

Appendix for “Calculating Correct Compilers”

PATRICK BAHR

Department of Computer Science, University of Copenhagen, Denmark

GRAHAM HUTTON

School of Computer Science, University of Nottingham, UK

A Global State

Recall the specification of the compiler:

$$\begin{aligned} \text{exec } (\text{comp}' x c) (s, q) = & \text{case eval } x \text{ q of} & (12) \\ & (\text{Just } n, q') \rightarrow \text{exec } c \text{ (VAL } n : s, q') \\ & (\text{Nothing}, q') \rightarrow \text{fail } (s, q') \end{aligned}$$

Below we give the full calculations for the *Add* and *Catch* cases.

$$\begin{aligned} & \text{exec } (\text{comp}' (\text{Add } x y) c) (s, q) \\ = & \{ \text{specification (12)} \} \\ & \text{case eval } x \text{ q of} \\ & (\text{Just } n, q') \rightarrow \text{case eval } y \text{ q' of} \\ & \quad (\text{Just } m, q'') \rightarrow \text{exec } c \text{ (VAL } (n + m) : s, q'') \\ & \quad (\text{Nothing}, q'') \rightarrow \text{fail } (s, q'') \\ & (\text{Nothing}, q') \rightarrow \text{fail } (s, q') \\ = & \{ \text{define: } \text{exec } (\text{ADD } c) \text{ (VAL } m : \text{VAL } n : s, q'') = \text{exec } c \text{ (VAL } (n + m) : s, q'') \} \\ & \text{case eval } x \text{ q of} \\ & (\text{Just } n, q') \rightarrow \text{case eval } y \text{ q' of} \\ & \quad (\text{Just } m, q'') \rightarrow \text{exec } (\text{ADD } c) \text{ (VAL } m : \text{VAL } n : s, q'') \\ & \quad (\text{Nothing}, q'') \rightarrow \text{fail } (s, q'') \\ & (\text{Nothing}, q') \rightarrow \text{fail } (s, q') \\ = & \{ \text{define: } \text{fail } (\text{VAL } n : s, q'') = \text{fail } (s, q'') \} \\ & \text{case eval } x \text{ q of} \\ & (\text{Just } n, q') \rightarrow \text{case eval } y \text{ q' of} \\ & \quad (\text{Just } m, q'') \rightarrow \text{exec } (\text{ADD } c) \text{ (VAL } m : \text{VAL } n : s, q'') \\ & \quad (\text{Nothing}, q'') \rightarrow \text{fail } (\text{VAL } n : s, q'') \\ & (\text{Nothing}, q') \rightarrow \text{fail } (s, q') \\ = & \{ \text{induction hypothesis for } y \} \\ & \text{case eval } x \text{ q of} \\ & (\text{Just } n, q') \rightarrow \text{exec } (\text{comp}' y (\text{ADD } c)) \text{ (VAL } n : s, q') \\ & (\text{Nothing}, q') \rightarrow \text{fail } (s, q') \\ = & \{ \text{induction hypothesis for } x \} \\ & \text{exec } (\text{comp}' x (\text{comp}' y (\text{ADD } c))) (s, q) \end{aligned}$$

2

Patrick Bahr and Graham Hutton

$$\begin{aligned}
& \text{exec } (\text{comp}' (\text{Catch } x \text{ h}) c) (s, q) \\
= & \{ \text{specification (12)} \} \\
& \text{case eval } x \text{ q of} \\
& \quad (\text{Just } n, q') \rightarrow \text{exec } c (\text{VAL } n : s, q') \\
& \quad (\text{Nothing}, q') \rightarrow \text{case eval } h \text{ q' of} \\
& \quad \quad (\text{Just } m, q'') \rightarrow \text{exec } c (\text{VAL } m : s, q'') \\
& \quad \quad (\text{Nothing}, q'') \rightarrow \text{fail } (s, q'') \\
= & \{ \text{induction hypothesis for } h \} \\
& \text{case eval } x \text{ q of} \\
& \quad (\text{Just } n, q') \rightarrow \text{exec } c (\text{VAL } n : s, q') \\
& \quad (\text{Nothing}, q') \rightarrow \text{exec } (\text{comp}' h c) (s, q') \\
= & \{ \text{define: fail } (\text{HAN } c' : s, q') = \text{exec } c' (s, q') \} \\
& \text{case eval } x \text{ q of} \\
& \quad (\text{Just } n, q') \rightarrow \text{exec } c (\text{VAL } n : s, q') \\
& \quad (\text{Nothing}, q') \rightarrow \text{fail } (\text{HAN } (\text{comp}' h c) : s, q') \\
= & \{ \text{define: exec } (\text{UNMARK } c) (\text{VAL } n : \text{HAN } _ : s, q') = \text{exec } c (\text{VAL } n : s, q') \} \\
& \text{case eval } x \text{ q of} \\
& \quad (\text{Just } n, q') \rightarrow \text{exec } (\text{UNMARK } c) (\text{VAL } n : \text{HAN } (\text{comp}' h c) : s, q') \\
& \quad (\text{Nothing}, q') \rightarrow \text{fail } (\text{HAN } (\text{comp}' h c) : s, q') \\
= & \{ \text{induction hypothesis for } x \} \\
& \text{exec } (\text{comp}' x (\text{UNMARK } c)) (\text{HAN } (\text{comp}' h c) : s, q) \\
= & \{ \text{define: exec } (\text{MARK } c'' c') (s, q) = \text{exec } c' (\text{HAN } c'' : s, q) \} \\
& \text{exec } (\text{MARK } (\text{comp}' h c) (\text{comp}' x (\text{UNMARK } c))) (s, q)
\end{aligned}$$

B Local State

We now consider the *local* approach to combining exceptions and state, in which the current state is discarded when an exception is thrown. This idea is reflected in the type for evaluation by moving the output state ‘inside’ the *Maybe* type:

$$\text{eval} :: \text{Expr} \rightarrow \text{State} \rightarrow \text{Maybe } (\text{Int}, \text{State})$$

That is, if evaluation succeeds then *eval* returns an integer value and a new state, and if an exception is thrown it returns *Nothing*. The definition for *eval* is similar to the previous section except there is now no state to propagate when evaluation fails, and in the case for *Catch* the handler uses the state from when the catch was entered:

$$\begin{aligned}
\text{eval } (\text{Val } n) \text{ q} &= \text{Just } (n, q) \\
\text{eval } (\text{Add } x \text{ y}) \text{ q} &= \text{case eval } x \text{ q of} \\
& \quad \text{Just } (n, q') \rightarrow \text{case eval } y \text{ q' of} \\
& \quad \quad \text{Just } (m, q'') \rightarrow \text{Just } (n + m, q'') \\
& \quad \quad \text{Nothing} \rightarrow \text{Nothing} \\
& \quad \text{Nothing} \rightarrow \text{Nothing} \\
\text{eval } \text{Throw } q &= \text{Nothing} \\
\text{eval } (\text{Catch } x \text{ h}) \text{ q} &= \text{case eval } x \text{ q of} \\
& \quad \text{Just } (n, q') \rightarrow \text{Just } (n, q')
\end{aligned}$$

$$\begin{aligned}
& \text{Nothing} && \rightarrow \text{eval } h \ q \\
\text{eval } \text{Get } q &= && \text{Just } (q, q) \\
\text{eval } (\text{Put } x \ y) \ q &= && \mathbf{case} \ \text{eval } x \ q \ \mathbf{of} \\
& \text{Just } (n, q') && \rightarrow \text{eval } y \ n \\
& \text{Nothing} && \rightarrow \text{Nothing}
\end{aligned}$$

For the purposes of the derivation of the compilation function $\text{comp}' :: \text{Expr} \rightarrow \text{Code} \rightarrow \text{Code}$ we use the same types as for the global state semantics:

```

exec :: Code → Conf → Conf
type Conf = (Stack, State)
type Stack = [Elem]
data Elem = VAL Int

```

The specification for the desired behaviour of comp' is essentially the same as for global state, except that when evaluation fails we no longer have an output state to consider and hence the function fail only takes a stack as argument:

$$\begin{aligned}
\text{exec } (\text{comp}' \ e \ c) \ (s, q) &= \mathbf{case} \ \text{eval } e \ q \ \mathbf{of} && (14) \\
& \text{Just } (n, q') && \rightarrow \text{exec } c \ (\text{VAL } n : s, q') \\
& \text{Nothing} && \rightarrow \text{fail } s
\end{aligned}$$

However, to ensure type correctness of the specification, fail must still return a configuration, i.e. $\text{fail} :: \text{Stack} \rightarrow \text{Conf}$. An alternative would be to supply the input state q as an argument to fail , which is a valid choice that would lead to a different compiler. We start the derivation for comp' with the cases for $\text{Val } n$, Throw and Get , which are easy:

$$\begin{aligned}
& \text{exec } (\text{comp}' \ (\text{Val } n) \ c) \ (s, q) \\
&= \{ \text{specification (14)} \} \\
& \text{exec } c \ (\text{VAL } n : s, q) \\
&= \{ \text{define: } \text{exec } (\text{PUSH } n \ c) \ (s, q) = \text{exec } c \ (\text{VAL } n : s, q) \} \\
& \text{exec } (\text{PUSH } n \ c) \ (s, q)
\end{aligned}$$

$$\begin{aligned}
& \text{exec } (\text{comp}' \ \text{Throw } c) \ (s, q) \\
&= \{ \text{specification (14)} \} \\
& \text{fail } s \\
&= \{ \text{define: } \text{exec } \text{FAIL} \ (s, q) = \text{fail } s \} \\
& \text{exec } \text{FAIL} \ (s, q)
\end{aligned}$$

$$\begin{aligned}
& \text{exec } (\text{comp}' \ \text{Get } c) \ (s, q) \\
&= \{ \text{specification (14)} \} \\
& \text{exec } c \ (\text{VAL } q : s, q) \\
&= \{ \text{define: } \text{exec } (\text{LOAD } c) \ (s, q) = \text{exec } c \ (\text{VAL } q : s, q) \} \\
& \text{exec } (\text{LOAD } c) \ (s, q)
\end{aligned}$$

The case for Add follows the now familiar pattern:

4

Patrick Bahr and Graham Hutton

$$\begin{aligned}
& \text{exec } (\text{comp}' (\text{Add } x \ y) \ c) \ (s, q) \\
= & \{ \text{specification (14)} \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{case eval } y \ q' \ \text{of} \\
& \quad \quad \text{Just } (m, q'') \rightarrow \text{exec } c \ (\text{VAL } (n + m) : s, q'') \\
& \quad \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{define: } \text{exec } (\text{ADD } c) \ (\text{VAL } m : \text{VAL } n : s, q'') = \text{exec } c \ (\text{VAL } (n + m) : s, q'') \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{case eval } y \ q' \ \text{of} \\
& \quad \quad \text{Just } (m, q'') \rightarrow \text{exec } (\text{ADD } c) \ (\text{VAL } m : \text{VAL } n : s, q'') \\
& \quad \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{define: } \text{fail } (\text{VAL } n : s) = \text{fail } s \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{case eval } y \ q' \ \text{of} \\
& \quad \quad \text{Just } (m, q'') \rightarrow \text{exec } (\text{ADD } c) \ (\text{VAL } m : \text{VAL } n : s, q'') \\
& \quad \quad \text{Nothing} \quad \rightarrow \text{fail } (\text{VAL } n : s) \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{induction hypothesis for } y \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } (\text{comp}' \ y \ (\text{ADD } c)) \ (\text{VAL } n : s, q') \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{induction hypothesis for } x \} \\
& \text{exec } (\text{comp}' \ x \ (\text{comp}' \ y \ (\text{ADD } c))) \ (s, q)
\end{aligned}$$

The case for *Catch* is more interesting this time. In the calculation for the global state semantics it was straightforward to bring the configuration arguments into the right form to apply the induction hypotheses. With local state, however, when an exception handler is invoked we require access to the state that was in place when the enclosing *Catch* was entered, which information we communicate via the stack:

$$\begin{aligned}
& \text{exec } (\text{comp}' (\text{Catch } x \ h) \ c) \ (s, q) \\
= & \{ \text{specification (14)} \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } c \ (\text{VAL } n : s, q') \\
& \quad \text{Nothing} \quad \rightarrow \text{case eval } h \ q \ \text{of} \\
& \quad \quad \text{Just } (m, q'') \rightarrow \text{exec } c \ (\text{VAL } m : s, q'') \\
& \quad \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{induction hypothesis for } h \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } c \ (\text{VAL } n : s, q') \\
& \quad \text{Nothing} \quad \rightarrow \text{exec } (\text{comp}' \ h \ c) \ (s, q) \\
= & \{ \text{define: } \text{fail } (\text{HAN } c' \ q : s) = \text{exec } c' \ (s, q) \} \\
& \text{case eval } x \ q \ \text{of} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } c \ (\text{VAL } n : s, q')
\end{aligned}$$

$$\begin{aligned}
& \text{Nothing} \quad \rightarrow \text{fail } (HAN \text{ (comp}' h c) q : s) \\
= & \{ \text{define: } \text{exec } (UNMARK c) (VAL n : HAN _ _ : s, q') = \text{exec } c (VAL n : s, q') \} \\
& \mathbf{case \textit{eval } x q \textit{ of}} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } (UNMARK c) (VAL n : HAN \text{ (comp}' h c) q : s, q') \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } (HAN \text{ (comp}' h c) q : s) \\
= & \{ \text{induction hypothesis for } x \} \\
& \text{exec } (comp' x (UNMARK c)) (HAN \text{ (comp}' h c) q : s, q) \\
= & \{ \text{define: } \text{exec } (MARK c'' c') (s, q) = \text{exec } c' (HAN c'' q : s, q) \} \\
& \text{exec } (MARK \text{ (comp}' h c) \text{ (comp}' x (UNMARK c))) (s, q)
\end{aligned}$$

Note that the new constructor HAN added to the $Elem$ type within this calculation now has two arguments: one for the handler code (as in previous calculations), and one for the state to be used if the handler is invoked (for local state). We conclude the calculation with the case for Put , which proceeds in the same manner as for global state:

$$\begin{aligned}
& \text{exec } (comp' (Put x y) c) (s, q) \\
= & \{ \text{specification (14)} \} \\
& \mathbf{case \textit{eval } x q \textit{ of}} \\
& \quad \text{Just } (n, q') \rightarrow \mathbf{case \textit{eval } y n \textit{ of}} \\
& \qquad \quad \text{Just } (m, q'') \rightarrow \text{exec } c (VAL m : s, q'') \\
& \qquad \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{induction hypothesis for } y \} \\
& \mathbf{case \textit{eval } x q \textit{ of}} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } (comp' y c) (s, n) \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{define: } \text{exec } (SAVE c') (VAL n : s, q') = \text{exec } c' (s, n) \} \\
& \mathbf{case \textit{eval } x q \textit{ of}} \\
& \quad \text{Just } (n, q') \rightarrow \text{exec } (SAVE \text{ (comp}' y c)) (VAL n : s, q') \\
& \quad \text{Nothing} \quad \rightarrow \text{fail } s \\
= & \{ \text{induction hypothesis for } x \} \\
& \text{exec } (comp' x (SAVE \text{ (comp}' y c))) (s, q)
\end{aligned}$$

In summary, collecting together everything that we have learned in the process of the above calculations, we obtained the following definitions.

Target language:

data $Code = HALT \mid PUSH \ Int \ Code \mid ADD \ Code \mid$
 $FAIL \mid MARK \ Code \ Code \mid UNMARK \ Code \mid$
 $LOAD \ Code \mid SAVE \ Code$

Compiler:

$comp \quad \quad \quad \quad :: Expr \rightarrow Code$
 $comp \ x \quad \quad \quad = comp' \ x \ HALT$
 $comp' \quad \quad \quad \quad :: Expr \rightarrow Code \rightarrow Code$
 $comp' \ (Val \ n) \ c \quad = PUSH \ n \ c$

6

Patrick Bahr and Graham Hutton

$$\begin{aligned}
\mathit{comp}' (\mathit{Add} \ x \ y) \ c &= \mathit{comp}' \ x (\mathit{comp}' \ y (\mathit{ADD} \ c)) \\
\mathit{comp}' \ \mathit{Throw} \ c &= \mathit{FAIL} \\
\mathit{comp}' (\mathit{Catch} \ x \ h) \ c &= \mathit{MARK} (\mathit{comp}' \ h \ c) (\mathit{comp}' \ x (\mathit{UNMARK} \ c)) \\
\mathit{comp}' \ \mathit{Get} \ c &= \mathit{LOAD} \ c \\
\mathit{comp}' (\mathit{Put} \ x \ y) \ c &= \mathit{comp}' \ x (\mathit{SAVE} (\mathit{comp}' \ y \ c))
\end{aligned}$$

Virtual machine:

$$\begin{aligned}
\mathbf{data} \ \mathit{Elem} &= \mathit{VAL} \ \mathit{Int} \mid \mathit{HAN} \ \mathit{Code} \ \mathit{State} \\
\mathit{exec} &:: \mathit{Code} \rightarrow \mathit{Conf} \rightarrow \mathit{Conf} \\
\mathit{exec} \ \mathit{HALT} \ (s, q) &= (s, q) \\
\mathit{exec} \ (\mathit{PUSH} \ n \ c) \ (s, q) &= \mathit{exec} \ c \ (\mathit{VAL} \ n : s, q) \\
\mathit{exec} \ (\mathit{ADD} \ c) \ (\mathit{VAL} \ m : \mathit{VAL} \ n : s, q) &= \mathit{exec} \ c \ (\mathit{VAL} \ (n + m) : s, q) \\
\mathit{exec} \ \mathit{FAIL} \ (s, q) &= \mathit{fail} \ s \\
\mathit{exec} \ (\mathit{MARK} \ h \ c) \ (s, q) &= \mathit{exec} \ c \ (\mathit{HAN} \ h \ q : s, q) \\
\mathit{exec} \ (\mathit{UNMARK} \ c) \ (\mathit{VAL} \ n : \mathit{HAN} \ _ _ : s, q) &= \mathit{exec} \ c \ (\mathit{VAL} \ n : s, q) \\
\mathit{exec} \ (\mathit{LOAD} \ c) \ (s, q) &= \mathit{exec} \ c \ (\mathit{VAL} \ q : s, q) \\
\mathit{exec} \ (\mathit{SAVE} \ c) \ (\mathit{VAL} \ n : s, q) &= \mathit{exec} \ c \ (s, n) \\
\mathit{fail} &:: \mathit{Stack} \rightarrow \mathit{Conf} \\
\mathit{fail} \ [] &= ([], 0) \\
\mathit{fail} \ (\mathit{VAL} \ n : s) &= \mathit{fail} \ s \\
\mathit{fail} \ (\mathit{HAN} \ h \ q : s) &= \mathit{exec} \ h \ (s, q)
\end{aligned}$$

Note that, as previously, we added an equation to *fail* for the case when the stack is empty in order to make the definition complete. Because *fail* does not take a state as an argument, we can only give a fixed output state as the result, for which purposes we simply return the value 0. As before, the choice for this additional equation has no impact on the correctness of the above calculations because they do not depend on this choice.