Proof-Producing Synthesis of CakeML with I/O and Local State from Monadic HOL Functions

Son Ho[1], Oskar Abrahamsson[2], Ramana Kumar[3], Magnus O. Myreen[2], Yong Kiam Tan[4], and Michael Norrish[5]

[1] MINES ParisTech, PSL Research University, France
[2] Chalmers University of Technology, Sweden
[3] Data61, CSIRO / UNSW, Australia
[4] Carnegie Mellon University, USA
[5] Data61, CSIRO / ANU, Australia

Abstract. We introduce an automatic method for producing stateful ML programs together with proofs of correctness from monadic functions in HOL. Our mechanism supports references, exceptions, and I/O operations, and can generate functions manipulating local state, which can then be encapsulated for use in a pure context. We apply this approach to several non-trivial examples, including the type inferencer and register allocator of the otherwise pure CakeML compiler, which now benefits from better runtime performance. This development has been carried out in the HOL4 theorem prover.

1 Introduction

This paper is about bridging the gap between programs verified in logic and verified implementations of those programs in a programming language (and ultimately machine code). As a toy example, consider computing the nth Fibonacci number. Here is a recursion equation for a function, fib, in higher-order logic (HOL) that does the job.

$$\mathsf{fib}\ n = \mathsf{if}\ n < 2\ \mathsf{then}\ n\ \mathsf{else}\ \mathsf{fib}\ (n-1) + \mathsf{fib}\ (n-2)$$

A hand-written implementation (shown here in CakeML [9], which has similar syntax and semantics to Standard ML) would look something like this:

```
fun fiba i j n = if n = 0 then i else fiba j (i+j) (n-1);
(print (n2s (fiba 0 1 (s2n (hd (CommandLine.arguments())))));
 print "\n")
handle _ => print_err ("usage: " ^ CommandLine.name() ^ " <n>\n");
```

In moving from mathematics to a real implementation, some issues are apparent:

(1) We use a tail-recursive linear-time algorithm, rather than the exponential-time recursion equation.
(2) The whole program is not a pure function: it does I/O, reading its argument from the command line and printing the answer to standard output.

(3) We use exception handling to deal with malformed inputs (if the arguments do not start with a string representing a natural number, `hd` or `s2n` may raise an exception).

The first of these issues (1) can easily be handled in the realm of logical functions: We define the tail-recursive version in logic

$$\mathsf{fiba}\ i\ j\ n = \mathsf{if}\ n = 0\ \mathsf{then}\ i\ \mathsf{else}\ \mathsf{fiba}\ j\ (i + j)\ (n - 1)$$

then produce a correctness theorem, $\vdash\ \forall n.\ \mathsf{fiba}\ 0\ 1\ n = \mathsf{fib}\ n$, with a simple inductive proof (a 5-line tactic proof in HOL4, not shown).

Now, because fiba is a logical function with an obvious computational counterpart, we can use proof-producing synthesis techniques [13] to automatically synthesise code verified to compute it. We thereby produce something like the first line of the CakeML code above, along with a theorem relating the semantics of the synthesised code back to the function in logic.

But when it comes to handling the other two issues, (2) and (3), and producing and verifying the remaining three lines of CakeML code, our options are less straightforward. The first issue was easy because we were working with a *shallow embedding*, where one writes the program as a function in logic and proves properties about that function directly. Shallow embeddings rely on an analogy between mathematical functions and procedures in a pure functional programming language. Effects, however, like state, I/O, and exceptions, can stretch this analogy too far. The alternative is a *deep embedding*: one writes the program as an input to a formal semantics, which can accurately model computational effects, and proves properties about its execution under those semantics.

Proofs about shallow embeddings are relatively easy since they are in the native language of the theorem prover, whereas proofs about deep embeddings are filled with tedious details because of the indirection through an explicit semantics. Still, the explicit semantics make deep embeddings more realistic. An intermediate option that is suitable for the effects we are interested in — state/references, exceptions, and I/O — is to use *monadic functions*: one writes (shallow) functions that represent computations, aided by a composition operator (monadic bind) for stitching together effects. The monadic approach to writing effectful code in a pure language may be familiar from the Haskell language which made it popular.

For our nth Fibonacci example, we can model the effects of the whole program with a monadic function, fibm, that calls the pure function fiba to do the calculation. Figure 1 shows how fibm can be written using do-notation familiar from Haskell. This is as close as we can get to capturing the effectful behaviour of the desired CakeML program while remaining in a shallow embedding. Now how can we produce real code along with a proof that it has the correct semantics? If we use the proof-producing synthesis techniques mentioned above [13], we produce *pure* CakeML code that exposes the monadic plumbing in an explicit state-passing style. But we would prefer verified *effectful* code that uses native features of the target language (CakeML) to implement the monadic effects.

```
fibm () =
do
  args ← commandline (arguments ());
  a ← hd args;
  n ← s2n a;
  stdio (print (n2s (fiba 0 1 n)));
  stdio (print "\n")
od otherwise
do
  name ← commandline (name ());
  stdio (print_err ("usage: " ^ name ^ " <n>\n"))
od
```

Fig. 1. The Fibonacci program written using do-notation in logic.

In this paper, we present an automated technique for producing verified effectful code that handles I/O, exceptions, and other issues arising in the move from mathematics to real implementations. Our technique systematically establishes a connection between shallowly embedded functions in HOL with monadic effects and deeply embedded programs in the impure functional language CakeML. The synthesised code is efficient insofar as it uses the native effects of the target language and is close to what a real implementer would write. For example, given the monadic fibm function above, our technique produces essentially the same CakeML program as on the first page (but with a `let` for every monad bind), together with a proof that the synthesised program is a refinement.

Contributions Our technique for producing verified effectful code from monadic functions builds on a previous limited approach [13]. The new generalised method adds support for the following features:

- global references and exceptions (as before, but generalised),
- mutable arrays (both fixed and variable size),
- input/output (I/O) effects,
- local mutable arrays and references, which can be integrated seamlessly with code synthesis for otherwise pure functions, and,
- composable effects, whereby different state and exception monads can be combined using a lifting operator.

As a result, we can now write *whole programs* as shallow embeddings and obtain real verified code via synthesis. Prior to this paper, whole program verification in CakeML involved manual deep embedding proofs for (at the very least) the I/O wrapper. To exercise our toolchain, we apply it to several examples:

- the nth Fibonacci example already seen (exceptions, I/O)
- the Floyd Warshall algorithm for finding shortest paths (arrays)
- the CakeML compiler's type inferencer (local refs, exceptions)
- the CakeML compiler's register allocator (local refs, arrays)

– the Candle theorem prover's kernel [8] (global refs, exceptions)
– an OpenTheory [7] article checker (global refs, exceptions, I/O)

In §5, we compare runtimes with the previous non-stateful versions of CakeML's register allocator and type inferencer; and for the OpenTheory reader we compare the amount of code/proof required before and after using our technique.

The HOL4 development is at `https://code.cakeml.org`; our new synthesis tool is at `https://code.cakeml.org/tree/master/translator/monadic`.

2 High-level ideas

This paper combines the following three concepts in order to deliver the contributions listed above. The main ideas will be described briefly in this section, while subsequent sections will provide details. The three concepts are:

(i) synthesis of stateful ML code as described in our previous work [13],
(ii) separation logic [15] as used by characteristic formulae for CakeML [5], and
(iii) a new abstract synthesis mode for the CakeML synthesis tools [13].

Our previous work on proof-producing synthesis of stateful ML (i) was severely limited by the requirement to have a hard-coded invariant on the program's state. There was no support for I/O and all references had to be declared globally. At the time of developing (i), we did not have a satisfactory way of generalising the hard-coded state invariant.

In this paper we show (in §3) that the separation logic of CF (ii) can be used to neatly generalise the hard-coded state invariant of our prior work (i). CF-style separation logic easily supports references and arrays, including resizable arrays, and, supports I/O too because it allows us to treat I/O components as if they are heap components. Furthermore, by carefully designing the integration of (i) and (ii), we retain the frame rule from the separation logic. In the context of code synthesis, this frame rule allows us to implement a lifting feature for changing the type of the state-and-exception monads. Being able to change types in the monads allows us to develop *reusable* libraries — e.g. verified file I/O functions — that users can lift into the monad that is appropriate for their application.

The combination of (i) and (ii) does not by itself support synthesis of code with local state due to inherited limitations of (i), wherein the generated code must be produced as a concrete list of global declarations. For example, if monadic functions, say `foo` and `bar`, refer to a common reference, say `r`, the reference `r` must be defined globally:

```
val r = ref 0;
fun foo n = ...; (* code that uses r *)
fun bar n = ...; (* code that uses r and calls foo *)
```

In this paper (in §4), we introduce a new *abstract* synthesis mode (iii) which removes the requirement of generating code that only consists of a list of global declarations, and, as a result, we are now able to synthesise code such as the following, where reference `r` is a local variable.

```
fun pure_bar k n =
  let
    val r = ref k
    fun foo n = ... (* code that uses r *)
    fun bar n = ... (* code that uses r and calls foo *)
  in Success (bar n) end
  handle e => Failure e;
```

In the input to the synthesis tool, this declaration and initialisation of local state corresponds to applying the state-and-exception monad. Expressions that fully apply the state-and-exception monad can subsequently be used in the synthesis of *pure* CakeML code: the monadic synthesis tools can prove a pure specification for such expressions, thereby encapsulating the monadic features.

3 Generalised approach to synthesis of stateful ML code

This section describes how our previous approach to proof-producing synthesis of stateful ML code [13] has been generalised. In particular, we explain how the separation logic from our previous work on characteristic formulae [5] has been used for the generalisation (§3.3); and how this new approach adds support for user-defined references, fixed- and variable-length arrays, I/O functions (§3.4), and a handy feature for reusing state-and-exception monads (§3.5).

In order to make this paper as self-contained as possible, we start with a brief look at how the semantics of CakeML is defined (§3.1) and how our previous work on synthesis of pure CakeML code works (§3.2), since the new synthesis method for stateful code is an evolution of the original approach for pure code.

3.1 Preliminaries: CakeML semantics

The semantics of the CakeML language is defined in the *functional big-step* style [14], which means that the semantics is an interpreter defined as a functional program in the logic of a theorem prover.

The definition of the semantics is layered. At the top-level the semantics function defines what the observable I/O events are for a given whole program. However, more relevant to the presentation in this paper is the next layer down: a function called evaluate that describes exactly how expressions evaluate. The type of the evaluate function is shown below. This function takes as arguments a state (with a type variable for the I/O environment), a value environment, and a list of expressions to evaluate. It returns a new state and a value result.

$$\begin{aligned}&\mathsf{evaluate} : \delta\ \mathtt{state} \rightarrow \\ &\qquad \mathtt{v\ sem\_env} \rightarrow \mathtt{exp\ list} \rightarrow \delta\ \mathtt{state} \times (\mathtt{v\ list}, \mathtt{v})\ \mathtt{result}\end{aligned}$$

The semantics state is defined as the record type below. The fields relevant for this presentation are: refs, clock and ffi. The refs field is a list of store values that acts as a mapping from reference names (list index) to reference and array

values (list element). The clock is a logical clock for the functional big-step style. The clock allows us to prove termination of evaluate and is, at the same time, used for reasoning about divergence. Lastly, ffi is the parametrised oracle model of the foreign function interface, i.e. I/O environment.

$$\delta\ \texttt{state} = \texttt{<|}\ \mathsf{clock} : \texttt{num}\ ;\ \mathsf{refs} : \texttt{store\_v list}\ ;\ \mathsf{ffi} : \delta\ \texttt{ffi\_state}\ ;\ \ldots\ \texttt{|>}$$

$$\text{where}\ \ \texttt{store\_v} = \mathsf{Refv\ v} \mid \mathsf{W8array}\ (\texttt{word8 list}) \mid \mathsf{Varray}\ (\mathsf{v}\ \texttt{list})$$

A call to the function evaluate returns one of two results: Rval *res* for successfully terminating computations, and Rerr *err* for stuck computations.

Successful computations, Rval *res*, return a list *res* of CakeML values. CakeML values are modelled in the semantics using a datatype called v. This datatype includes (among other things) constructors for (mutually recursive) closures (Closure and Recclosure), datatype constructor values (Conv), and literal values (Litv) such as integers, strings, characters etc. These will be explained when needed in the rest of the paper.

Stuck computations, Rerr *err*, carry an error value *err* that is one of the following. For this paper, Rraise *exc* is the most relevant case.

- Rraise *exc* indicates that evaluation results in an uncaught exception *exc*. These exceptions can be caught with a `handle` in CakeML.
- Rabort Rtimeout_error indicates that evaluation of the expression consumes all of the logical clock. Programs that hit this error for all initial values of the clock are considered diverging.
- Rabort Rtype_error, for other kinds of errors, e.g. when evaluating ill-typed expressions, or attempting to access unbound variables.

3.2 Preliminaries: Synthesis of pure ML code

Our previous work [13] describes a *proof-producing* algorithm for synthesising CakeML functions from functions in higher-order logic. Here proof-producing means that each execution proves a theorem (called a certificate theorem) guaranteeing correctness of that execution of the algorithm. In our setting, these theorems relate the CakeML semantics of the synthesised code with the given HOL function.

The whole approach is centred around a systematic way of proving theorems relating HOL functions (i.e. HOL terms) with CakeML expressions. In order for us to state relations between HOL terms and CakeML expressions, we need a way to state relations between HOL terms and CakeML values. For this we use relations (int, list _, _ ⟶ _, etc.) which we call refinement invariants. The definition of the simple int refinement invariant is shown below: int i v is true if CakeML value v of type v represents the HOL integer i of type `int`.

$$\mathsf{int}\ i = (\lambda v.\ v = \mathsf{Litv}\ (\mathsf{IntLit}\ i))$$

Most refinement invariants are more complicated, e.g. list (list int) xs v states that CakeML value v represents lists of int lists xs of HOL type `int list list`.

We now turn to CakeML expressions: we define a predicate called Eval in order to conveniently state relationships between HOL terms and CakeML expressions. The intuition is that Eval *env exp P* is true if *exp* evaluates (in environment *env*) to some result *res* (of HOL type v) such that *P* holds for *res*, i.e. *P res*. The formal definition below is cluttered by details regarding the clock and references: there must be a large enough clock and *exp* may allocate new references, *refs′*, but must not modify any existing references, *refs*. We express this restriction on the references using list append ++. Note that any list index that can be looked up in *refs* has the same look up in *refs* ++ *refs′*.

$$
\begin{array}{l}
\mathsf{Eval}\ env\ exp\ P \iff \\
\forall\, refs. \\
\quad \exists\, res\ refs'\ ck. \\
\qquad (\mathsf{evaluate}\ (\mathsf{empty}\ with\ \texttt{<|}\,\mathsf{refs}\ :=\ refs;\ \mathsf{clock}\ :=\ ck\,\texttt{|>})\ env\ [exp] = \\
\qquad\ (\mathsf{empty}\ with\ \mathsf{refs}\ :=\ refs +\!\!+ refs',\mathsf{Rval}\ [res])) \wedge P\ res
\end{array}
$$

The use of Eval and the main idea behind the synthesis algorithm is most conveniently described using an example. The example we consider here is the following HOL function:

$$\mathsf{add1} = \lambda\, x.\ x + 1$$

The main part of the synthesis algorithm proceeds as a syntactic bottom-up pass over the given HOL term. In this case, the bottom-up pass traverses HOL term $\lambda\, x.\ x + 1$. The result of each stage of the pass is a theorem stated in terms of Eval in the format shown below. Such theorems state a connection between a HOL term *t* and some generated *code* w.r.t. a refinement invariant *ref_inv* that is appropriate for the type of *t*.

$$\text{general format:} \quad \mathit{assumptions} \Rightarrow \mathsf{Eval}\ env\ code\ (\mathit{ref\_inv}\ t)$$

For our little example, the algorithm derives the following theorems for the subterms *x* and 1, which are the leaves of the HOL term. Here and elsewhere in this paper, we display CakeML abstract syntax as concrete syntax inside $\lfloor \cdots \rfloor$, i.e. $\lfloor \texttt{1} \rfloor$ is actually the CakeML expression Lit (IntLit 1) in the theorem prover HOL4; similarly $\lfloor \texttt{x} \rfloor$ is actually displayed as Var (Short "x") in HOL4. Note that both theorems below are of the required form.

$$
\begin{array}{l}
\vdash \mathsf{T} \Rightarrow \mathsf{Eval}\ env\ \lfloor \texttt{1} \rfloor\ (\mathsf{int}\ 1) \\
\vdash \mathsf{Eval}\ env\ \lfloor \texttt{x} \rfloor\ (\mathsf{int}\ x) \Rightarrow \mathsf{Eval}\ env\ \lfloor \texttt{x} \rfloor\ (\mathsf{int}\ x)
\end{array}
\tag{1}
$$

The algorithm uses theorems (1) when proving a theorem for the compound expression $x + 1$. The process is aided by an auxiliary lemma for integer addition, shown below. The synthesis algorithm is supported by several such pre-proved lemmas for various common operations.

$$
\begin{array}{l}
\vdash \mathsf{Eval}\ env\ x_1\ (\mathsf{int}\ n_1) \Rightarrow \\
\quad \mathsf{Eval}\ env\ x_2\ (\mathsf{int}\ n_2) \Rightarrow \\
\quad \mathsf{Eval}\ env\ \lfloor x_1 \texttt{ + } x_2 \rfloor\ (\mathsf{int}\ (n_1 + n_2))
\end{array}
$$

By choosing the right specialisations for the variables, x_1, x_2, n_1, n_2, the algorithm derives the following theorem for the body of the running example. Here the assumption on evaluation of $\lfloor \texttt{x} \rfloor$ was inherited from (1).

$$\vdash \mathsf{Eval}\ env\ \lfloor \texttt{x} \rfloor\ (\mathsf{int}\ x) \Rightarrow \mathsf{Eval}\ env\ \lfloor \texttt{x + 1} \rfloor\ (\mathsf{int}\ (x+1)) \tag{2}$$

Next, the algorithm needs to introduce the λ-binder in $\lambda\, x.\ x + 1$. This can be done by instantiation of the following pre-proved lemma. Note that the lemma below introduces a refinement invariant for function types, $\longrightarrow$, which combines refinement invariants for the input and output types of the function [13].

$$\begin{array}{l} \vdash (\forall\, v\ x.\ a\ x\ v \Rightarrow \mathsf{Eval}\ (env\,[n \mapsto v])\ body\ (b\ (f\ x))) \Rightarrow \\ \quad \mathsf{Eval}\ env\ \lfloor \texttt{fn n => } body \rfloor\ ((a \longrightarrow b)\, f) \end{array}$$

An appropriate instantiation and combination with (2) produces the following:

$$\vdash \mathsf{T} \Rightarrow \mathsf{Eval}\ env\ \lfloor \texttt{fn x => x + 1} \rfloor\ ((\mathsf{int} \longrightarrow \mathsf{int})\ (\lambda\, x.\ x+1))$$

which, after only minor reformulation, becomes a certificate theorem for the given HOL function add1:

$$\vdash \mathsf{Eval}\ env\ \lfloor \texttt{fn x => x + 1} \rfloor\ ((\mathsf{int} \longrightarrow \mathsf{int})\ \mathsf{add1})$$

Additional notes. The main part of the synthesis algorithm is always a bottom-up traversal as described above. However, synthesis of recursive functions requires an additional post-processing phase which involves an automatic induction proof. We omit a description of such induction proofs since the solution described previously in [13] is not important for understanding this paper, and works in essentially the same way for synthesis of recursive stateful functions.

3.3 Synthesis of stateful ML code

Our algorithm for synthesis of stateful ML is very similar to the algorithm described above for synthesis of pure CakeML code. The main differences are:

- the input HOL terms must be written in a state-and-exception monad, and
- instead of Eval and $\longrightarrow$, the derived theorems use EvalM and $\longrightarrow^M$,

where EvalM and $\longrightarrow^M$ relate the monad's state to the references and foreign function interface of the underlying CakeML state (fields refs and ffi). These concepts will be described below.

Generic state-and-exception monad. The new generalised synthesis work-flow uses the following state-and-exception monad $(\alpha,\ \beta,\ \gamma)\ \texttt{M}$, where α is the state type, β is the return type, and γ is the exception type.

$$\begin{array}{l} (\alpha,\ \beta,\ \gamma)\ \texttt{M} \ = \ \alpha \to (\beta,\ \gamma)\ \texttt{exc} \times \alpha \\ \text{where}\ \ (\beta,\ \gamma)\ \texttt{exc} \ = \ \mathsf{Success}\ \beta \mid \mathsf{Failure}\ \gamma \end{array}$$

We define the following interface for this monad type. Note that syntactic sugar is often used: in our case, we write do $n \leftarrow$ *foo*; return (*bar* n) od (as was done in §1) when we mean bind *foo* ($\lambda\, n$. return (*bar* n)).

return $x = (\lambda\, s.\ (\mathsf{Success}\ x,s))$

bind $x\ f =$
$(\lambda\, s.\ \mathsf{case}\ x\ s\ \mathsf{of}\ (\mathsf{Success}\ y,s)\ \Rightarrow\ f\ y\ s \mid (\mathsf{Failure}\ x,s)\ \Rightarrow\ (\mathsf{Failure}\ x,s))$

x otherwise $y =$
$(\lambda\, s.\ \mathsf{case}\ x\ s\ \mathsf{of}\ (\mathsf{Success}\ v,s)\ \Rightarrow\ (\mathsf{Success}\ v,s) \mid (\mathsf{Failure}\ e,s)\ \Rightarrow\ y\ s)$

Functions that update the content of state can only be defined once the state type is instantiated. A function for changing a monad `M` to have a different state type is introduced in §3.5.

Definitions and lemmas for synthesis. We define EvalM as follows. A CakeML source expression *exp* is considered to satisfy an execution relation P if for any CakeML state s, which is related by state_rel to the state monad state *st* and state assertion H, the CakeML expression *exp* evaluates to a result *res* such that the relation P accepts the transition and state_rel_frame holds for state assertion H. The auxiliary functions state_rel and state_rel_frame will be described below. The first argument *ro* can be used to restrict effects to *references only*, as described a few paragraphs further down.

$$
\begin{array}{l}
\mathsf{EvalM}\ ro\ env\ st\ exp\ P\ H\ \Longleftrightarrow \\
\forall\, s. \\
\quad \mathsf{state\_rel}\ H\ st\ s\ \Rightarrow \\
\quad \exists\, s_2\ res\ st_2\ ck. \\
\qquad (\mathsf{evaluate}\ (s\ with\ \mathsf{clock}\ :=\ ck)\ env\ [exp] = (s_2,res)) \wedge \\
\qquad P\ st\ (st_2,res) \wedge \mathsf{state\_rel\_frame}\ ro\ H\ (st,s)\ (st_2,s_2)
\end{array}
$$

In the definition above, state_rel and state_rel_frame are used to check that the user-specified state assertion H relates the CakeML states and the monad states. Furthermore, state_rel_frame ensures that the separation logic frame rule is true. Both use the separation logic set-up from our previous work on characteristic formulae for CakeML [5], where we define a function st2heap which, given a projection p and CakeML state s, turns the CakeML state into a set representation of the reference store and foreign-function interface (used for I/O).

The H in the definition above is a pair (h,p) containing a heap assertion h and the projection p. We define state_rel (h,p) *st* s to state that the heap assertion produced by applying h to the current monad state *st* must be true for some subset produced by st2heap when applied to the CakeML state s. Here $(*)$ is the separating conjunction and T is true for any heap.

$$\mathsf{state\_rel}\ (h,p)\ st\ s\ \Longleftrightarrow\ (h\ st * \mathsf{T})\ (\mathsf{st2heap}\ p\ s)$$

The relation state_rel_frame states: any frame F that is true separately from $h\ st_1$ for the initial state is also true for the final state; and if the references-only *ro* configuration is set, then the only difference in the states must be in

the references and clock, i.e. no I/O operations are permitted. The *ro* flag is instantiated to true when a pure specification (Eval) is proved for local state §4.

$$\begin{array}{l}\mathsf{state\_rel\_frame}\ ro\ (h,p)\ (st_1,s_1)\ (st_2,s_2)\iff\\(ro \Rightarrow \exists\, refs.\ s_2 = s_1\ with\ \mathsf{refs}\ :=\ refs) \wedge\\ \forall\, F.\ (h\ st_1 * F)\ (\mathsf{st2heap}\ p\ s_1) \Rightarrow (h\ st_2 * F * \mathsf{T})\ (\mathsf{st2heap}\ p\ s_2)\end{array}$$

We prove lemmas to aid the synthesis algorithm in construction of proofs. The lemmas shown in this paper use the following definition of monad.

$$\begin{array}{l}\mathsf{monad}\ a\ b\ x\ st_1\ (st_2,res)\iff\\ \mathsf{case}\ (x\ st_1,res)\ \mathsf{of}\\ \quad((\mathsf{Success}\ y,st),\mathsf{Rval}\ [v])\ \Rightarrow\ (st = st_2) \wedge a\ y\ v\\ |\ ((\mathsf{Failure}\ e,st),\mathsf{Rerr}\ (\mathsf{Rraise}\ v'))\ \Rightarrow\ (st = st_2) \wedge b\ e\ v'\\ |\ \_\ \Rightarrow\ \mathsf{F}\end{array}$$

Synthesis makes use of the following two lemmas in proofs involving monadic return and bind. For return x, synthesis proves an Eval-theorem for x. For bind, it proves a theorem that fits the shape of the first four lines of the lemma and returns a theorem consisting of the last two lines, appropriately instantiated.

$$\vdash \mathsf{Eval}\ env\ exp\ (a\ x) \Rightarrow \mathsf{EvalM}\ ro\ env\ st\ exp\ (\mathsf{monad}\ a\ b\ (\mathsf{return}\ x))\ H$$

$$\begin{array}{l}\vdash ((assums_1 \Rightarrow \mathsf{EvalM}\ ro\ env\ st\ e_1\ (\mathsf{monad}\ b\ c\ x)\ H) \wedge\\ \quad\forall\, z\ v.\\ \quad\quad b\ z\ v \wedge assums_2\ z \Rightarrow\\ \quad\quad \mathsf{EvalM}\ ro\ (env\,[n \mapsto v])\ (\mathsf{snd}\ (x\ st))\ e_2\ (\mathsf{monad}\ a\ c\ (f\ z))\ H) \Rightarrow\\ \quad assums_1 \wedge (\forall\, z.\ (\mathsf{fst}\ (x\ st) = \mathsf{Success}\ z) \Rightarrow assums_2\ z) \Rightarrow\\ \quad \mathsf{EvalM}\ ro\ env\ st\ \lfloor \texttt{let n = } e_1\ \texttt{in}\ e_2 \rfloor\ (\mathsf{monad}\ a\ c\ (\mathsf{bind}\ x\ f))\ H\end{array}$$

3.4 References, Arrays and I/O

The synthesis algorithm uses specialised lemmas when the generic state-and-exception monad has been instantiated. Consider the following instantiation of the monad's state type to a record type. The programmer's intention is that the lists are to be synthesised to arrays in CakeML and the I/O component `IO_fs` is a model of a file system (taken from a library).

```
example_state =
 <| ref1 : int; farray1 : int list; rarray1 : int list; stdio : IO_fs |>
```

With the help of getter- and setter-functions and library functions for file I/O, users can conveniently write monadic functions that operate over this state type.

When it comes to synthesis, the automation instantiates H with an appropriate heap assertion, in this instance: ASSERT. The user has informed the synthesis tool that farray1 is to be a fixed-size array and rarray1 is to be a resizable-size array. A resizable-array is implemented as a reference that contains an array, since CakeML (like SML) does not directly support resizing arrays. Below,

REF_REL int ref1_loc st.ref1 asserts that int relates the value held in a reference at a fixed store location ref1_loc to the integer in st.ref1. Similarly, ARRAY_REL and RARRAY_REL specify a connection for the array fields. Lastly, STDIO is a heap assertion for the file I/O taken from a library.

ASSERT st =
REF_REL int ref1_loc st.ref1 * RARRAY_REL int rarray1_loc st.rarray1 *
ARRAY_REL int farray1_loc st.farray1 * STDIO st.stdio

Automation specialises pre-proved EvalM lemmas for each term that might be encountered in the monadic functions. As an example, a monadic function might contain an automatically defined function update_farray1 for updating array farray1. Anticipating this, synthesis automation can, at set-up time, automatically derive the following lemma which it can use when it encounters update_farray1.

$$\begin{array}{l}\vdash \mathsf{Eval}\ env\ e_1\ (\mathsf{num}\ n) \wedge \mathsf{Eval}\ env\ e_2\ (\mathsf{int}\ x) \wedge \\ \quad (\mathsf{lookup\_var}\ \lfloor\texttt{farray1}\rfloor\ env = \mathsf{Some\ farray1\_loc}) \Rightarrow \\ \quad \mathsf{EvalM}\ ro\ env\ st\ \lfloor\texttt{Array.update (farray1,}e_1\texttt{,}e_2\texttt{)}\rfloor \\ \qquad (\mathsf{monad\ unit\ exc}\ (\mathsf{update\_farray1}\ n\ x))\ (\mathsf{ASSERT},p)\end{array}$$

3.5 Changing monad types

The possibility to change the types of the monad is useful when previously developed monadic functions (e.g. from an existing library) are to be used as part of a larger context. Consider the case of the file I/O in the example from above. The following EvalM theorem has been proved in the CakeML basis library.

$$\begin{array}{l}\vdash \mathsf{Eval}\ env\ e\ (\mathsf{string}\ x) \wedge \\ \quad (\mathsf{lookup\_var}\ \lfloor\texttt{print}\rfloor\ env = \mathsf{Some\ print\_v}) \Rightarrow \\ \quad \mathsf{EvalM\ F}\ env\ st\ \lfloor\texttt{print}\ e\rfloor\ (\mathsf{monad\ unit}\ b\ (\mathsf{print}\ x))\ (\mathsf{STDIO},p)\end{array}$$

This can be used directly if the state type of the monad is the `IO_fs` type. However, our example above uses `example_state` as the state type.

To overcome such type mismatches, we define a function liftM which can bring a monadic operation defined in libraries into the required context. The type of liftM r w is $(\alpha,\ \beta,\ \gamma)\ \texttt{M} \rightarrow (\epsilon,\ \beta,\ \gamma)\ \texttt{M}$, for appropriate r and w.

$$\mathsf{liftM}\ read\ write\ op = (\lambda\, s.\ (\mathsf{let}\ (ret,new)\ =\ op\ (read\ s)\ \mathsf{in}\ (ret,write\ new\ s)))$$

Our liftM function changes the state type. A simpler lifting operation can be used to change the exception type.

For our example, we define stdio f as a function that performs f on the `IO_fs`-part of a `example_state`. (The fib example §1 used a similar stdio.)

$$\mathsf{stdio} = \mathsf{liftM}\ (\lambda\, s.\ s.\mathsf{stdio})\ (\lambda\, n\ s.\ s\ with\ \mathsf{stdio}\ :=\ n)$$

For synthesis, we prove a lemma that can transfer any EvalM result for the file I/O model to a similar EvalM result wrapped in the stdio function. Such

lemmas are possible because of the separation logic frame rule that is part of EvalM. The generic lemma is the following:

$$\vdash (\forall\, st.\ \mathsf{EvalM}\ ro\ env\ st\ exp\ (\mathsf{monad}\ a\ b\ op)\ (\mathsf{STDIO},p)) \Rightarrow \\ \forall\, st.\ \mathsf{EvalM}\ ro\ env\ st\ exp\ (\mathsf{monad}\ a\ b\ (\mathsf{stdio}\ op))\ (\mathsf{ASSERT},p)$$

And the following is the transferred lemma, which enables synthesis of HOL terms of the form stdio (print x) for Eval-synthesisable x.

$$\vdash \mathsf{Eval}\ env\ e\ (\mathsf{string}\ x) \wedge \\ (\mathsf{lookup\_var}\ \lfloor \texttt{print} \rfloor\ env = \mathsf{Some\ print\_v}) \Rightarrow \\ \mathsf{EvalM\ F}\ env\ st\ \lfloor \texttt{print}\ e \rfloor\ (\mathsf{monad\ unit\ exc}\ (\mathsf{stdio}\ (\mathsf{print}\ x)))\ (\mathsf{ASSERT},p)$$

4 Local state and the abstract synthesis mode

This section explains how we have adapted the method described above to also support generation of code that uses local state and local exceptions. These features enable use of stateful code (EvalM) in a pure context (Eval). We used these features to significantly speed up parts of the CakeML compiler (see §5).

In the monadic functions, users indicate that they want local state to be generated by using the following run function. In the logic, the run function essentially just applies a monadic function m to an explicitly provided state st.

$$\mathsf{run}\ :\ (\alpha,\ \beta,\ \gamma)\ \texttt{M}\ \rightarrow\ \alpha\ \rightarrow\ (\beta,\ \gamma)\ \texttt{exc} \\ \mathsf{run}\ m\ st = \mathsf{fst}\ (m\ st)$$

In the generated code, an application of run to a concrete monadic function, say bar, results in code of the following form:

```
fun run_bar k n =
  let
    val r = ref ... (* allocate, initialise, let-bind all local state *)
    fun foo n = ... (* all auxiliary funs that depend on local state *)
    fun bar n = ... (* define the main monadic function *)
  in Success (bar n) end (* wrap normal result in Success constructor *)
  handle e => Failure e; (* wrap any exception in Failure constructor *)
```

Synthesis of locally effectful code is made complicated in our setting for two reasons: (1) there are no fixed locations where the references and arrays are stored, e.g. we cannot define ref1_loc as used in the definition of ASSERT in §3.4; and (2) the local names of state components must be in scope for all of the function definitions that depend on local state.

Our solution to challenge (1) is to leave the location values as variables (loc_1, loc_2, loc_3) in the heap assertion when synthesising local state. To illustrate, we will adapt the `example_state` from §3.4: we omit `IO_fs` in the state because I/O cannot be made local. The local-state enabled heap assertion is:

$$\mathsf{LOCAL\_ASSERT}\ loc_1\ loc_2\ loc_3\ st = \\ \mathsf{REF\_REL\ int}\ loc_1\ st.\mathsf{ref1} * \mathsf{RARRAY\_REL\ int}\ loc_2\ st.\mathsf{rarray1} * \\ \mathsf{ARRAY\_REL\ int}\ loc_3\ st.\mathsf{farray1}$$

The lemmas referring to local state now assume they can find the right variable locations with variable look-ups.

$$\begin{array}{l}\vdash \mathsf{Eval}\ env\ e_1\ (\mathsf{num}\ n) \wedge \mathsf{Eval}\ env\ e_2\ (\mathsf{int}\ x) \wedge \\ \quad (\mathsf{lookup\_var}\ \lfloor \texttt{farray1} \rfloor)\ env = \mathsf{Some}\ loc_3) \Rightarrow \\ \quad \mathsf{EvalM}\ ro\ env\ st\ \lfloor \texttt{Array.update (farray1,}e_1\texttt{,}e_2\texttt{)} \rfloor \\ \qquad (\mathsf{monad\ unit\ exc}\ (\mathsf{update\_farray1}\ n\ x))\ (\mathsf{LOCAL\_ASSERT}\ loc_1\ loc_2\ loc_3{,}p)\end{array}$$

Challenge (2) was caused by technical details of our previous synthesis methods. The previous version was set up to only produce top-level declarations, which is incompatible with the requirement to have local (not globally fixed) state declarations shared between several functions. The requirement to only have top-level declarations arose from our desire to keep things simple: each synthesised function is attached to the end of a concrete linear program that is being built. It is beneficial to be concrete because then each assumption on the lexical environment where the function is defined can be proved immediately on definition. We will call this old approach the *concrete mode* of synthesis, since it eagerly builds a concrete program.

In order to support having functions access local state, we implement a new *abstract mode* of synthesis. In the abstract mode, each assumption on the lexical environment is left as an unproved side condition as long as possible. This allows us to define functions in a dynamic environment.

To prove a pure specification (Eval) from the EvalM theorems, the automation first proves that the generated state-allocation and -initialisation code establishes the relevant heap assertion (e.g. LOCAL_ASSERT); it then composes the abstractly synthesised code while proving the environment-related side conditions (e.g. presence of loc_3). The final proof of an Eval theorem requires instantiating the references-only *ro* flag to true, in order to know that no I/O occurs (§3.3).

5 Case studies and experiments

In this section we present the runtime and proof size results of applying our method to some case studies. Performance experiments were carried out on an Intel i7-2600 running at 3.4GHz with 16 GB of RAM. Full data is available at `https://cakeml.org/ijcar18.zip`.

Type Inference and Register Allocation. Both of these phases of the CakeML compiler are written with a state (and exception) monad, but were previously synthesised into pure CakeML code. We updated them to use the new synthesis tool, resulting in performant, stateful CakeML code. The allocator underwent more significant changes, because we could now use CakeML arrays via the synthesis tool. It was previously confined to using tree-like functional arrays for its internal state, leading to logarithmic access overheads. This is not a specific issue for the CakeML compiler; a verified register allocator for CompCert [3] also reported log-factor overheads due to (functional) array accesses.

Tests were carried out using versions of the bootstrapped CakeML compiler. We ran each test 50 times on the same input program, recording time elapsed in each compiler phase. For each test in the register allocation benchmark, we also compared the resulting executables 10 times, to confirm that both compilers generated code of comparable quality (i.e. runtime performance).

In the largest program (`knuth-bendix`), the new register allocator ran 15 times faster (with a wide 95% CI of 11.76–20.93 due in turn to a high standard deviation on the runtimes for the old code). In the smaller `pidigits` benchmark, the new register allocator ran 9.01 times faster (95% CI of 9.01–9.02). Across 6 example input programs, we saw ratios of runtimes between 7.58 and 15.06. Register allocation was previously such a significant part of the compiler runtime that this improvement results in runtime improvements for the whole compiler (on these benchmark programs) of factors between 2 and 9 times.

In contrast, the type inferencer became slower. We compared the performance of commit `28aba93` (incorporating the monadic inference code) against the same baseline. The slowdowns ranged between factors of approximately 3 and 1.17. However, the case with the most dramatic slowdown as a ratio still only represents a tiny proportion of the total time spent compiling. In this case (`pidigits`), the new code takes 10ms out of a total elapsed time of 2.05s (roughly 0.5% of the total). The best (least bad) case was in an artificial program exemplifying the worst-case for Hindley-Milner where types grow exponentially. There, the old code took 251ms and the new took 295ms. The extra indirection through references in the new code seems to cost performance. We intend to keep using the purely synthesised version until the compiler optimises the references better.

OpenTheory Article Checker. The type changing feature from §3.5 enabled us to produce an OpenTheory [7] article checker with our new synthesis approach, and reduce the amount of manual proof required in a previous version. The checker reads articles from the file system, and performs each logical inference in the OpenTheory framework using the verified Candle kernel [8]. Previously, the I/O code for the checker was implemented in stateful CakeML, and verified manually using characteristic formulae. By replacing the manually verified I/O wrapper by monadic code we removed 400 lines of tedious manual proof.

6 Related Work

Effectful code using monads. Our work on encapsulating stateful computations (§4) in pure programs is similar in purpose to that of the ST monad [11]. The main difference is how this encapsulation is performed: the ST monad relies on parametric polymorphism to prevent references from escaping their scope, whereas we utilise lexical scoping in synthesised code to achieve a similar effect.

Imperative HOL by Bulwahn et al. [4] is a framework for implementing and reasoning about effectful programs in Isabelle/HOL. Monadic functions are used to describe stateful computations which act on the heap, in a similar way as §3 but with some important differences. Instead of using a state monad, the authors

introduce a polymorphic *heap monad* – similar in spirit to the ST monad of Launchbury and Jones [11], but without encapsulation – where polymorphism is achieved by mapping HOL types to the natural numbers. Contrary to our approach, this allows for heap elements (e.g. references) to be declared on-the-fly and used as first-class values. The drawback, however, is that only countable types can be stored on the heap; in particular, the heap monad does not admit function-typed values, which our work supports.

More recently, Lammich [10] has built a framework for the refinement of pure data structures into imperative counterparts, in Imperative HOL. The refinement process is automated, and refinements are verified using a program logic based on separation logic, which comes with proof-tools to aid the user in verification.

Both developments [4, 10] differ from ours in that they lack a verified mechanism for extracting executable code from shallow embeddings. Although stateful computations are implemented and verified within the confines of higher-order logic, Imperative HOL relies on the unverified code-generation mechanisms of Isabelle/HOL. Moreover, neither work presents a way to deal with I/O effects.

Verified Compilation. Mechanisms for synthesising programs from shallow embeddings defined in the logics of interactive theorem provers exist as components of several verified compiler projects [1, 12, 6, 13]. Although the main contribution of our work is proof-producing synthesis, comparisons are relevant as our synthesis tool plays an important part in the CakeML compiler [9]. To the best of our knowledge, ours is the first work combining effectful computations with proof-producing synthesis and fully verified compilation.

CertiCoq by Anand et al. [1] strives to be a fully verified optimising compiler for functional programs implemented in Coq. The compiler front-end supports the full syntax of the dependently typed logic Gallina, which is reified into a deep embedding and compiled to Cminor through a series of verified compilation steps [1]. Contrary to the approach we have taken [13] (see §3.2), this reification is neither verified nor proof-producing, and the resulting embedding has no formal semantics (although there are attempts to resolve this issue [2]). Moreover, as of yet, no support exists for expressing effectful computations (such as in §3.4) in the logic. Instead, effects are deferred to wrapper code from which the compiled functions can be called, and this wrapper code must be manually verified.

The Œuf compiler by Mullen et al. [12] is similar in spirit to CertiCoq in that it compiles pure Coq functions to Cminor through a verified process. Similarly, compiled functions are pure, and effects must be performed by wrapper code. Unlike CertiCoq, Œuf supports only a limited subset of Gallina, from which it synthesises deeply embedded functions in the Œuf-language. The Œuf language has both denotational and operational semantics, and the resulting syntax is automatically proven equivalent with the corresponding logical functions through a process of computational denotation [12].

Hupel and Nipkow [6] have developed a compiler from Isabelle/HOL to CakeML AST. The compiler satisfies a partial correctness guarantee: if the generated CakeML code terminates, then the result of execution is guaranteed to relate to an equality in HOL. Our approach proves termination of the code.

7 Summary

This paper describes a technique that makes it possible to synthesise whole programs from monadic functions in HOL, with automatic proofs relating the generated effectful code to the original functions. Using the separation logic from characteristic formulae for CakeML, the synthesis mechanism supports references, exceptions, I/O, reusable library developments, and encapsulation of locally stateful computations inside pure functions. To our knowledge, this is the first proof-producing synthesis technique with the aforementioned features.

Acknowledgements. The second and fourth authors were partly supported by the Swedish Foundation for Strategic Research. The fifth author was supported by an A*STAR National Science Scholarship (PhD), Singapore.

References

1. Anand, A., Appel, A., Morrisett, G., Paraskevopoulou, Z., Pollack, R., Belanger, O.S., Sozeau, M., Weaver, M.: CertiCoq: A verified compiler for Coq. In: CoqPL (2017)
2. Anand, A., Boulier, S., Tabareau, N., Sozeau, M.: Typed Template Coq – Certified Meta-Programming in Coq. In: CoqPL (2018)
3. Blazy, S., Robillard, B., Appel, A.W.: Formal verification of coalescing graph-coloring register allocation. In: ESOP. LNCS, vol. 6012 (2010)
4. Bulwahn, L., Krauss, A., Haftmann, F., Erkök, L., Matthews, J.: Imperative functional programming with Isabelle/HOL. In: Mohamed, O.A., Muñoz, C.A., Tahar, S. (eds.) TPHOLs. LNCS, vol. 5170, pp. 134–149 (2008)
5. Guéneau, A., Myreen, M.O., Kumar, R., Norrish, M.: Verified characteristic formulae for CakeML. In: Yang, H. (ed.) ESOP. LNCS, vol. 10201, pp. 584–610 (2017)
6. Hupel, L., Nipkow, T.: A verified compiler from Isabelle/HOL to CakeML. In: Ahmed, A. (ed.) European Symposium on Programming (ESOP). Springer (2018)
7. Hurd, J.: The OpenTheory standard theory library. In: Bobaru, M.G., Havelund, K., Holzmann, G.J., Joshi, R. (eds.) NFM. LNCS, vol. 6617, pp. 177–191 (2011)
8. Kumar, R., Arthan, R., Myreen, M.O., Owens, S.: Self-formalisation of higher-order logic - semantics, soundness, and a verified implementation. J. Autom. Reasoning 56(3), 221–259 (2016)
9. Kumar, R., Myreen, M.O., Norrish, M., Owens, S.: CakeML: a verified implementation of ML. In: Jagannathan, S., Sewell, P. (eds.) POPL. pp. 179–192 (2014)
10. Lammich, P.: Refinement to Imperative/HOL. In: ITP. LNCS, vol. 9236 (2015)
11. Launchbury, J., Jones, S.L.P.: Lazy functional state threads. In: Sarkar, V., Ryder, B.G., Soffa, M.L. (eds.) PLDI. pp. 24–35 (1994)
12. Mullen, E., Pernsteiner, S., Wilcox, J.R., Tatlock, Z., Grossman, D.: Œuf: minimizing the Coq extraction TCB. In: CPP (2018)
13. Myreen, M.O., Owens, S.: Proof-producing translation of higher-order logic into pure and stateful ML. J. Funct. Program. 24(2-3), 284–315 (2014)
14. Owens, S., Myreen, M.O., Kumar, R., Tan, Y.K.: Functional big-step semantics. In: Thiemann, P. (ed.) ESOP. LNCS, vol. 9632, pp. 589–615 (2016)
15. Reynolds, J.C.: Separation logic: A logic for shared mutable data structures. In: LICS. pp. 55–74 (2002)