

\[
\langle \text{do } \ell \ V \rangle^f \mid \gamma \ (\sigma, \chi) :: (\sigma', \chi') :: \kappa | (\kappa')_{\text{op}} \Rightarrow \\
\langle M \mid \gamma' \ (x \mapsto [V]) \gamma, \ k \mapsto (\kappa')^d \mid (\sigma + \sigma', \chi) :: \kappa \rangle
\]

6. Related Work

**Faking Row Polymorphism in Haskell** Haskell provides a rather rich type system that allows one to simulate many aspects of row polymorphism. Perhaps the biggest mismatch between rows and other typing features is that rows are inherently unordered, whereas other typing features are usually inherently ordered.

One approach is Swierstra’s \textit{data types a la carte} technique [34]. This amounts to encoding a row type as a sum, and then leveraging the type class system to automatically navigate through the sum type as if it were unordered. In practice, this encoding is a little fragile (e.g. sometimes additional type annotations are required), although recent improvements can make it somewhat more robust [25], particularly if one adds support for instance chains [26].

Another approach is to take advantage of the property that type class constraints are unordered. Early work on monad transformers [21] uses this idea to write modular abstract computations, as do Kammar et al. [15], Kiselyov et al. [17], and Kiselyov and Ishii [16] in their handler libraries. However, without some form of higher-order constraint solving (not supported by Haskell), one must still materialise ordered lists of effects when composing handlers. For many useful examples this is not a problem, but if we wish to build a list of handlers from disparate sources, then we must carefully ensure that their effects are composed in the same order.

Orchard and Petricek [28] make some progress towards encoding unordered effect rows by performing a sorting algorithm at the level of types, taking advantage of GHC’s support for dependently-typed programming [37]. However, this approach can fail in practice, as the type system cannot always infer that two types are equivalent in the presence of effect polymorphism.
Handler

\[
\text{handler } h[p] \equiv \text{handler}[m] h[p], \text{ where } m \text{ is fresh.}
\]

\[
D \left( \text{handler}[m] h[p] \{ \tau \} \right) = \begin{cases} 
\text{fun } h(m) \{ \ \text{handle}(m) \{ D (\tau) \} \} & \text{if } [p] = 0 \\
\text{fun } h[p](m) \{ \ \text{handle}(m) \{ D [p] (\tau) \} [p] \} & \text{otherwise}
\end{cases}
\]

Handler cases

\[
D_{[p]} (\tau) = D_{[p]} (c_1) \cdots D_{[p]} (c_n)
\]

\[
D_{[p]} (\text{case } q -> M) = \begin{cases} 
\text{case } q -> D (M) & \text{if } [p] = 0 \\
\text{case } q -> \text{fun}[p] \{ D (M) \} & \text{otherwise}
\end{cases}
\]

Figure 11: Desugaring Handlers

Implementations Any signature of abstract operations can be understood as a free algebra and represented as a functor. In particular, every such functor gives rise to a free monad. Thus, free monads provide a natural basis for implementing effect handlers. Many of the library implementations of effect handlers include implementations based on free monads [6, 15–17, 36].

Kammar et al. [15] provide an implementation of effect handlers using a continuation monad, which avoids materialising any data constructors. Wu and Schrijvers [35] explain how it works, by taking advantage of Haskell’s fusion optimisations. This approach does appear to depend rather critically on the handlers being deep rather than shallow, and in Haskell it relies on them being type classes, and hence not really first class.

The Idris effects library [6] takes advantage of dependent types to provide effect handlers for a form of effects corresponding to parameterised monads [1].

We are aware of three languages that are specifically designed with effect handlers in mind.

- Eff [3] is a strict language with Hindley-Milner type inference similar in spirit to ML, but extended with effect handlers. It includes a novel feature for supporting fresh generation of effects in order to support effects such as ML-style higher-order state (which has an operation for generating new references). The original version of Eff [3] does not include an effect type system. Later variants do include an effect type system [2, 31] which is considerably more complicated than ours. It makes essential use of subtyping, includes a region system, and a form of effect polymorphism which uses a form of row polymorphism.

- Frank [23] is a language with effect handlers but no separate notion of function: a function is but a special case of a handler. Frank has a bidirectional type system. It includes an effect type system and a novel form of effect polymorphism in which the programmer need never read or write effect variables. Frank’s effect system can be viewed as implementing a form of row polymorphism. Unlike Links, but much like Koka [18], Frank allows multiple occurrences of the same label in a row. In contrast rows in Links are based on Remy’s design in which duplicates are not allowed, but negative information is.

- Shonky [24] amounts to a dynamically-typed variant of Frank. Though it is not statically typed, handlers must be annotated with the names of the effects that they handle. The implementation of Shonky is quite similar to ours in that it uses a generalisation of the CEK machine. The main differences are that Shonky does not use an ANF representation, so has more forms of continuation to handle, and in contrast to our nested continuation structure, Shonky uses a completely flat structure.

Although OCaml itself has no support for effect handlers, a development branch, Multicore OCaml [9], does. Multicore OCaml does not include an effect type system, and handlers are restricted so that continuations can be invoked at most once. This design admits an efficient implementation, as continuations need never be copied, so they can simply be stored on the stack.

7. Conclusions and Future Work

We have implemented algebraic effects and handlers using row polymorphism and demonstrated that rows allow us to compose effectful computations seamlessly. We have formalised our system as the core calculus \( C^n eff \) and shown a correspondence between two semantics: a small-step operational semantics and an abstract machine semantics, the latter of which matches the implementation. We conclude by discussing ongoing and future work.

Effects are pervasive in modern web applications, so we would like to extend our implementation to the client backend of Links. The client backend already produces JavaScript in continuation-passing style in order to implement concurrency. We plan to extend this representation to support handlers.

The overhead incurred by the Links interpreter is significant [12, 14]. To improve performance, we are working on building a compiler backend with support for handlers [13]. One performance bottleneck we envisage is the need to support copying of continuations. It is well-known that one-shot continuations can be implemented efficiently [7]. Links already has a linear type system, which in the future we plan to take advantage of in order to track the linearity of handlers. During code generation we can specialise the run-time representation of handlers according to their linearity.

Links employs a message-passing concurrency model, similar to Erlang, but typed. Taking ideas from Multicore OCaml [9], we are investigating whether we can rebuild the Links concurrency implementation directly in terms of handlers.

Acknowledgments

Nicolas Oury originally suggested the Nim game as an example to demonstrate programming with handlers. Thanks to Simon Fowler, Ohad Kammar, Caoimhin Laoide-Kemp, Craig McLaughlin, Matthias R. Pedersen, and Gabriel Scherer for useful feedback, suggestions, and discussions. The first author was supported by EPSRC grant EP/L01503X/1 (The University of Edinburgh CDT in Pervasive Parallelism). The second author was supported by EPSRC grant EP/K034413/1 (A Basis for Concurrency and Distribution).

References


