TCMC: A Novel Way of Compiling Functional Languages

Rafael D. Lins & Bruno O. Lira
Dept. de Informática - Universidade Federal de Pernambuco - Recife - Brasil
Computing Laboratory - The University of Kent - Canterbury - England

Abstract

The efficient compilation of functional languages has been shown to be a difficult task. The most successful implementations so far generate code in assembly language. This makes implementation extremely hard and machine dependent. In this paper we present TCMC, a new abstract machine, in which we transfer the control of the execution flow to C, as much as possible. TCMC takes advantage of the extremely low costs of procedure calls in modern RISC architectures. This produces a substantial improvement in performance, as we show here.

Introduction

Due to their semantic elegance, expressive power and ease in proving the correctness of programs, functional languages have been pointed out as a possible solution for the problem of programming known as the Software Crisis. In such languages programs are written as a set of function definitions and an expression, whose value is the result of the program. The evaluation is accomplished through consecutively rewriting the expression according to the functions definitions. Functional languages seem to be harder to implement than conventional imperative ones. At execution time, we must maintain complicated structures, such as unevaluated function applications, which allow us to work with higher-order functions and infinite lists.

The traditional way to implement lazy functional languages was graph interpretation of combinators, as introduced by Turner in [12]. The understanding of the evaluation mechanisms of these languages allowed implementation to move from interpretation towards compilation, with substantial gain in performance. Cardelli’s abstract machine FAM [3] developed for the compilation of strict functional languages was an important step in this search for efficiency.

Johnsson [10] developed a strategy for compiling lazy functional languages, described as an abstract stack machine, called the G-Machine. The basic principle of the G-Machine is to avoid generating graphs. The code generated by the G-Machine when executed produces time and space performance at least an order of magnitude faster than interpreted functional languages. The original G-Machine implemented at Chalmers, Göteborg, by Johnsson and his colleagues [10], generated code in VAX-780 Assembly language, which made implementation extremely hard and machine dependent. It was common sense in the community of implementation of functional languages that assembly language implementation was the price to pay if one wanted efficiency. The Chalmers LML compiler is still a reference in terms of performance of lazy functional languages. The G-machine way of controlling the execution flow and evaluation was followed by most of other implementations, even the ones based on different abstract machines as the Spineless G-Machine [1], the Spineless Tagless G-machine [9], TIM [2], and GM-C [8].

The first author has made several implementations of compiled functional languages in C [8, 6, 11, 7], which were close, but worse, in performance to the best assembly language implementations. All these C-based implementations were portable and simpler than the assembly ones. C was used as a macro-assembler and all “execution flow control” was made on a higher-level abstract machine.

In this paper we present TCMC, a new abstract machine, in which we transfer the execution flow control to C, as much as possible. The key idea behind TCMC is to take advantage of efficient context switching in modern architectures based on RISC, which is able to implement function calls at a very low cost. We also observed that the object code generated by C compilers is extremely neat and very fast. These factors lead us to try to translate each function definition into a procedure in C. It is obvious that not all scripts could be translated into C, if we wanted to have a lazy functional
language. However, it is safe to translate strict functions on all arguments that produce results of
ground type as procedures in C. The same is also true for arithmetic expressions wherever they appear.
This is the key for the efficiency of \( \Gamma \text{CMC} \). A higher-level abstract machine is still needed to glue
together procedure calls, unevaluated expressions and functions, data-structures, etc. Categorical
Multi-Combinators [4, 7] served as a basis for the evaluation model of the \( \Gamma \text{CMC} \) abstract machine.
Our experience with GM-C [8] and CM-CM [6, 11] was fundamental for the design, implementation,
and optimisation of \( \Gamma \text{CMC} \).

In this paper we also compare the performance of \( \Gamma \text{CMC} \) with the Chalmers LML compiler and
with GMC [8].

Categorical Multi-Combinators

In this section we present a brief introduction to Categorical Multi-Combinators [4], a rewriting system
which provides the computational model for \( \Gamma \text{CMC} \). Later, we show how to compile a functional
language directly into \( \Gamma \text{CMC} \) code.

The Source Language

A program is taken to be a sequence of combinator definitions together with an expression to be
evaluated, which will involve these combinators.

\[
c_1 = \text{def combinator}_1 \\
\ldots \\
c_n = \text{def combinator}_n
\]

\[
\text{main-expression}
\]

A program when compiled will generate a script which is formed by a sequence of combinators linked
to their code thus,

\[
\rho = \begin{bmatrix}
\text{c}_1 & \mapsto [\text{combinator}_1] \\
\vdots \\
\text{c}_n & \mapsto [\text{combinator}_n]
\end{bmatrix}
\]

The main-expression is compiled separately as,

\[
[[\text{main-expression}] \rho]
\]

In order properly to interpret recursion, we assume that the environment \( \rho \) contains the definition of
all combinators, so that recursive combinators produce recursive references through the environment.
The notation we use is: with each combinator \( c \) there is associated code \( c_r \), we supress the environment
\( \rho \) when no confusion is possible.

Compiling into Categorical Multi-Combinators

In Categorical Multi-Combinators function application is denoted by juxtaposition, taken to be
left-associative. The compilation algorithm for translating \( \lambda \)-expressions into Categorical Multi-
Combinators is given by the function \( R^{e \ldots x_j} \) where each \( x_i \) is a variable and the corresponding \( i \)
its depth in the environment, i.e. the corresponding DeBruijn number. Top level expressions are
translated using an empty environment, so by \( R^{1} \). For a matter of uniformity combinators will be
represented as composed with a dummy frame, (.), which can be seen as the identity frame.

(T.1) \[ R^{1} \lambda x_k \ldots x_1. a = (L^{m-1}(R^{e \ldots x_j} a). ()) \]

(T.2) \[ R^{e \ldots x_j} a \ldots b = R^{e \ldots x_j} a \ldots R^{e \ldots x_j} b \]

(T.3) \[ R^{e \ldots x_j} b = b , \text{ if } b \text{ is a constant} \]

(T.4) \[ R^{e \ldots x_j} x_i = i \]

Combinator names are treated as constants.
**Example of Compilation**

The script:

\[
S = \lambda a. \lambda b. \lambda c. ac(bc) \\
K = \lambda k. \lambda l. k \\
I = \lambda i. i \\
SKKI
\]

forms the following environment:

\[
S \mapsto R^1[\lambda a. \lambda b. \lambda c. ac (bc)] \\
K \mapsto R^1[\lambda k. \lambda l. k] \\
I \mapsto R^1[\lambda i. i]
\]

which by application of the compilation rules above translates to:

\[
S \mapsto \langle L^2(2 0 (1 0)), () \rangle \\
K \mapsto \langle L^1(1), () \rangle \\
I \mapsto \langle L^1(0), () \rangle
\]

The expression to be evaluated is translated as

\[
R^1[[SKKI]]
\]

which generates \( SKKI \) as compiled code.

**Categorical Multi-Combinator Rewriting Laws**

The core of the Categorical Multi-Combinator machine is presented on page 71 of [4]. For a matter of convenience we will represent the multi-pair combinator, which forms evaluation environments as \((x_0, \ldots, x_n)\) and compositions, which represent closures, will be written as \(\langle a, b \rangle\). Using this notation the kernel of the Categorical Multi-Combinator rewriting laws is:

(\(M^*.1\)) \(\langle n, (x_m, \cdots, x_1, x_0) \rangle \Rightarrow x_n\)

(\(M^*.2\)) \(\langle x_0 x_1 x_2 \cdots x_n, y \rangle \Rightarrow \langle x_0, y \rangle \cdots \langle x_n, y \rangle\)

(\(M^*.3\)) \(\langle L^r(y), (w_0, \ldots, w_m) \rangle x_0 x_1 \cdots x_n x_{n+1} \cdots x_z \Rightarrow \langle y, (x_0, \cdots, x_n) \rangle x_{n+1} \cdots x_z\)

The state of computation of a Categorical Multi-Combinator expression is represented by the expression itself. Rule (\(M^*.1\)) performs environment look-up, this is the mechanism by which a variable fetches its value in the corresponding environment. (\(M^*.2\)) is responsible for environment distribution. The rule (\(M^*.3\)) performs environment formation: if during rewriting a combinator reaches the leftmost position of the code we proceed a script look-up and enter the corresponding code in the definition environment. This can be expressed as

\(\langle l, y \rangle \Rightarrow \langle l, y \rangle\)

**From CM-C into \(\Gamma\)CMC**

In this section we give an overview of the \(\Gamma\)CMC evaluation mechanism.

If one observes the rewriting rules for Categorical Multi-Combinators above we see that rule (\(M^*.3\)) is equivalent to \(\lambda\)-Calculus \(\beta\)-reduction in which substitutions are performed on demand. For
a matter of convenience we will structure the Categorical Multi-Combinator expression in two parts: the reduction stack $T$ and the heap $H$, where we place evaluation environments. The transition

$$\langle T, H \rangle \Rightarrow \langle T', H' \rangle$$

must be interpreted as: “when the machine arrives at state $\langle T, H \rangle$, it can get to state $\langle T', H' \rangle$. It is easy to see that the rewriting laws above can be rewritten as state transition rules:

1. $\langle (y, e_1), c, H[e_1 = (x_m, \ldots, x_0)] \rangle \Rightarrow \langle x_m, H[e_1 = (x_m, \ldots, x_0)] \rangle$
2. $\langle \langle x_0 \ldots x_n, e_1 \rangle, c, H[e_1 = \ldots] \rangle \Rightarrow \langle \langle x_0, e_1 \rangle \ldots \langle x_n, e_1 \rangle, c, H[e_1 = \ldots] \rangle$
3. $\langle \langle L^n(y), e_1 \rangle x_0 \ldots x_n, c, H \rangle \Rightarrow \langle \langle y, e_1 \rangle, c, H[e_1 = (x_0, \ldots, x_n)] \rangle$
4. $\langle \langle l, e_1 \rangle, c, H \rangle \Rightarrow \langle \langle l, e_1 \rangle, c, H \rangle$

Instead of manipulating references to environments directly as above we have a stack which keeps references to the current environment. Variables on the top position of the reduction stack fetch their values from the current environment. The current environment changes whenever a variable fetches a closure from the current environment or by creating a new environment via $\beta$-reduction. We call the environment stack $E$.

1. $\langle u.c, H[c = ((x_m, e_m), \ldots, (x_0, e_0)), e, E] \rangle \Rightarrow \langle x_m, c, H[c_1 = ((x_m, e_m), \ldots, (x_0, e_0)), e_1, E] \rangle$
2. $\langle \langle L^n(y), e_1 \rangle x_0 \ldots x_n, c, H \rangle \Rightarrow \langle \langle y, e_1 \rangle, c, H[e_1 = ((x_0, e_1), \ldots, (x_n, e_1))], e_2, E \rangle$
3. $\langle \langle l, e_1 \rangle, c, H \rangle \Rightarrow \langle \langle l, e_1 \rangle, c, H \rangle$

Example of Evaluation

The expression $SKKI$, where $S, K,$ and $I$ correspond to the following entries in the script

$$S \mapsto L^3(2\ 0\ (1\ 0))$$
$$K \mapsto L^1(1)$$
$$I \mapsto L^3(0)$$

is evaluated as,

$$\langle SKKI, H, E \rangle \xrightarrow{\beta} \langle L^3(2\ 0\ (1\ 0))\ K\ K\ I, H, E \rangle$$
$$\xrightarrow{\gamma} \langle 2\ 0\ (1\ 0), H[e_1 = (K\ K\ I)], e_1, E \rangle$$
$$\xrightarrow{\delta} \langle K\ 0\ (1\ 0), H[e_1 = (K\ K\ I)], e_1, E \rangle$$
$$\xrightarrow{\gamma} \langle L^1(1)\ 0\ (1\ 0), H[e_1 = (K\ K\ I)], e_1, E \rangle$$
$$\xrightarrow{\gamma} \langle 1, H[e_2 = ((0, e_1), (1\ 0), e_1))][e_1 = (K\ K\ I)], e_2, e_1, E \rangle$$
$$\xrightarrow{\delta} \langle 0, H[e_2 = ((0, e_1), (1\ 0), e_1))][e_1 = (K\ K\ I)], e_1, e_2, e_1, E \rangle$$
$$\xrightarrow{\gamma} \langle I, H[e_2 = ((0, e_1), (1\ 0), e_1))][e_1 = (K\ K\ I)], e_1, e_2, e_1, E \rangle$$

As there are no arguments on the evaluation stack we stop evaluation. The abstract machine presented above resembles the evaluation mechanism of CM-CM [6].

Special Functions

Strict functions on all arguments which produce results of ground type are called special. These functions will fetch their arguments from the evaluation stack and return the result of evaluation to the top of $T$. The evaluation of special functions happens outside $\Gamma CM C$. All $\Gamma CM C$ does is to prepare the arguments for them and receive the result. We introduce a new state transition law for special functions:

4. $\langle f^n x_0 \ldots x_n, c, H, e, E \rangle \Rightarrow \langle f, c, H, e, E \rangle$

where $f_s = f^n(x_0, \ldots, x_n)$ and $x\_i$ is the weak head normal form of $x_i$. 

4
Arithmetic expressions, in general, will be lifted from the code and will be treated in a similar way to special functions. For instance,

\[ S' = \lambda a. \lambda b. \lambda c. a(b + c) \]

generates the following script:

\[ S' \mapsto L^2(2 \ (f_1 \ 1 \ 0)) \]
\[ f_1 \mapsto (1' + 0') \]

From Interpretation to Compilation

The Categorical Multi-Combinator structures which appear on T are now replaced by code which when executed will generate a corresponding data structure on T. This data structure is interpreted by using the state transition laws above. A variable \( n \), for example, is generated on top of T by using a MKTvar\( (n) \) instruction.

The code will always try to predict the behaviour of evaluation and avoid generating intermediate expressions, as much as possible. The novel aspect of CMC if compared with its predecessors lies on translating special functions into procedures in C. In the next section we see CMC in more details.

We must keep in mind the code sequences generated will perform operations equivalent to the naive CMC machine in this section.

Compiling into CMC Code

We present here the complete set of direct compilation rules for the kernel of CMC.

A program in CMC is formed by a set of function definitions plus an expression, which we want to evaluate, as follows:

\[
\begin{align*}
   f_1 x_1 \ldots x_n &= \text{body} - \text{of} - f_i \\
   f_2 x_1 \ldots x_n &= \text{body} - \text{of} - f_i \\
   &\vdots \\
   f_n x_1 \ldots x_n &= \text{body} - \text{of} - f_i \\
   \text{expression?}
\end{align*}
\]

The expression to be evaluated is compiled by scheme \( \mathcal{E} \) called as:

\[ \mathcal{E}[\text{expression}] \]

Strict functions on all arguments which produce results of ground type are called *special*. These functions will be compiled directly as procedures in C. Special functions are compiled as:

\[
\begin{align*}
   f_i x_1 \ldots x_n &= \text{body} - \text{of} - f_i \\
   f_i \mapsto \mathcal{S}[\text{body} - \text{of} - f_i]
\end{align*}
\]

Ordinary functions are compiled as:

\[
\begin{align*}
   f_j x_1 \ldots x_l &= \text{body} - \text{of} - f_j \\
   f_j \mapsto \mathcal{T}[x_1 \ldots x_l][\text{body} - \text{of} - f_j]
\end{align*}
\]

Scheme \( \mathcal{E} \)

This scheme is responsible for the printing routine and driving the evaluation mechanism.

1. \( \mathcal{E}[k] = \text{printf}(k); \) if \( k \) is a constant
2. \( \mathcal{E}[a + b] = t1 = S'[a]; t2 = S'[b]; \text{printf}(t1 + t2); \)
3. \( \mathcal{E}[\text{if } a > b \text{ then } c \text{ else } d] = \text{if}_{\text{true}}(A, C, D); \) where \( A \mapsto T^{[a]}; t1 = \ast (\text{top T} - 1) \mapsto \text{rem-value}; T^{[b]}; t2 = \ast (\text{top T} - 1) \mapsto \text{rem-value}; t1 > t2; \)
\( C \mapsto \mathcal{E}[c]; D \mapsto \mathcal{E}[d]; \)
4. \( \mathcal{E}[f_i x_0 \ldots x_n] = \text{printf}(f_i[S'[x_0], \ldots, S'[x_n]]); \)
5. \( \mathcal{E}[f_i \ldots] = \mathcal{T}[f_i \ldots] \text{ printf}(); \)
Scheme $S$

This scheme is responsible for starting-up the compilation of special functions generating procedures in C.

1. $S[k] = \text{return}(k)$; if $k$ is a constant
2. $S[x] = \text{return}(x)$; if $x$ is a variable
3. $S[a + b] = \text{return}(S'[a] + S'[b])$
4. $S[if \ (a > b) \ then \ c \ else \ d] = if \ (S'[a] > S'[b]) \ \{S[c]\}; \ else\{S[d]\}$
5. $S[f, x_0 \ldots x_j] = \text{return}(f(S'[x_0], \ldots, S'[x_j]));$ if $f_i$ is a special function
6. $S'[f, x_0 \ldots x_j] = T[[f, x_0 \ldots x_j] \ \text{return}(((\ast\text{topT} - -)) \rightarrow \text{rem.value})$;

Scheme $S'$

This scheme is ancillary to $S$ and is responsible for the compilation of inner parts of the body of a special function, generating parts of procedure code in C.

1. $S'[k] = k$ if $k$ is a constant
2. $S'[x] = x$ if $x$ is a variable
3. $S'[a + b] = (S'[a] + S'[b])$
4. $S'[if \ (a > b) \ then \ c \ else \ d] = if \ (S'[a] > S'[b]) \ \{S'[c]\}; \ else\{S'[d]\}$
5. $S'[f, x_0 \ldots x_j] = f(S'[x_0], \ldots, S'[x_j]);$ if $f_i$ is a special function.
6. $S'[f, x_0 \ldots x_j] = T[[f, x_0 \ldots x_j] \ \text{return}(((\ast\text{topT} - -)) \rightarrow \text{rem.value})$;

Example of Compilation

Let us show an example of special function compilation. If we have the script:

\[
\text{fib } n = \text{if } n < 2 \text{ then } 1 \text{ else fib}(n-1) + \text{fib}(n-2) \\
\text{fib } 20?
\]

it will be compiled as,

\[
\mathcal{E}[\text{fib } 20] \xrightarrow{\Delta} \text{printf(fib(S'[20])}; \\
\xrightarrow{S'_{1.1}} \text{printf(fib(20))};
\]

Now we compile

\[
\text{fib } n = \text{if } n < 2 \text{ then } 1 \text{ else fib}(n-1) + \text{fib}(n-2)
\]

by using scheme $S$.

\[
\text{fib } \mapsto S[if \ n < 2 \ then \ 1 \ else \ fib(n - 1) + fib(n - 2)]
\]

\[
S_{1.4} \xrightarrow{\Delta} \text{if } (S'[n] < S'[2]) \ \{S'[1]\}; \ else\{S'[fib(n - 1) + fib(n - 2)]\}$
\]

\[
S_{1.2} \xrightarrow{\Delta} \text{if } (n < S'[2]) \ \{S'[1]\}; \ else\{S'[fib(n - 1) + fib(n - 2)]\}$
\]

\[
S_{1.1} \xrightarrow{\Delta} \text{if } (n < 2) \ \{S'[1]\}; \ else\{S'[fib(n - 1) + fib(n - 2)]\}$
\]

\[
S_{1.0} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(n - 1) + fib(n - 2)]\}$
\]

\[
S_{1.5} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(S'[n - 1]) + S'[fib(n - 2)]]\}$
\]

\[
S_{1.4} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(S'[n - 1]) + fib(S'[n - 2])]\}$
\]

\[
S_{1.3} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(n - S'[1]) + \text{fib}(S'[n - 2])]\}$
\]

\[
S_{1.2} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(n - S'[1]) + \text{fib}(S'[n - 2])]\}$
\]

\[
S_{1.1} \xrightarrow{\Delta} \text{if } (n < 2) \ \{return(1)\}; \ else\{S'[fib(n - 1) + \text{fib}(n - 2)]\}$
\]
As we can see the result of compilation is a procedure in C, which needs only a heading with type declarations to be compiled and executed by the C machine.

Scheme $T$

This scheme is responsible for the compilation of ordinary functions and generates code which is handled by the abstract machine. We assume the arity of a function $f_n$ to be $n + 1$.

1. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n x_0 \ldots x_n] = T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n] T^{x_0\rightarrow\ldots\rightarrow x_n}[x_0] \ldots T^{x_0\rightarrow\ldots\rightarrow x_n}[x_n]$
2. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n x_0 \ldots x_n] = \text{MKTc}(f_n (Z^{x_0\rightarrow\ldots\rightarrow x_n} x_0, \ldots, Z^{x_0\rightarrow\ldots\rightarrow x_n} x_n))$; if $f_n$ is a special function
3. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n x_0 \ldots x_n] = \text{MKE}(n + 1); G^{x_0\rightarrow\ldots\rightarrow x_n}[x_0] \ldots G^{x_0\rightarrow\ldots\rightarrow x_n}[x_n] \text{Push}(f_n); \text{Pop}(\text{env});$
4. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n] = \text{eval}(0) \ldots \text{eval}(n); \text{MKT}(n, f_n ((\text{top} - n) \rightarrow \text{rem}.value), \ldots, ((\text{top} - n) \rightarrow \text{rem}.value))$; if $f_n$ is a special function.
5. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n] = \text{MKE}(n + 1); \text{Push}(f_n); \text{Pop}(\text{env});$
6. $T^{x_0\rightarrow\ldots\rightarrow x_n}[k] = \text{MKTc}(k)$; if $k$ is a constant
7. $T^{x_0\rightarrow\ldots\rightarrow x_n}[x_i] = \text{MKV}(x_i)$; if $x_i$ is evaluated
8. $T^{x_0\rightarrow\ldots\rightarrow x_n}[y_i] = \text{eval}(i)$
9. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a > b] = T^{x_0\rightarrow\ldots\rightarrow x_n}[a] T^{x_0\rightarrow\ldots\rightarrow x_n}[b] \text{rem}.value; T^{x_0\rightarrow\ldots\rightarrow x_n}[t_1 = (s(\text{top} - n) - s(\text{top} - n)) \rightarrow \text{rem}.value; t_1 = t_2; C = T^{x_0\rightarrow\ldots\rightarrow x_n}[c] D = T^{x_0\rightarrow\ldots\rightarrow x_n}[d]$
10. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = T^{x_0\rightarrow\ldots\rightarrow x_n}[a] T^{x_0\rightarrow\ldots\rightarrow x_n}[b] \text{MKTc}(t_1 + t_2);$
11. $T^{x_0\rightarrow\ldots\rightarrow x_n}[x_0 x_1 \ldots x_m] = \text{MKT}(\text{top}, (\text{MKT} T^{x_0\rightarrow\ldots\rightarrow x_n}[x_0] \ldots T^{x_0\rightarrow\ldots\rightarrow x_n}[x_m]));$
12. $T^{x_0\rightarrow\ldots\rightarrow x_n}[x_0 \ldots x_n] = T^{x_0\rightarrow\ldots\rightarrow x_n}[x_0] \ldots T^{x_0\rightarrow\ldots\rightarrow x_n}[x_n].$

Scheme $G$

This scheme generates code which when executed fills the fields of a cell in an evaluation environment.

1. $G^{x_0\rightarrow\ldots\rightarrow x_n}[k] = \text{MKE}(k)$
2. $G^{x_0\rightarrow\ldots\rightarrow x_n}[x_i] = \text{MKE}(x_i)$
3. $G^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(Z^{x_0\rightarrow\ldots\rightarrow x_n}[a] + Z^{x_0\rightarrow\ldots\rightarrow x_n}[b])$
4. $G^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKT}(A, i)$
5. $G^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(A, i)$
6. $G^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(A, i)$
7. $G^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(A, i)$
8. $G^{x_0\rightarrow\ldots\rightarrow x_n}[f_n] = \text{MKE}(f_n, i)$

Scheme $T'$

This scheme produces code which when executed generates cells on the top of the $T$-stack.

1. $T^{x_0\rightarrow\ldots\rightarrow x_n}[k] = \text{MKTc}(k)$
2. $T^{x_0\rightarrow\ldots\rightarrow x_n}[x_i] = \text{MKTvar}(i)$
3. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKTc}(Z^{x_0\rightarrow\ldots\rightarrow x_n}[a] + Z^{x_0\rightarrow\ldots\rightarrow x_n}[b])$
4. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKT}(A)$
5. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKT}(A)$
6. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(A)$
7. $T^{x_0\rightarrow\ldots\rightarrow x_n}[a + b] = \text{MKE}(A)$
8. $T^{x_0\rightarrow\ldots\rightarrow x_n}[f_n] = \text{MKT}(f_n)$
Scheme $\mathcal{Z}$
This scheme makes parameters ready for special functions or arithmetic expressions whenever called inside an ordinary function.
1. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[k] = k$;
2. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[x_i] = (((\text{topE} + (*\text{topE}) \to \text{tio} + 1 - i) \to \text{rem} \cdot \text{graph}) \to \text{rem} \cdot \text{value;}}$ if $x_i$ is already evaluated.
3. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[x_i] = \text{evalenv}(i); \text{return}(*\text{topT} - -) \to \text{rem} \cdot \text{value};$)
4. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[f \ a \ldots b] = f_i(\mathcal{Z}^{x_0 \ldots x_{i-1}}[a], \ldots, \mathcal{Z}^{x_0 \ldots x_{i-1}}[b])$;
5. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[a \ldots b] = \text{eval}(\text{topT} - -) \to \text{rem} \cdot \text{value};$ if $f_i$ is a special function.
6. $\mathcal{Z}^{x_0 \ldots x_{i-1}}[a + b] = \mathcal{Z}^{x_0 \ldots x_{i-1}}[a] + \mathcal{Z}^{x_0 \ldots x_{i-1}}[b]$;

Scheme $\mathcal{Z'}$
This scheme makes parameters ready for special functions or arithmetic expressions whenever called inside a cell generating scheme.
1. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[k] = k$;
2. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[x_i] = (((\text{topE} + (*\text{topE}) \to \text{tio} + 1 - i) \to \text{rem} \cdot \text{graph}) \to \text{rem} \cdot \text{value;}}$ if $x_i$ is already evaluated.
3. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[x_i] = \text{evalenv}(i); \text{return}(*\text{topT} - -) \to \text{rem} \cdot \text{value};$)
4. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[f \ a \ldots b] = f_i(\mathcal{Z'}^{x_0 \ldots x_{i-1}}[a], \ldots, \mathcal{Z'}^{x_0 \ldots x_{i-1}}[b])$;
5. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[a \ldots b] = \text{eval}(\text{topT} - -) \to \text{rem} \cdot \text{value};$ if $f_i$ is a special function.
6. $\mathcal{Z'}^{x_0 \ldots x_{i-1}}[a + b] = \mathcal{Z'}^{x_0 \ldots x_{i-1}}[a] + \mathcal{Z'}^{x_0 \ldots x_{i-1}}[b]$;

Example of Compilation
Let us show an example of compilation of an ordinary function. If we have the script:

```plaintext
def n = if n<2 then 1 else fib(n-1) + fib(n-2)
def twice f x = f (f x)
def fib 5?
```

it will be compiled as,

\[
\mathcal{E}[\text{twice fib 5}] \xrightarrow{\mathcal{E}} \mathcal{T}^{\text{twice fib 5}}; \text{print}(1);
\]

where A is:

\[
A \mapsto \mathcal{T}^{A}^{\text{fib}}
\]

Now we compile

```plaintext
def twice f x = f (f x)
def T as,
```

\[
\text{twice} \mapsto \mathcal{T}^{\text{twice}}[f (f x)]
\]

where A' is:

\[
A' \mapsto \mathcal{T}^{A'}^{\text{fib}}
\]

fib is as in the previous example of compilation above.

\[
\begin{align*}
\text{fib} & \quad \text{is a special function.} \\
\end{align*}
\]

\[
\begin{align*}
\text{inside an ordinary function.} & \\
\text{This scheme makes parameters ready for special functions or arithmetic expressions whenever called inside an ordinary function.} & \\
\end{align*}
\]

\[
\begin{align*}
\text{inside a cell generating scheme.} & \\
\text{This scheme makes parameters ready for special functions or arithmetic expressions whenever called inside a cell generating scheme.} & \\
\end{align*}
\]
State Transition Laws

We present GC\textsc{mc} as a state transition machine. A state of GC\textsc{mc} is a 5-uple

\[ \langle C, T, H, O, E \rangle \]

in which each component is interpreted in the following way:

\textbf{C}: The code to be executed.

This code is generated by the translation rules presented by the compilation schemes above.

\textbf{T}: The reduction stack. The top of \textbf{T} points to the part of the graph to be evaluated.

\textbf{H}: The heap where graphs are stored. The notation \( H[d = e_1 \ldots e_n] \) means that there is in \( H \) a \( n \)-component cell named \( d \). The fields of \( d \) are filled with \( e_1 \ldots e_n \), in this order.

Cells are fully-boxed.

\textbf{O}: The output.

\textbf{E}: The environment stack. Its top contains a reference to the current environment.

GC\textsc{mc} is defined as a set of transition rules. The transition

\[ \langle C, T, H, O, E \rangle \Rightarrow \langle C', T', H', O', E' \rangle \]

must be interpreted as: “when the machine arrives at state \( \langle C, T, H, O, E \rangle \), it can get to state \( \langle C', T', H', O', E' \rangle \)”.

We present here the complete set of state transition laws for the kernel of GC\textsc{mc}:

1. \( \text{print}(c, d.T, H[d=k], O, E) \Rightarrow \langle c, T, H[d=k], k, O, E \rangle \)
2. \( \text{eval}(c, d.T, H[d=k], O, E) \Rightarrow \langle c, d.T, H[d=k], O, E \rangle \)
3. \( \text{eval}(c, d.T, H[d = (A,c)], O, E) \Rightarrow \langle A, \text{Popenv}(c, T, H[d = (A,c)], O, E) \rangle \)
4. \( \text{eval}(c, d.T, H[d = f_n], O, E) \Rightarrow \langle f_n.c, T, H[d = f_n], O, E \rangle \)
5. \( \text{eval}(i).c, d, \ldots d, \ldots, T, H[d = k], O, E \Rightarrow \langle c, \ldots d, \ldots T, H[d = k], O, E \rangle \)
6. \( \text{eval}(i).c, d, \ldots d, \ldots, T, H[d = (A,c)], O, E \Rightarrow \langle A, \text{Pop}(i).c, d, \ldots d, \ldots T, H[d = (A,c)], O, E \rangle \)
7. \( \text{eval}(i).c, d, \ldots d, \ldots, T, H[d = f_n], O, E \Rightarrow \langle c, \ldots d, \ldots T, H[d = f_n], O, E \rangle \)
8. \( \text{eval}(i).c, T, H[e = \ldots f_n], O, E \Rightarrow \langle f_n, T, H[e = \ldots f_n], O, E \rangle \)
9. \( \text{eval}(i).c, T, H[e = \ldots u_\ldots], O, E \Rightarrow \langle \text{eval}(c, u.T, H[e = \ldots u_\ldots]), O, E \rangle \)
10. \( \text{MKenv}(n).c, d_1, \ldots d_n, \ldots d_m, T, H, O, E \Rightarrow \langle c, d_{n+1}, \ldots d_m, T, H[e = d_1, \ldots d_n], O, E \rangle \)
11. \( \text{Popenv}(c, T, H, O, E) \Rightarrow \langle c, T, H, O, E \rangle \)
12. \( \text{Pop}(i).c, d, d_1, \ldots d, \ldots T, H, O, E \Rightarrow \langle c, d_1, \ldots d, \ldots T, H, O, E \rangle \)
13. \( \text{Pushfun}(f).c, T, H, O, E \Rightarrow \langle f.c, T, H, O, E \rangle \)
14. \( \text{If} \text{true}(\text{True}, A, B).c, T, H, O, E \Rightarrow \langle A.c, T, H, O, E \rangle \)
15. \( \text{If} \text{true}(\text{False}, A, B).c, T, H, O, E \Rightarrow \langle B.c, T, H, O, E \rangle \)
16. \( \text{MKTvar}(k).c, T, H[e = e_0 \ldots e_m], O, E \Rightarrow \langle c, d.T, H[d = e_{m+1}] = e_0 \ldots e_m], O, E \rangle \)
17. \( \text{MKTcomp}(P).c, T, H, O, E \Rightarrow \langle c, d.T, H[d = (P, c)], O, E \rangle \)
18. \( \text{MKTpc}(A).c, T, H, O, E \Rightarrow \langle c, d.T, H[d = A], O, E \rangle \)
19. \( \text{MKTcte}(k).c, T, H, O, E \Rightarrow \langle c, d.T, H[d = k], O, E \rangle \)
20. \( \text{MKTk}(n).c, d_0, \ldots d_n, T, H[d_n = \ldots], O, E \Rightarrow \langle c, d_n, \ldots d_m.T, H[d_n = A], O, E \rangle \)
21. \( \text{MKEvar}(j.i).c, T, H[e_1 = \ldots u_\ldots] = e_0 = \ldots e_j \ldots] = e_0 = \ldots e_j \ldots] = c, c_1.c_0.E \Rightarrow \langle c, T, H[e_1 = \ldots u_\ldots] = e_0 = \ldots e_j \ldots] = c, c_1.c_0.E \rangle \)
22. \( \text{MKEcte}(j.i).c, T, H[e_1 = \ldots u_\ldots] = c, E \Rightarrow \langle c, T, H[e_1 = \ldots k_\ldots], O, E \rangle \)
23. \( \text{MKEcell}(n).c, T, H, O, E \Rightarrow \langle c, d.T, H[e = u_1 \ldots u_n], O, E \rangle \)
24. \( \text{MKEcomp}(A).c, T, H[e_1 = \ldots u_\ldots], O, c_1.c_0.E \Rightarrow \langle c, T, H[e_1 = \ldots (A, c_1), O, c_1.c_0.E \rangle \)
25. \( \text{MKEp}(A).c, T, H[e = \ldots i_\ldots], O, E \Rightarrow \langle c, d.T, H[e = A \ldots], O, E \rangle \)
26. \( \text{If} \text{true}(\text{True}, B, C).c, T, H[e_1 = \ldots u_\ldots], O, c_1.c_0.E \Rightarrow \langle c, T, H[e_1 = \ldots (B, c_1), O, c_1.c_0.E \rangle \)
27. \( \text{If} \text{true}(\text{False}, B, C).c, T, H[e_1 = \ldots u_\ldots], O, c_1.c_0.E \Rightarrow \langle c, T, H[e_1 = \ldots (C, c_1), O, c_1.c_0.E \rangle \)
Example of Evaluation

As we saw in the examples of compilation above, the program:

\[
\begin{align*}
\text{fib } n &= \text{if } n < 2 \text{ then } 1 \text{ else fib}(n-1) + \text{fib}(n-2) \\
twice \ f \ x &= f (f \ x) \\
twice \ \text{fib} \ 5 &? \\
\end{align*}
\]

compiled as,

\[
\begin{align*}
twice \ \text{fib} \ 5 &\Rightarrow \text{MKEcell}(2) \cdot \text{MKEp}(A, 1) \cdot \text{MKEcte}(5, 0) \cdot \text{Pushfun}(\text{twice}) \cdot \text{Popenv} \cdot \text{print}() ; \\
A &\Rightarrow \text{eval}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \rightarrow \text{rem.value}) ; \\
twice &\Rightarrow \text{MKTcomp}(A') ; \text{eval.env}(1) ; \\
A' &\Rightarrow \text{MKTvar}(0) \cdot \text{eval.env}(1) ; \\
\text{fib} &\Rightarrow \text{if } (n < 2) \{ \text{return}(1) \} \text{ else } \{ \text{return}(\text{fib}(n-1) + \text{fib}(n-2)) \} ;
\end{align*}
\]

The initial state of the machine is:

\[
\langle \text{MKEcell}(2) ; \text{MKEp}(A, 1) ; \text{MKEcte}(5, 0) ; \text{Pushfun}(\text{twice}) ; \text{Popenv} ; \text{print}() ; , T, H, O, E \rangle
\]

executing this code using the state transition laws above we have,

\[
\begin{align*}
11 &\Rightarrow \langle \text{MKEp}(A, 1) ; \text{MKEcte}(5, 0) ; \text{Pushfun}(\text{twice}) ; \text{Popenv} ; \text{print}() ; , T, H[d_0 = 1 \circ d_0], O, E \rangle \\
12 &\Rightarrow \langle \text{MKEcte}(5, 0) ; \text{Pushfun}(\text{twice}) ; \text{Popenv} ; \text{print}() ; , T, H[d_0 = A \circ d_0], O, E \rangle \\
13 &\Rightarrow \langle \text{Pushfun}(\text{twice}) ; \text{Popenv} ; \text{print}() ; , T, H[d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
14 &\Rightarrow \langle \text{MKTcomp}(A') ; \text{eval.env}(1) ; \text{Popenv} ; \text{print}() ; , T, H[d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
15 &\Rightarrow \langle \text{eval}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d, T, H[d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
16 &\Rightarrow \langle \text{MKTvar}(0) ; \text{eval.env}(1) ; \text{Pop}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d, T, H[d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
17 &\Rightarrow \langle \text{eval.env}(1) ; \text{Pop}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d_0, d, T, H[d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
18 &\Rightarrow \langle \text{eval}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Pop}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d_0, d, T, H[d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
19 &\Rightarrow \langle \text{eval}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Pop}(0) ; \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d_2, d, T, H[d_2 = 8][d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
20 &\Rightarrow \langle \text{MKT}_1(0, \text{fib}((\ast \text{topT})) \Rightarrow \text{rem.value}) ; \text{Popenv} ; \text{print}() ; , \\
\quad d_2, T, H[d_2 = 8][d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
21 &\Rightarrow \langle \text{Popenv} ; \text{print}() ; , d_2, T, H[d_2 = 34][d_2 = 8][d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
22 &\Rightarrow \langle \text{print}() ; , d_2, T, H[d_2 = 34][d_2 = 8][d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], O, E \rangle \\
23 &\Rightarrow \langle , T, H[d_3 = 34][d_2 = 8][d_1 = (A', c_0)] [d_0 = 5][c_0 = A \circ d_0], 34, O, E \rangle
\]

Compiling Lists

Now we enrich \( \Gamma \)CMC with lists. A new compilation scheme, called \( \mathcal{L} \), is introduced. Some of the previous compilation schemes need to be extended.
Scheme $\mathcal{E}$

6. $\mathcal{E}[] = \text{printf}([]);$
7. $\mathcal{E}[a : b] = \mathcal{E}[a] \mathcal{E}[b] \text{print}();$
8. $\mathcal{E}[\text{Hd}(a \ldots b)] = T^0[a \ldots b] \text{Hd}; \text{printf}(\langle s(\text{topT} - -) \rangle \mapsto \text{rem value});$
9. $\mathcal{E}[\text{Th}(a \ldots b)] = T^3[a \ldots b] \text{Th}; \text{print}();$
10. $\mathcal{E}[\text{Hd}(a : b)] = \mathcal{E}[a]$
11. $\mathcal{E}[\text{Th}(a : b)] = \mathcal{E}[b]$

Scheme $\mathcal{T}$

13. $T^{x_0 \ldots x_j}[\ ] = \text{MKTlv};$
14. $T^{x_0 \ldots x_j}[\text{Hd } x_i] = \text{Ehd}(i);$
15. $T^{x_0 \ldots x_j}[\text{Hd}(a : b)] = T^{x_0 \ldots x_j}[a]$
16. $T^{x_0 \ldots x_j}[\text{Th}(a \ldots b)] = T^{x_0 \ldots x_j}[a \ldots b] \text{Th};$
17. $T^{x_0 \ldots x_j}[a \ (\text{Hd } b \ldots)] = T^{x_0 \ldots x_j}[b \ldots] \text{Hd}; T^{x_0 \ldots x_j}[a]$
18. $T^{x_0 \ldots x_j}[\text{Th } x_i] = \text{Etl}(i);$
19. $T^{x_0 \ldots x_j}[\text{Th}(a : b)] = T^{x_0 \ldots x_j}[b]$
20. $T^{x_0 \ldots x_j}[\text{Th}(a \ldots b)] = T^{x_0 \ldots x_j}[a \ldots b] \text{Th};$
21. $T^{x_0 \ldots x_j}[a \ (\text{Th } b \ldots)] = T^{x_0 \ldots x_j}[b \ldots] \text{Th}; T^{x_0 \ldots x_j}[a]$
22. $T^{x_0 \ldots x_j}[a : b] = T^{x_0 \ldots x_j}[a] T^{x_0 \ldots x_j}[b] \text{MKcons};$
23. $T^{x_0 \ldots x_j}[\text{if } i = [ ] \text{ then } b \text{ else } c] = \text{Enull}(i, l_1); T^{x_0 \ldots x_j}[l_2] \text{ Jmp } l_2; l_2 : T^{x_0 \ldots x_j}[c] l_2 :$

Scheme $\mathcal{G}$

9. $G^{x_0 \ldots x_j}[\ ] j = \text{MKElv}(j);$
10. $G^{x_0 \ldots x_j}[\text{Hd } x_i] j = Ehd(i, j);$
11. $G^{x_0 \ldots x_j}[\text{Hd}(a : b)] j = G^{x_0 \ldots x_j}[a] j$
12. $G^{x_0 \ldots x_j}[\text{Hd}(a \ldots b)] j = G^{x_0 \ldots x_j}[a] j$
13. $G^{x_0 \ldots x_j}[\text{Th } x_i] j = \text{Etl}(i, j);$
14. $G^{x_0 \ldots x_j}[\text{Hd}(a : b)] j = G^{x_0 \ldots x_j}[a] j$
15. $G^{x_0 \ldots x_j}[\text{Th}(a : b)] j = G^{x_0 \ldots x_j}[a] j$
16. $G^{x_0 \ldots x_j}[a : b] j = L^{x_0 \ldots x_j}[a] L^{x_0 \ldots x_j}[b] \text{MKcons1}(j);$

Scheme $T'$

9. $T'^{x_0 \ldots x_j}[\ ] = \text{MKTlv};$
10. $T'^{x_0 \ldots x_j}[\text{Hd } x_i] = \text{Ehd}(i);$
11. $T'^{x_0 \ldots x_j}[\text{Hd}(a : b)] = T^{x_0 \ldots x_j}[a]$
12. $T'^{x_0 \ldots x_j}[\text{Th } x_i] = \text{Etl}(i);$
13. $T'^{x_0 \ldots x_j}[\text{Th}(a : b)] = T^{x_0 \ldots x_j}[b]$
14. $T'^{x_0 \ldots x_j}[a : b] = \text{MKTcomp}(A);$

Scheme $\mathcal{L}$

1. $L^{x_0 \ldots x_j}[x_i] = \text{MKTvar}(i);$
2. $L^{x_0 \ldots x_j}[k] = \text{MKTcte}(k);$
3. $L^{x_0 \ldots x_j}[a + b] = \text{MKTcte}(L^{x_0 \ldots x_j}[a] + L^{x_0 \ldots x_j}[b]);$
4. $L^{x_0 \ldots x_j}[a + b] = \text{MKTcomp}(A);$
5. $L^{x_0 \ldots x_j}[\text{if } a \text{ then } b \text{ else } c] = \text{IFtrue}(L^{x_0 \ldots x_j}[a], B, C)$
where $B \mapsto T^{x_0 \ldots x_j}[b] \quad C \mapsto T^{x_0 \ldots x_j}[c]$
6. $L^{x_0 \ldots x_j}[\text{if } a \text{ then } b \text{ else } c] = \text{MKTcomp}(A);$
7. $L^{x_0 \ldots x_j}[a : b] = L^{x_0 \ldots x_j}[a] L^{x_0 \ldots x_j}[b] \text{MKcons};$
8. $L^{x_0 \ldots x_j}[a + b] = \text{MKTcomp}(A);$

where $A \mapsto T^{x_0 \ldots x_j}[a : b]$

if $a$ and $b$ are evaluated.

if $a$ is evaluated. 

if $a$ is evaluated.
Example of Compilation

Let us present an example of compilation involving lists. If we have the script:

```plaintext
map f x = if x=[] then [] else f(Hd x):map f (Tl x)
fib n = if n<2 then 1 else fib(n-1) + fib(n-2)
twice f x = f (f x)
map (twice fib) (1:(2:[]))
```

it will be compiled as,

```
L[map (twice fib) (1:(2:[]))]
⇒ T^3[map (twice fib) (1:(2:[]))]; print();
⇒ T^3[map f x = if x=[] then [] else f(Hd x):map f (Tl x)]
```
$T^{\phi_0}$: $\text{Null}(0, 1)$; $\text{MKThv}; \text{Jmpl}_{2}; l_{2} : T_{T^{\phi_0}}[f(Hd x)]T_{T^{\phi_0}}[\text{map } f (T l x)]; \text{MKcons}; l_{2} :$

$T^{\phi_1}$: $\text{Null}(0, 1)$; $\text{MKThv}; \text{Jmpl}_{2}; l_{1} : T_{T^{\phi_1}}[x] \text{ Hd}; T_{T^{\phi_1}}[f]T_{T^{\phi_1}}[\text{map } f (T l x)]; \text{MKcons}; l_{2} :$

$T^{\phi_2}$: $\text{Null}(0, 1)$; $\text{MKThv}; \text{Jmpl}_{2}; l_{1} : \text{MKvar}(0); \text{ Hd}; T_{T^{\phi_2}}[f]T_{T^{\phi_2}}[\text{map } f (T l x)]; \text{MKcons}; l_{2} :$

$T^{\phi_3}$: $\text{Null}(0, 1)$; $\text{MKThv}; \text{Jmpl}_{2}; l_{1} : \text{MKvar}(0); \text{ Hd}; \text{eval}(1); T_{T^{\phi_3}}[\text{map } f (T l x)]; \text{MKcons}; l_{2} :$

$T^{\phi_4}$: $\text{Null}(0, 1)$; $\text{MKThv}; \text{Jmpl}_{2}; l_{1} : \text{MKvar}(0); \text{ Hd}; \text{eval}(1); \text{MKcomp}(C); \text{MKcons}; l_{2} :$

where,

\[
C \mapsto T_{T^{\phi_0}}[\text{map } f (T l x)]
\]

$\Rightarrow$ $\text{MKEcell}(2); G_{T^{\phi_0}}[f]^1 G_{T^{\phi_0}}[T l x]0 \text{ Pushfun}(\text{map}); \text{Popenv};$

$\Rightarrow$ $\text{MKEcell}(2); \text{MKEvar}(1, 1); G_{T^{\phi_0}}[T l x]0 \text{ Pushfun}(\text{map}); \text{Popenv};$

$\Rightarrow$ $\text{MKEcell}(2); \text{MKEvar}(1, 1); \text{Etl}(0, 0); \text{Pushfun}(\text{map}); \text{Popenv};$

**New State Transition Laws**

I\CMC with lists also makes use of the following state transition laws:

28. \(\langle \text{print}, c, d, T, H[d = a : b], O, E, H[d = \lambda : b], O, E \rangle \Rightarrow \langle \text{eval} \cdot \text{print}, c, a, b, T, H[d = a : b], O, E \rangle\)

29. \(\langle \text{print}, c, d, T, H[d = \lambda], O, E \rangle \Rightarrow \langle c, \text{ T}, H[d = \lambda], O, E \rangle\)

30. \(\langle \text{Hd}, c, d, T, H[d = a : b], O, E \rangle \Rightarrow \langle c, a, T, H[d = a : b], O, E \rangle\)

31. \(\langle \text{Tl}, c, d, T, H[d = a : b], O, E \rangle \Rightarrow \langle c, b, T, H[d = a : b], O, E \rangle\)

32. \(\langle \text{MKcons}, c, d_{1}, d_{2}, T, H[d_{2} = b][d_{1} = a], O, E \rangle \Rightarrow \langle c, d, T, H[d = d_{2} : d_{1}], O, E \rangle\)

33. \(\langle \text{MKThv}, c, d, T, H[d = \lambda], O, E \rangle \Rightarrow \langle c, d, T, H[d = \lambda], O, E \rangle\)

34. \(\langle \text{MKElv}(i), c, T, H[e_{0} = \ldots w_{i}], O, e_{0}, E \rangle \Rightarrow \langle c, T, H[d = \lambda][e_{0} = \ldots d_{i}], O, e_{0}, E \rangle\)

35. \(\langle \text{MKEcons}(i), c, d_{1}, d_{2}, T, H[d_{1} = a][d_{2} = b][e_{i} = \ldots w_{i}], O, e_{i}, E \rangle \Rightarrow \langle c, d, T, H[d = d_{2} : d_{1}][e_{i} = \ldots d_{i}], O, e_{i}, E \rangle\)

36. \(\langle \text{MKEcons}(i), c, d, T, H[e_{0} = \ldots w_{i}], O, e_{0}, E \rangle \Rightarrow \langle c, d, T, H[d = w_{i}][e_{0} = \ldots w_{i}], O, e_{0}, E \rangle\)

37. \(\langle \text{MKEcons}(i), c, d, T, H[e_{0} = \ldots w_{i}], O, e_{0}, E \rangle \Rightarrow \langle c, d, T, H[d = \lambda], O, e_{0}, E \rangle\)

38. \(\langle \text{End}(i), c, T, H[e_{i} = \ldots a : b], O, e_{i}, E \rangle \Rightarrow \langle c, a, T, H[e_{i} = \ldots a : b], O, e_{i}, E \rangle\)

39. \(\langle \text{Etl}(i), c, T, H[e_{i} = \ldots a : b], O, e_{i}, E \rangle \Rightarrow \langle c, b, T, H[e_{i} = \ldots a : b], O, e_{i}, E \rangle\)

40. \(\langle \text{ETd}(i), c, T, H[e_{0} = \ldots a : b], O, e_{0}, E \rangle \Rightarrow \langle c, a, T, H[e_{0} = \ldots a : b], O, e_{0}, E \rangle\)

41. \(\langle \text{ETd}(i), c, T, H[e_{0} = \ldots a : b], O, e_{0}, E \rangle \Rightarrow \langle c, b, T, H[e_{0} = \ldots a : b], O, e_{0}, E \rangle\)

42. \(\langle \text{End}(1), c, T, H[e_{1} = \ldots x], O, e_{1}, E \rangle \Rightarrow \langle c, x, T, H[e_{1} = \ldots x], O, e_{1}, E \rangle\)

43. \(\langle \text{Etl}(1, i), c, T, H[e_{1} = \ldots x], O, e_{1}, E \rangle \Rightarrow \langle c, x, T, H[e_{1} = \ldots x], O, e_{1}, E \rangle\)

44. \(\langle \text{Null}(n, t), c, T, H[e_{0} = \ldots a : b], O, e_{0}, E \rangle \Rightarrow \langle c, T, H[d_{0} = \lambda][e_{0} = \ldots d_{n}, \ldots], O, e_{0}, E \rangle\)

45. \(\langle \text{Jmp} \ldots l_{x}, c, T, O, E \rangle \Rightarrow \langle c, T, H, O, E \rangle\)

46. \(\langle \text{Jfalse}(\text{False}, l), c, T, H, O, E \rangle \Rightarrow \langle c, T, H, O, E \rangle\)

47. \(\langle \text{Jfalse}(\text{True}, l), c, T, H, O, E \rangle \Rightarrow \langle c, T, H, O, E \rangle\)

**Example of Evaluation**

As we saw in the examples of compilation above, the program:

\[
\text{map } f x = \text{if } x = {} \square \text{ then } {} \square \text{ else } f(/(\text{Hd } x)) : \text{map } f (/T l x)
\]

\[
\text{fib } n = \text{if } n \leq 2 \text{ then } 1 \text{ else } \text{fib}(n-1) + \text{fib}(n-2)
\]

\[
\text{twice } f x = f (f x)
\]

\[
\text{map } (\text{twice } \text{fib}) (1 : (2 : \square))?
\]
compiled as,

\[
\text{map\ (twice\ fib\ (1\ :\ [(])\)} \mapsto \text{MKEcell(2);\ MKEvcomp(B,\ 1);\ MKTcte(1);\ MKTcte(2);\ MKTlv;}
\]

\[
\text{MKcons;\ MKKcons;\ MKKcons(0);\ Pushfun(map);\ Popenv;\ print(1);}
\]

\[
\text{B} \mapsto \text{MKTp(A);\ MKEnv(2);\ Pushfun(twice);\ Popenv;}
\]

\[
\text{map} \mapsto \text{Enull(0,\ 1];\ MKTlv;\ Jmpl(2;\ 1)\ :\ MKTvar(0);\ Hd;\ eval\ Env(1);\ MKTcomp(C);\ MKcons;\ l2;\ }
\]

\[
\text{C} \mapsto \text{MKTenv(2);\ MKEvar(1,\ 1);\ Etl(0,\ 0);\ Pushfun(map);\ Popenv;}
\]

\[
\text{A} \mapsto \text{eval(0);\ MKTK(0,\ fib\ (*\ (topT)) \mapsto \text{remvalue});}
\]

\[
\text{twice} \mapsto \text{MKTcomp(A');\ eval\ Env(1);}
\]

\[
\text{fib} \mapsto \text{if\ (n < 2)\ \{\text{return(1);}\\}\}\text{else\{\text{return(fib(n - 1) + fib(n - 2))}}\}
\]

The initial state of the machine is:

\[
\langle\ \text{MKEcell(2);\ MKEvcomp(B,\ 1);\ MKTcte(1);\ MKTcte(2);\ MKTlv;\ \ldots;\ T,\ H, T, O, E, E} \rangle
\]

executing this code using the state transition laws above we have,

\[
\begin{align*}
\text{\small\text{Pushfun(map);\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d1,\ d2];\ d3 = d1 = d2;\ d2 = [1]} & \mapsto [B,\ c0][c0 = c,\ 0,\ 0];\ c1 = c,\ 0,\ 0,\ E) \\
\text{Enull(0,\ 1];\ MKTlv;\ Jmpl(2,\ 1)\ :\ MKTvar(0);\ Hd;\ eval\ Env(1);\ MKTcomp(C);\ MKcons;\ l2;\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d0,\ d2];\ d3 = d1 = d2;\ d2, d1 = [1]} & \mapsto (B,\ c0)(c0 = c,\ d4);\ c1 = c,\ 0,\ 0,\ E) \\
\text{MKTvar(0);\ Hd;\ eval\ Env(1);\ MKTcomp(C);\ MKcons;\ l2;\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d0,\ d2];\ d3 = d1 = d2;\ d2, d1 = [1]} & \mapsto (B,\ c0)(c0 = c,\ d4);\ c1 = c,\ 0,\ 0,\ E) \\
\text{eval\ Env(1);\ MKTcomp(C);\ MKcons;\ l2;\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d0,\ d2];\ d3 = d1 = d2;\ d2, d1 = [1]} & \mapsto (B,\ c0)(c0 = c,\ d4);\ c1 = c,\ 0,\ 0,\ E) \\
\text{eval\ Env(1);\ MKTcomp(C);\ MKcons;\ l2;\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d0,\ d2];\ d3 = d1 = d2;\ d2, d1 = [1]} & \mapsto (B,\ c0)(c0 = c,\ d4);\ c1 = c,\ 0,\ 0,\ E) \\
\end{align*}
\]

at this point of execution the graph for the list is complete and we enter the code for map.

\[
\begin{align*}
\text{\small\text{Pushfun(map);\ Popenv;\ print(1);\ ;\ d4,\ T,\ H[d4 = d1,\ d2];\ d3 = d1 = d2;\ d2 = [1]} & \mapsto [B,\ c0][c0 = c,\ 0,\ 0];\ c1 = c,\ 0,\ 0,\ E) \\
\end{align*}
\]

Now the code for twice is called taking as arguments fib and 1, which are referenced by the frame on the top of the environment stack. The reduction sequence above gives an idea of how ΓCMC evaluates lists.
Optimisations

A number of code optimisations should be introduced to GCVM in order to obtain a better performance. In this section we present the most important of them.

Sharing

Sharing of computation can bring substantial improvement to the performance of the machine. There is a number of ways sharing can be incorporated to GCVM. Although the authors are still experimenting to know the best possible way, the sharing mechanism implemented at the moment is similar to the one in CMCM [6, 11], which is inspired in the frame update mechanism of TFM [2]. Now, the user provides annotations (U combinator) to specify variables one wants to share. The U combinator performs the following state transition:

48. \( \langle U(i).c.d.T, H[c_0 = \ldots a_i \ldots], O, c_0.E \rangle \Rightarrow \langle c, d.T, H[c_0 = \ldots d \ldots], O, c_0.E \rangle \)

As Categorical Multi-Combinators do not allow for partial applications to be reduced we think of using Partial Categorical Multi-Combinators [5] to deal with sharing of partial applications.

Tail Recursion

Functions over lists recursively defined as

\[ f_n x_0 \ldots x_n = \text{if } a \text{ then } b \text{ else } z : (f_n y_0 \ldots y_n) \]

such as \texttt{map}, are of widespread use in functional programs. The compilation schemes we have generate an environment every time we make a recursive call and discard the environment used for the previous call. To increase the performance of GCVM we avoid garbage generation by compiling tail recursive functions as

\[ [f_n x_0 \ldots x_n = \text{if } x_i = [ ] \text{ then } b \text{ else } z : (f_n y_0 \ldots y_n)] \]

by the following entries in the script:

\[
\begin{align*}
  f_n & \mapsto \text{Enull}(i, l_1) ; T^{x_0 \ldots x_n}[b] ; \text{Jmp} l_2 ; l_1 : T^{x_0 \ldots x_n}[z] \ \text{MKTcomp}(A) ; \text{MKenv}(n + 1) ; \text{MKcons} ; l_2 : \\
  A & \mapsto G^{x_0 \ldots x_n}[y_0] n \ldots G^{x_0 \ldots x_n}[y_0] 0 \ \text{Swap} ; \text{Pushfun}(f_n) ; \text{Popenv} ; \\
  f_n & \mapsto \text{Enull}(i, l_1) ; T^{x_0 \ldots x_n}[b] ; \text{Jmp} l_2 ; l_1 : T^{x_0 \ldots x_n}[z] \ \text{MKTcomp}(A) ; \text{MKcons} ; l_2 : \\
\end{align*}
\]

The state transition law for \texttt{Swap} is:

49. \( \langle \text{Swap}, c, T, H, O, c_1, E \rangle \Rightarrow \langle c, A.T, H, O, c_0, c_1, E \rangle \)

Recursive Functions

Recursion is fundamental for functional programming languages. Many functions are not special thus can not benefit from the very efficient handling of recursion made by the C compiler, which takes advantage of the fast context switching mechanism of RISC architectures. Better performance can be obtained if we introduce a stack to handle recursion. Thus we translate functions defined as

\[ f_n x_0 \ldots x_n = \text{if } a < b \text{ then } c \text{ else } f_n y_0 \ldots y_n \]

by using the following scheme:

\[
\begin{align*}
  f_n & \mapsto \text{pushR}(0) ; \\
  \text{LP1} & : \text{false}(T^{x_0 \ldots x_n}[a]) \rightarrow \text{rem.value} ; \\
  T^{x_0 \ldots x_n}[b] & ; t1 = (*topT--) \rightarrow \text{rem.value} ; t1 > t2, l_1) ; T^{x_0 \ldots x_n}[c] ; \\
  \text{Jmp} l_1 ; \\
  l_1 : & \text{MKEcell}(n + 1) ; \\
  G^{x_0 \ldots x_n}[y_0] n & ; \\
  \vdots \\
  G^{x_0 \ldots x_n}[y_0] 0 & \text{pushR}(1) ; \text{Jmp}(\text{LP1}) ; \\
  \text{LP2} & : \text{Popenv} ; \\
  l_2 : & \text{if}(*\text{topR}--) == 1) \ \text{Jmp}(\text{LP2}) ; \\
\end{align*}
\]
Avoiding Indirections

Functions which take only one parameter are frequent. Because we adopted a fully-boxed representation the cell which represents the environment of a function to one argument works as an indirection cell. One can avoid the generation of this indirection cell by making the environment stack point directly to its argument. New operators are needed for this optimisation. For a matter of simplicity we will call them as before suffixed by 0. For instance, instruction Ehd(i) becomes Ehd0.

Monomorphic Print

Instead of having a general (polymorphic) printing routine, which at run-time tests the data produced to output it, we use information provided by the type-checker to choose statically which printing function is suitable for printing the output.

Performance

In this section we present the performance figures obtained for the benchmark programs below running on a SUN Sparcstation II under UNIX.

Fib 30: the Fibonacci number of 30

Rev: reverse reverse reverse of a list of 300 numbers.

Sieve: generates a list of prime numbers smaller than 2000 by using Erathosthenes’ sieve.

Insord: sorting by insertion of a list of 600 random numbers.

Simlog: takes a list of 4000 random numbers and produces 4000 boolean values.

Map: maps (twice twice twice successor) on a list of 2000 integers.

Tak: Takeuchi function of 30 25 15.

<table>
<thead>
<tr>
<th>Prog</th>
<th>Fib 30</th>
<th>Rev</th>
<th>Sieve</th>
<th>Insord</th>
<th>Simlog</th>
<th>Map</th>
<th>Tak</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM-C</td>
<td>10.6</td>
<td>—</td>
<td>2.8</td>
<td>5.0</td>
<td>1.6</td>
<td>2.2</td>
<td>2.382</td>
</tr>
<tr>
<td>ΓCMC</td>
<td>1.5</td>
<td>2.5</td>
<td>1.4</td>
<td>2.9</td>
<td>0.2</td>
<td>0.5</td>
<td>2.8</td>
</tr>
<tr>
<td>LML</td>
<td>8.9</td>
<td>4.0</td>
<td>1.6</td>
<td>2.5</td>
<td>0.6</td>
<td>0.4</td>
<td>42.8</td>
</tr>
</tbody>
</table>

GM-C corresponds to the last version of GM-C [8] done by Musicante and Lins. ΓCMC refers to our best implementation of ΓCMC. LML presents the performance of the Chalmers Lazy ML compiler version 0.99.3.

As we can observe from the table above the performance of ΓCMC is far better than GM-C, for all our benchmark programs. ΓCMC presented a performance close to LML in the benchmark programs which made intensive use of higher-order-functions and lazy evaluation. In the case of the use of strict functions under recursion ΓCMC performed far better than LML.

Further Work & Conclusions

The authors are currently working on a front-end for ΓCMC and also analysing a number of optimisations to the back-end. We believe the benchmark figures we presented are representative of the performance of ΓCMC and that they scale-up for larger programs.

In our opinion, ΓCMC has already shown that transferring the flow of execution from a higher-level abstract machine to C, by using procedure calls, brings not only portability but also efficiency on RISC architectures. The performance of ΓCMC is in the worst case as good as the Chalmers LML compiler. In the best case, ΓCMC can run several times faster than LML.
Acknowledgements

We express gratitude for several discussions with Ricardo Massa and Genésio Neto. Research reported herein has been sponsored jointly by The British Council, CNPq (Brazil) grants 40.9110/88-4, 46.0782/89.4, and 80.4520/88-7, and CAPES (Brazil) grant 2487/91-08.

References