Ravi Programming Language Documentation

Release 0.1

Dibyendu Majumdar

January 13, 2017

Contents

Contents:

CHAPTER 1

Ravi Programming Language

Ravi is a derivative/dialect of Lua 5.3 with limited optional static typing and an LLVM powered JIT compiler. The name Ravi comes from the Sanskrit word for the Sun. Interestingly a precursor to Lua was Sol which had support for static types; Sol means the Sun in Portugese.

Lua is perfect as a small embeddable dynamic language so why a derivative? Ravi extends Lua with static typing for greater performance under JIT compilation. However, the static typing is optional and therefore Lua programs are also valid Ravi programs.

There are other attempts to add static typing to Lua - e.g. Typed Lua but these efforts are mostly about adding static type checks in the language while leaving the VM unmodified. The Typed Lua effort is very similar to the approach taken by Typescript in the JavaScript world. The static typing is to aid programming in the large - the code is eventually translated to standard Lua and executed in the unmodified Lua VM.

My motivation is somewhat different - I want to enhance the VM to support more efficient operations when types are known. Type information can be exploited by JIT compilation technology to improve performance. At the same time, I want to keep the language safe and therefore usable by non-expert programmers.

Of course there is also the fantastic LuaJIT implementation. Ravi has a different goal compared to LuaJIT. Ravi prioritizes ease of maintenance and support, language safety, and compatibility with Lua 5.3, over maximum performance. For more detailed comparison please refer to the documentation links below.

Table of Contents

1.1 Features

- Optional static typing
- Type specific bytecodes to improve performance
- Compatibility with Lua 5.3 (see Compatibility section below)
- LLVM powered JIT compiler
- Additionally a libgccjit based alternative JIT compiler is also available
- LLVM bindings exposed in Lua

1.2 Documentation

See Ravi Documentation. As more stuff is built I will keep updating the documentation so please revisit for latest information.

Also see the slides I presented at the Lua 2015 Workshop.

1.3 JIT Implementation

The LLVM JIT compiler is functional. The Lua and Ravi bytecodes currently implemented in LLVM are described in JIT Status page.

Ravi also provides an LLVM binding; this is still work in progress so please check documentation for the latest status.

There is also a libgccjit based JIT implementation but this implementation is lagging behind the LLVM based implementation. Further development of this is currently not planned.

1.4 Ravi Extensions to Lua 5.3

1.4.1 Optional Static Typing

Ravi allows you to annotate `local` variables and function parameters with static types. The supported types and the resulting behaviour are as follows:

`integer` denotes an integral value of 64-bits.

`number` denotes a double (floating point) value of 8 bytes.

`integer[]` denotes an array of integers

`number[]` denotes an array of numbers

`table` a Lua table

Declaring the types of `local` variables and function parameters has following advantages.

- `integer` and `number` types are automatically initialized to zero
- Arithmetic operations on numeric types make use of type specific bytecodes which leads to more efficient JIT compilation
- Specialised operators to get/set from array types are implemented; this makes array access more efficient in JIT mode as the access can be inlined
- Declared tables allow specialized opcodes for table gets involving integer and short literal string keys; these opcodes result in more efficient JIT code
- Values assigned to typed variables are checked statically when possible; if the values are results from a function call then runtime type checking is performed
- The standard table operations on arrays are checked to ensure that the type is not subverted
- Even if a typed variable is captured in a closure its type must be respected
- When function parameters are decorated with types, Ravi performs an implicit coersion of those parameters to the required types. If the coersion fails a runtime error occurs.

The array types (`number[]` and `integer[]`) are specializations of Lua table with some additional special behaviour:

- Array types are not compatible with declared table variables, i.e. following is not allowed:

```
local t: table = {}
local t2: number[] = t  -- error!

local t3: number[] = {}
local t4: table = t3    -- error!
```

  But following is okay:

```
local t5: number[] = {}
local t6 = t5           -- t6 treated as table
```

  The reason for these restrictions is that declared table types generate optimized JIT code which assumes that the keys are integers or literal short strings. Similarly declared array types result in specialized JIT code that assume integer keys and numeric values. The generated JIT code would be incorrect if the types were not as expected.

- Indices >= 1 should be used when accessing array elements. Ravi arrays (and slices) have a hidden slot at index 0 for performance reasons, but this is not visible in `pairs()` or `ipairs()`, or when initializing an array using a literal initializer; only direct access via the `[]` operator can see this slot.

- Arrays must always be initialized:

```
local t: number[] = {} -- okay
local t2: number[]     -- error!
```

  This restriction is placed as otherwise the JIT code would need to insert tests to validate that the variable is not nil.

- An array will grow automatically (unless the array was created as fixed length using `table.intarray()` or `table.numarray()`) if the user sets the element just past the array length:

```
local t: number[] = {} -- dynamic array
t[1] = 4.2             -- okay, array grows by 1
t[5] = 2.4             -- error! as attempt to set value
```

- It is an error to attempt to set an element that is beyond len+1 on dynamic arrays; for fixed length arrays attempting to set elements at positions greater than len will cause an error.

- The current used length of the array is recorded and returned by len operations

- The array only permits the right type of value to be assigned (this is also checked at runtime to allow compatibility with Lua)

- Accessing out of bounds elements will cause an error, except for setting the len+1 element on dynamic arrays

- It is possible to pass arrays to functions and return arrays from functions. Arrays passed to functions appear as Lua tables inside those functions if the parameters are untyped - however the tables will still be subject to restrictions as above. If the parameters are typed then the arrays will be recognized at compile time:

```
local function f(a, b: integer[], c)
  -- Here a is dynamic type
  -- b is declared as integer[]
  -- c is also a dynamic type
  b[1] = a[1] -- Okay only if a is actually also integer[]
  b[1] = c[1] -- Will fail if c[1] cannot be converted to an integer
end

local a : integer[] = {1}
local b : integer[] = {}
local c = {1}
```

```
    f(a,b,c)        -- ok as c[1] is integer
    f(a,b, {'hi'})  -- error!
```

- Arrays returned from functions can be stored into appropriately typed local variables - there is validation that the types match:

```
    local t: number[] = f() -- type will be checked at runtime
```

- Operations on array types can be optimised to special bytecode and JIT only when the array type is statically known. Otherwise regular table access will be used subject to runtime checks.
- Array types ignore `__index`, `__newindex` and `__len` metamethods.
- Array types cannot be set as metatables for other values.
- `pairs()` and `ipairs()` work on arrays as normal
- There is no way to delete an array element.
- The array data is stored in contiguous memory just like native C arrays; morever the garbage collector does not scan the array data

A declared table (as shown below) has some additional nuances:

```
local t: table = {}
```

- Like array types, a variable of `table` type must be initialized
- Array types are not compatible with declared table variables, i.e. following is not allowed:

```
    local t: table = {}
    local t2: number[] = t -- error!
```

- When short string literals are used to access a table element, specialized bytecodes are generated that are more efficiently JIT compiled:

```
    local t: table = { name='dibyendu'}
    print(t.name) -- The GETTABLE opcode is specialized in this case
```

- As with array types, specialized bytecodes are generated when integer keys are used

Following library functions allow creation of array types of defined length.

`table.intarray(num_elements, initial_value)` creates an integer array of specified size, and initializes with initial value. The return type is integer[]. The size of the array cannot be changed dynamically, i.e. it is fixed to the initial specified size. This allows slices to be created on such arrays.

`table.numarray(num_elements, initial_value)` creates an number array of specified size, and initializes with initial value. The return type is number[]. The size of the array cannot be changed dynamically, i.e. it is fixed to the initial specified size. This allows slices to be created on such arrays.

1.4.2 Type Assertions

Ravi does not support defining new types, or structured types based on tables. This creates some practical issues when dynamic types are mixed with static types. For example:

```
local t = { 1,2,3 }
local i: integer = t[1] -- generates an error
```

Above code generates an error as the compiler does not know that the value in `t[1]` is an integer. However often we as programmers know the type that is expected and it would be nice to be able to tell the compiler what the expected

type of `t[1]` is above. To enable this Ravi supports type assertion operators. A type assertion is introduced by the '`@`' symbol, which must be followed by the type name. So we can rewrite the above example as:

```
local t = { 1,2,3 }
local i: integer = @integer( t[1] )
```

The type assertion operator is a unary operator and binds to the expression following the operator. We use the parenthesis above to enure that the type assertion is applied to `t[1]` rather than `t`. More examples are shown below:

```
local a: number[] = @number[] { 1,2,3 }
local t = { @number[] { 4,5,6 }, @integer[] { 6,7,8 } }
local a1: number[] = @number[]( t[1] )
local a2: integer[] = @integer[]( t[2] )
```

For a real example of how type assertions can be used, please have a look at the test program gaussian2.lua

1.4.3 Array Slices

Since release 0.6 Ravi supports array slices. An array slice allows a portion of a Ravi array to be treated as if it is an array - this allows efficient access to the underlying array elements. Following new functions are available:

`table.slice(array, start_index, num_elements)` creates a slice from an existing *fixed size* array - allowing efficient access to the underlying array elements.

Slices access the memory of the underlying array; hence a slice can only be created on fixed size arrays (constructed by `table.numarray()` or `table.intarray()`). This ensures that the array memory cannot be reallocated while a slice is referring to it. Ravi does not track the slices that refer to arrays - slices get garbage collected as normal.

Slices cannot extend the array size for the same reasons above.

The type of a slice is the same as that of the underlying array - hence slices get the same optimized JIT operations for array access.

Each slice holds an internal reference to the underlying array to ensure that the garbage collector does not reclaim the array while there are slices pointing to it.

For an example use of slices please see the matmul1.ravi benchmark program in the repository. Note that this feature is highly experimental and not very well tested.

1.4.4 Examples

Example of code that works - you can copy this to the command line input:

```
function tryme()
  local i,j = 5,6
  return i,j
end
local i:integer, j:integer = tryme(); print(i+j)
```

When values from a function call are assigned to a typed variable, an implicit type coersion takes place. In above example an error would occur if the function returned values that could not converted to integers.

In the following example, the parameter `j` is defined as a `number`, hence it is an error to pass a value that cannot be converted to a `number`:

```
function tryme(j: number)
  for i=1,1000000000 do
    j = j+1
  end
```

```
  return j
end
print(tryme(0.0))
```

An example with arrays:

```
function tryme()
  local a : number[], j:number = {}
  for i=1,10 do
    a[i] = i
    j = j + a[i]
  end
  return j
end
print(tryme())
```

Another example using arrays. Here the function receives a parameter `arr` of type `number[]` - it would be an error to pass any other type to the function because only `number[]` types can be converted to `number[]` types:

```
function sum(arr: number[])
  local n: number = 0.0
  for i = 1,#arr do
    n = n + arr[i]
  end
  return n
end

print(sum(table.numarray(10, 2.0)))
```

The `table.numarray(n, initial_value)` creates a `number[]` of specified size and initializes the array with the given initial value.

1.4.5 All type checks are at runtime

To keep with Lua's dynamic nature Ravi uses a mix of compile type checking and runtime type checks. However due to the dynamic nature of Lua, compilation happens at runtime anyway so effectually all checks are at runtime.

1.4.6 JIT API

The LLVM based JIT compiler is functional. There are two modes of JIT compilation.

auto mode in this mode the compiler decides when to compile a Lua function. The current implementation is very simple - any Lua function call is checked to see if the bytecodes contained in it can be compiled. If this is true then the function is compiled provided either a) function has a fornum loop, or b) it is largish (greater than 150 bytecodes) or c) it is being executed many times (> 50). Because of the simplistic behaviour performance the benefit of JIT compilation is only available if the JIT compiled functions will be executed many times so that the cost of JIT compilation can be amortized.

manual mode in this mode user must explicitly request compilation. This is the default mode. This mode is suitable for library developers who can pre compile the functions in library module table.

A JIT api is available with following functions:

`ravi.jit([b])` returns enabled setting of JIT compiler; also enables/disables the JIT compiler; defaults to true

`ravi.auto([b [, min_size [, min_executions]]])` returns setting of auto compilation and compilation thresholds; also sets the new settings if values are supplied; defaults are false, 150, 50.

ravi.compile(func_or_table[, options]) compiles a Lua function (or functions if a table is supplied) if possible, returns `true` if compilation was successful for at least one function. `options` is an optional table with compilation options - in particular `omitArrayGetRangeCheck` - which disables range checks in array get operations to improve performance in some cases. Note that at present if the first argument is a table of functions and has more than 100 functions then only the first 100 will be compiled. You can invoke compile() repeatedly on the table until it returns false. Each invocation leads to a new module being created; any functions already compiled are skipped.

ravi.iscompiled(func) returns the JIT status of a function

ravi.dumplua(func) dumps the Lua bytecode of the function

ravi.dumpir(func) dumps the IR of the compiled function (only if function was compiled; only LLVM version)

ravi.dumpasm(func) dumps the machine code using the currently set optimization level (only if function was compiled; only LLVM)

ravi.optlevel([n]) sets LLVM optimization level (0, 1, 2, 3); defaults to 2

ravi.sizelevel([n]) sets LLVM size level (0, 1, 2); defaults to 0

ravi.tracehook([b]) Enables support for line hooks via the debug api. Note that enabling this option will result in inefficient JIT as a call to a C function will be inserted at beginning of every Lua bytecode boundary; use this option only when you want to use the debug api to step through code line by line

1.5 Performance

For performance benchmarks please visit the Ravi Performance Benchmarks page.

To obtain the best possible performance, types must be annotated so that Ravi's JIT compiler can generate efficient code. Additionally function calls are expensive - as the JIT compiler cannot inline function calls, all function calls go via the Lua call protocol which has a large overhead. This is true for both Lua functions and C functions. For best performance avoid function calls inside loops.

1.6 Compatibility with Lua

Ravi should be able to run all Lua 5.3 programs in interpreted mode, but following should be noted:

- Ravi supports optional typing and enhanced types such as arrays (described above). Programs using these features cannot be run by standard Lua. However all types in Ravi can be passed to Lua functions; operations on Ravi arrays within Lua code will be subject to restrictions as described in the section above on arrays.
- Values crossing from Lua to Ravi will be subjected to typechecks should these values be assigned to typed variables.
- Upvalues cannot subvert the static typing of local variables (issue #26) when types are annotated.
- Certain Lua limits are reduced due to changed byte code structure. These are described below.

Limit name	Lua value	Ravi value
MAXUPVAL	255	125
LUAI_MAXCCALLS	200	125
MAXREGS	255	125
MAXVARS	200	125
MAXARGLINE	250	120

When JIT compilation is enabled there are following additional constraints:

- Ravi will only execute JITed code from the main Lua thread; any secondary threads (coroutines) execute in interpreter mode.
- In JITed code tailcalls are implemented as regular calls so unlike the interpreter VM which supports infinite tail recursion JIT compiled code only supports tail recursion to a depth of about 110 (issue #17)

1.7 Building Ravi

1.7.1 Build Dependencies

- CMake

Ravi can be built with or without LLVM. Following versions of LLVM work with Ravi.

- LLVM 3.7 or 3.8 or 3.9
- LLVM 3.5 and 3.6 should also work but have not been recently tested

Unless otherwise noted the instructions below should work for LLVM 3.7 or later.

1.7.2 Building LLVM on Windows

I built LLVM from source. I used the following sequence from the VS2015 command window:

```
cd \github\llvm
mkdir build
cd build
cmake -DCMAKE_INSTALL_PREFIX=c:\LLVM -DLLVM_TARGETS_TO_BUILD="X86" -G "Visual Studio 14 Win64" ..
```

I then opened the generated solution in VS2015 and performed a INSTALL build from there. Above will build the 64-bit version of LLVM libraries. To build a 32-bit version omit the `Win64` parameter.

Note: Note that if you perform a Release build of LLVM then you will also need to do a Release build of Ravi otherwise you will get link errors.

1.7.3 Building LLVM on Ubuntu

On Ubuntu I found that the official LLVM distributions don't work with CMake. The CMake config files appear to be broken. So I ended up downloading and building LLVM from source and that worked. The approach is similar to that described for MAC OS X below.

1.7.4 Building LLVM on MAC OS X

I am using Max OSX El Capitan. Pre-requisites are XCode 7.x and CMake. Ensure cmake is on the path. Assuming that LLVM source has been extracted to `$HOME/llvm-3.7.0.src` I follow these steps:

```
cd llvm-3.7.0.src
mkdir build
cd build
cmake -DCMAKE_BUILD_TYPE=Release -DCMAKE_INSTALL_PREFIX=$HOME/LLVM -DLLVM_TARGETS_TO_BUILD="X86" ..
make install
```

1.7.5 Building Ravi with JIT enabled

I am developing Ravi using Visual Studio 2015 Community Edition on Windows 10 64bit, gcc on Unbuntu 64-bit, and clang/Xcode on MAC OS X. I was also able to successfully build a Ubuntu version on Windows 10 using the newly released Ubuntu/Linux sub-system for Windows 10.

Note: Location of cmake files has moved in LLVM 3.9; the new path is `$LLVM_INSTALL_DIR/lib/cmake/llvm`.

Assuming that LLVM has been installed as described above, then on Windows I invoke the cmake config as follows:

```
cd build
cmake -DLLVM_JIT=ON -DCMAKE_INSTALL_PREFIX=c:\ravi -DLLVM_DIR=c:\LLVM37\share\llvm\cmake -G "Visual S
```

I then open the solution in VS2015 and do a build from there.

On Ubuntu I use:

```
cd build
cmake -DLLVM_JIT=ON -DCMAKE_INSTALL_PREFIX=$HOME/ravi -DLLVM_DIR=$HOME/LLVM/share/llvm/cmake -DCMAKE_
make
```

Note that on a clean install of Ubuntu 15.10 I had to install following packages:

- cmake
- git
- libreadline-dev

On MAC OS X I use:

```
cd build
cmake -DLLVM_JIT=ON -DCMAKE_INSTALL_PREFIX=$HOME/ravi -DLLVM_DIR=$HOME/LLVM/share/llvm/cmake -DCMAKE_
```

I open the generated project in Xcode and do a build from there. You can also use the command line build tools if you wish - generate the make files in the same way as for Linux.

1.7.6 Building without JIT

You can omit `-DLLVM_JIT=ON` option above to build Ravi with a null JIT implementation.

1.7.7 Building Static Libraries

By default the build generates a shared library for Ravi. You can choose to create a static library and statically linked executables by supplying the argument `-DSTATIC_BUILD=ON` to CMake.

1.7.8 Build Artifacts

The Ravi build creates a shared or static depending upon options supplied to CMake, the Ravi executable and some test programs. Additionally when JIT compilation is switched off, the `ravidebug` executable is generated which is the debug adapter for use by Visual Studio Code.

The `ravi` command recognizes following environment variables. Note that these are only for internal debugging purposes.

RAVI_DEBUG_EXPR if set to a value this triggers debug output of expression parsing

RAVI_DEBUG_CODEGEN if set to a value this triggers a dump of the code being generated

RAVI_DEBUG_VARS if set this triggers a dump of local variables construction and destruction

Also see section above on available API for dumping either Lua bytecode or LLVM IR for compiled code.

1.7.9 Testing

I test the build by running a modified version of Lua 5.3.3 test suite. These tests are located in the `lua-tests` folder. Additionally I have ravi specific tests in the `ravi-tests` folder. There is a also a travis build that occurs upon commits - this build runs the tests as well.

Note: To thoroughly test changes, you need to invoke CMake with `-DCMAKE_BUILD_TYPE=Debug` option. This turns on assertions, memory checking, and also enables an internal module used by Lua tests.

1.8 Roadmap

- 2015 - Implemented JIT compilation using LLVM
- 2015 - Implemented libgccjit based alternative JIT
- 2016 - Implemented debugger for Ravi and Lua 5.3 for Visual Studio Code
- 2017 - Main priorities are:
  - I would like Ravi to be backward compatible with Lua 5.1 and 5.2 as far as possible
  - Lua function inlining
  - Improve performance of Ravi
  - Additional type annotations

1.9 License

MIT License for LLVM version.

CHAPTER 2

Lua 5.3 Bytecode Reference

This is my attempt to bring up to date the Lua bytecode reference. Note that this is work in progress. Following copyrights are acknowledged:

```
A No-Frills Introduction to Lua 5.1 VM Instructions
  by Kein-Hong Man, esq. <khman AT users.sf.net>
  Version 0.1, 2006-03-13
```

A No-Frills Introduction to Lua 5.1 VM Instructions is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike License 2.0. You are free to copy, distribute and display the work, and make derivative works as long as you give the original author credit, you do not use this work for commercial purposes, and if you alter, transform, or build upon this work, you distribute the resulting work only under a license identical to this one. See the following URLs for more information:

```
http://creativecommons.org/licenses/by-nc-sa/2.0/
http://creativecommons.org/licenses/by-nc-sa/2.0/legalcode
```

2.1 Lua Stack and Registers

Lua employs two stacks. The `Callinfo` stack tracks activation frames. There is the secondary stack `L->stack` that is an array of `TValue` objects. The `Callinfo` objects index into this array. Registers are basically slots in the `L->stack` array.

When a function is called - the stack is setup as follows:

```
stack
|             function reference
|             var arg 1
|             ...
|             var arg n
| base->      fixed arg 1
|             ...
|             fixed arg n
|             local 1
|             ...
|             local n
|             temporaries
|             ...
|  top->
|
V
```

So `top` is just past the registers needed by the function. The number of registers is determined based on parameters, locals and temporaries.

For each Lua function, the `base` of the stack is set to the first fixed parameter or local. All register addressing is done as offset from `base` - so `R(0)` is at `base+0` on the stack.

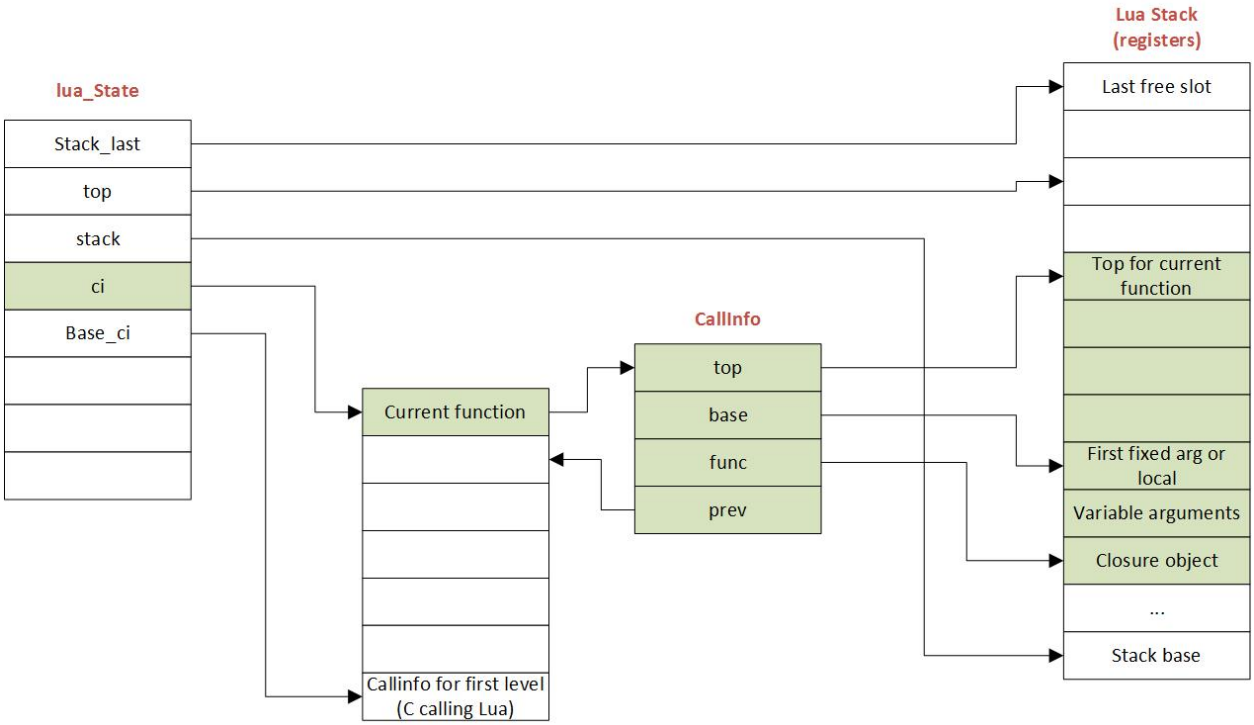

Fig. 2.1: The figure above shows how the stack is related to other Lua objects.

2.2 Instruction Notation

R(A) Register A (specified in instruction field A)

R(B) Register B (specified in instruction field B)

R(C) Register C (specified in instruction field C)

PC Program Counter

Kst(n) Element n in the constant list

Upvalue[n] Name of upvalue with index n

Gbl[sym] Global variable indexed by symbol sym

RK(B) Register B or a constant index

RK(C) Register C or a constant index

sBx Signed displacement (in field sBx) for all kinds of jumps

2.3 Instruction Summary

Lua bytecode instructions are 32-bits in size. All instructions have an opcode in the first 6 bits. Instructions can have the following fields:

```
'A' : 8 bits
'B' : 9 bits
'C' : 9 bits
'Ax' : 26 bits ('A', 'B', and 'C' together)
'Bx' : 18 bits ('B' and 'C' together)
'sBx' : signed Bx
```

A signed argument is represented in excess K; that is, the number value is the unsigned value minus K. K is exactly the maximum value for that argument (so that -max is represented by 0, and +max is represented by 2*max), which is half the maximum for the corresponding unsigned argument.

Note that B and C operands need to have an extra bit compared to A. This is because B and A can reference registers or constants, and the extra bit is used to decide which one. But A always references registers so it doesn't need the extra bit.

Opcode	Description
MOVE	Copy a value between registers
LOADK	Load a constant into a register
LOADKX	Load a constant into a register
LOADBOOL	Load a boolean into a register
LOADNIL	Load nil values into a range of registers
GETUPVAL	Read an upvalue into a register
GETTABUP	Read a value from table in up-value into a register
GETTABLE	Read a table element into a register
SETTABUP	Write a register value into table in up-value
SETUPVAL	Write a register value into an upvalue
SETTABLE	Write a register value into a table element
NEWTABLE	Create a new table
SELF	Prepare an object method for calling
ADD	Addition operator
SUB	Subtraction operator
MUL	Multiplication operator
MOD	Modulus (remainder) operator
POW	Exponentation operator
DIV	Division operator
IDIV	Integer division operator
BAND	Bit-wise AND operator
BOR	Bit-wise OR operator
BXOR	Bit-wise Exclusive OR operator
SHL	Shift bits left
SHR	Shift bits right
UNM	Unary minus
BNOT	Bit-wise NOT operator
NOT	Logical NOT operator
LEN	Length operator
CONCAT	Concatenate a range of registers
JMP	Unconditional jump
EQ	Equality test, with conditional jump

Continued on next page

Table 2.1 – continued from previous page

Opcode	Description
LT	Less than test, with conditional jump
LE	Less than or equal to test, with conditional jump
TEST	Boolean test, with conditional jump
TESTSET	Boolean test, with conditional jump and assignment
CALL	Call a closure
TAILCALL	Perform a tail call
RETURN	Return from function call
FORLOOP	Iterate a numeric for loop
FORPREP	Initialization for a numeric for loop
TFORLOOP	Iterate a generic for loop
TFORCALL	Initialization for a generic for loop
SETLIST	Set a range of array elements for a table
CLOSURE	Create a closure of a function prototype
VARARG	Assign vararg function arguments to registers

2.4 OP_CALL instruction

2.4.1 Syntax

```
CALL A B C    R(A), ... ,R(A+C-2) := R(A)(R(A+1), ... ,R(A+B-1))
```

2.4.2 Description

Performs a function call, with register R(A) holding the reference to the function object to be called. Parameters to the function are placed in the registers following R(A). If B is 1, the function has no parameters. If B is 2 or more, there are (B-1) parameters. If B >= 2, then upon entry to the called function, R(A+1) will become the `base`.

If B is 0, then B = 'top', i.e., the function parameters range from R(A+1) to the top of the stack. This form is used when the number of parameters to pass is set by the previous VM instruction, which has to be one of `OP_CALL` or `OP_VARARG`.

If C is 1, no return results are saved. If C is 2 or more, (C-1) return values are saved. If C == 0, then 'top' is set to last_result+1, so that the next open instruction (`OP_CALL`, `OP_RETURN`, `OP_SETLIST`) can use 'top'.

2.4.3 Examples

Example of `OP_VARARG` followed by `OP_CALL`:

```
function y(...) print(...) end

1 [1] GETTABUP  0 0 -1  ; _ENV "print"
2 [1] VARARG    1 0     ; VARARG will set L->top
3 [1] CALL      0 0 1   ; B=0 so L->top set by previous instruction
4 [1] RETURN    0 1
```

Example of `OP_CALL` followed by `OP_CALL`:

```
function z1() y(x()) end

1 [1] GETTABUP  0 0 -1  ; _ENV "y"
```

```
2 [1] GETTABUP  1 0 -2  ; _ENV "x"
3 [1] CALL      1 1 0   ; C=0 so return values indicated by L->top
4 [1] CALL      0 0 1   ; B=0 so L->top set by previous instruction
5 [1] RETURN    0 1
```

Thus upon entry to a function `base` is always the location of the first fixed parameter if any or else `local` if any. The three possibilities are shown below.

```
                                    Two variable args and 1     Two variable args and no
Caller   One fixed arg              fixed arg                   fixed args
R(A)     CI->func  [ function    ]  CI->func  [ function    ]   CI->func [ function   ]
R(A+1)   CI->base  [ fixed arg 1 ]            [ var arg 1   ]            [ var arg 1  ]
R(A+2)             [ local 1     ]            [ var arg 2   ]            [ var arg 2  ]
R(A+3)                              CI->base  [ fixed arg 1 ]   CI->base [ local 1    ]
R(A+4)                                        [ local 1     ]
```

Results returned by the function call are placed in a range of registers starting from R(A). If C is 1, no return results are saved. If C is 2 or more, (C-1) return values are saved. If C is 0, then multiple return results are saved. In this case the number of values to save is determined by one of following ways:

- A C function returns an integer value indicating number of results returned so for C function calls this is used (see the value of `n` passed to luaD_poscall() in luaD_precall())
- For Lua functions, the results are saved by the called function's `OP_RETURN` instruction.

2.4.4 More examples

```
x=function() y() end
```

Produces:

```
function <stdin:1,1> (3 instructions at 000000CECB2BE040)
0 params, 2 slots, 1 upvalue, 0 locals, 1 constant, 0 functions
  1       [1]     GETTABUP        0 0 -1  ; _ENV "y"
  2       [1]     CALL            0 1 1
  3       [1]     RETURN          0 1
constants (1) for 000000CECB2BE040:
  1       "y"
locals (0) for 000000CECB2BE040:
upvalues (1) for 000000CECB2BE040:
  0       _ENV    0       0
```

In line [2], the call has zero parameters (field B is 1), zero results are retained (field C is 1), while register 0 temporarily holds the reference to the function object from global y. Next we see a function call with multiple parameters or arguments:

```
x=function() z(1,2,3) end
```

Generates:

```
function <stdin:1,1> (6 instructions at 000000CECB2D7BC0)
0 params, 4 slots, 1 upvalue, 0 locals, 4 constants, 0 functions
  1       [1]     GETTABUP        0 0 -1  ; _ENV "z"
  2       [1]     LOADK           1 -2    ; 1
  3       [1]     LOADK           2 -3    ; 2
  4       [1]     LOADK           3 -4    ; 3
  5       [1]     CALL            0 4 1
  6       [1]     RETURN          0 1
```

```
constants (4) for 000000CECB2D7BC0:
  1       "z"
  2       1
  3       2
  4       3
locals (0) for 000000CECB2D7BC0:
upvalues (1) for 000000CECB2D7BC0:
  0       _ENV    0       0
```

Lines [1] to [4] loads the function reference and the arguments in order, then line [5] makes the call with an operand B value of 4, which means there are 3 parameters. Since the call statement is not assigned to anything, no return results need to be retained, hence field C is 1. Here is an example that uses multiple parameters and multiple return values:

```
x=function() local p,q,r,s = z(y()) end
```

Produces:

```
function <stdin:1,1> (5 instructions at 000000CECB2D6CC0)
0 params, 4 slots, 1 upvalue, 4 locals, 2 constants, 0 functions
  1       [1]     GETTABUP        0 0 -1  ; _ENV "z"
  2       [1]     GETTABUP        1 0 -2  ; _ENV "y"
  3       [1]     CALL            1 1 0
  4       [1]     CALL            0 0 5
  5       [1]     RETURN          0 1
constants (2) for 000000CECB2D6CC0:
  1       "z"
  2       "y"
locals (4) for 000000CECB2D6CC0:
  0       p       5       6
  1       q       5       6
  2       r       5       6
  3       s       5       6
upvalues (1) for 000000CECB2D6CC0:
  0       _ENV    0       0
```

First, the function references are retrieved (lines [1] and [2]), then function y is called first (temporary register 1). The CALL has a field C of 0, meaning multiple return values are accepted. These return values become the parameters to function z, and so in line [4], field B of the CALL instruction is 0, signifying multiple parameters. After the call to function z, 4 results are retained, so field C in line [4] is 5. Finally, here is an example with calls to standard library functions:

```
x=function() print(string.char(64)) end
```

Leads to:

```
function <stdin:1,1> (7 instructions at 000000CECB2D6220)
0 params, 3 slots, 1 upvalue, 0 locals, 4 constants, 0 functions
  1       [1]     GETTABUP        0 0 -1  ; _ENV "print"
  2       [1]     GETTABUP        1 0 -2  ; _ENV "string"
  3       [1]     GETTABLE        1 1 -3  ; "char"
  4       [1]     LOADK           2 -4    ; 64
  5       [1]     CALL            1 2 0
  6       [1]     CALL            0 0 1
  7       [1]     RETURN          0 1
constants (4) for 000000CECB2D6220:
  1       "print"
  2       "string"
  3       "char"
  4       64
```

```
locals (0) for 000000CECB2D6220:
upvalues (1) for 000000CECB2D6220:
  0       _ENV    0       0
```

When a function call is the last parameter to another function call, the former can pass multiple return values, while the latter can accept multiple parameters.

2.5 OP_TAILCALL instruction

2.5.1 Syntax

```
TAILCALL  A B C return R(A)(R(A+1), ... ,R(A+B-1))
```

2.5.2 Description

Performs a tail call, which happens when a return statement has a single function call as the expression, e.g. return foo(bar). A tail call results in the function being interpreted within the same call frame as the caller - the stack is replaced and then a 'goto' executed to start at the entry point in the VM. Only Lua functions can be tailcalled. Tailcalls allow infinite recursion without growing the stack.

Like `OP_CALL`, register R(A) holds the reference to the function object to be called. B encodes the number of parameters in the same manner as a `OP_CALL` instruction.

C isn't used by TAILCALL, since all return results are significant. In any case, Lua always generates a 0 for C, to denote multiple return results.

2.5.3 Examples

An `OP_TAILCALL` is used only for one specific return style, described above. Multiple return results are always produced by a tail call. Here is an example:

```
function y() return x('foo', 'bar') end
```

Generates:

```
function <stdin:1,1> (6 instructions at 000000C3C24DE4A0)
0 params, 3 slots, 1 upvalue, 0 locals, 3 constants, 0 functions
  1       [1]     GETTABUP        0 0 -1  ; _ENV "x"
  2       [1]     LOADK           1 -2    ; "foo"
  3       [1]     LOADK           2 -3    ; "bar"
  4       [1]     TAILCALL        0 3 0
  5       [1]     RETURN          0 0
  6       [1]     RETURN          0 1
constants (3) for 000000C3C24DE4A0:
  1       "x"
  2       "foo"
  3       "bar"
locals (0) for 000000C3C24DE4A0:
upvalues (1) for 000000C3C24DE4A0:
  0       _ENV    0       0
```

Arguments for a tail call are handled in exactly the same way as arguments for a normal call, so in line [4], the tail call has a field B value of 3, signifying 2 parameters. Field C is 0, for multiple returns; this due to the constant

LUA_MULTRET in lua.h. In practice, field C is not used by the virtual machine (except as an assert) since the syntax guarantees multiple return results. Line [5] is a OP_RETURN instruction specifying multiple return results. This is required when the function called by OP_TAILCALL is a C function. In the case of a C function, execution continues to line [5] upon return, thus the RETURN is necessary. Line [6] is redundant. When Lua functions are tailcalled, the virtual machine does not return to line [5] at all.

2.6 OP_RETURN instruction

2.6.1 Syntax

```
RETURN  A B return R(A), ... ,R(A+B-2)
```

2.6.2 Description

Returns to the calling function, with optional return values.

First OP_RETURN closes any open upvalues by calling luaF_close().

If B is 1, there are no return values. If B is 2 or more, there are (B-1) return values, located in consecutive registers from R(A) onwards. If B is 0, the set of values range from R(A) to the top of the stack.

It is assumed that if the VM is returning to a Lua function then it is within the same invocation of the luaV_execute(). Else it is assumed that luaV_execute() is being invoked from a C function.

If B is 0 then the previous instruction (which must be either OP_CALL or OP_VARARG) would have set L->top to indicate how many values to return. The number of values to be returned in this case is R(A) to L->top.

If B > 0 then the number of values to be returned is simply B-1.

OP_RETURN calls luaD_poscall() which is responsible for copying return values to the caller - the first result is placed at the current closure's address. luaD_poscall() leaves L->top just past the last result that was copied.

If OP_RETURN is returning to a Lua function and if the number of return values expected was indeterminate - i.e. OP_CALL had operand C = 0, then L->top is left where luaD_poscall() placed it - just beyond the top of the result list. This allows the OP_CALL instruction to figure out how many results were returned. If however OP_CALL had invoked with a value of C > 0 then the expected number of results is known, and in that case, L->top is reset to the calling function's C->top.

If luaV_execute() was called externally then OP_RETURN leaves L->top unchanged - so it will continue to be just past the top of the results list. This is because luaV_execute() does not have a way of informing callers how many values were returned; so the caller can determine the number of results by inspecting L->top.

2.6.3 Examples

Example of OP_VARARG followed by OP_RETURN:

```
function x(...) return ... end

1 [1]  VARARG          0 0
2 [1]  RETURN          0 0
```

Suppose we call x(1,2,3); then, observe the setting of L->top when OP_RETURN executes:

```
(LOADK A=1 Bx=-2)       L->top = 4, ci->top = 4
(LOADK A=2 Bx=-3)       L->top = 4, ci->top = 4
(LOADK A=3 Bx=-4)       L->top = 4, ci->top = 4
(TAILCALL A=0 B=4 C=0) L->top = 4, ci->top = 4
(VARARG A=0 B=0)        L->top = 2, ci->top = 2  ; we are in x()
(RETURN A=0 B=0)        L->top = 3, ci->top = 2
```

Observe that `OP_VARARG` set `L->top` to `base+3`.

But if we call `x(1)` instead:

```
(LOADK A=1 Bx=-2)       L->top = 4, ci->top = 4
(LOADK A=2 Bx=-3)       L->top = 4, ci->top = 4
(LOADK A=3 Bx=-4)       L->top = 4, ci->top = 4
(TAILCALL A=0 B=4 C=0) L->top = 4, ci->top = 4
(VARARG A=0 B=0)        L->top = 2, ci->top = 2 ; we are in x()
(RETURN A=0 B=0)        L->top = 1, ci->top = 2
```

Notice that this time `OP_VARARG` set `L->top` to `base+1`.

2.7 OP_JMP instruction

2.7.1 Syntax

```
JMP A sBx   pc+=sBx; if (A) close all upvalues >= R(A - 1)
```

2.7.2 Description

Performs an unconditional jump, with sBx as a signed displacement. sBx is added to the program counter (PC), which points to the next instruction to be executed. If sBx is 0, the VM will proceed to the next instruction.

If R(A) is not 0 then all upvalues >= R(A-1) will be closed by calling luaF_close().

`OP_JMP` is used in loops, conditional statements, and in expressions when a boolean true/false need to be generated.

2.7.3 Examples

For example, since a relational test instruction makes conditional jumps rather than generate a boolean result, a JMP is used in the code sequence for loading either a true or a false:

```
function x() local m, n; return m >= n end
```

Generates:

```
function <stdin:1,1> (7 instructions at 00000034D2ABE340)
0 params, 3 slots, 0 upvalues, 2 locals, 0 constants, 0 functions
  1       [1]     LOADNIL         0 1
  2       [1]     LE              1 1 0   ; to 4 if false    (n <= m)
  3       [1]     JMP             0 1     ; to 5
  4       [1]     LOADBOOL        2 0 1
  5       [1]     LOADBOOL        2 1 0
  6       [1]     RETURN          2 2
  7       [1]     RETURN          0 1
constants (0) for 00000034D2ABE340:
```

```
locals (2) for 00000034D2ABE340:
  0       m       2       8
  1       n       2       8
upvalues (0) for 00000034D2ABE340:
```

Line[2] performs the relational test. In line [3], the JMP skips over the false path (line [4]) to the true path (line [5]). The result is placed into temporary local 2, and returned to the caller by RETURN in line [6].

2.8 OP_VARARG instruction

2.8.1 Syntax

```
VARARG  A B R(A), R(A+1), ..., R(A+B-1) = vararg
```

2.8.2 Description

VARARG implements the vararg operator `...` in expressions. VARARG copies B-1 parameters into a number of registers starting from R(A), padding with nils if there aren't enough values. If B is 0, VARARG copies as many values as it can based on the number of parameters passed. If a fixed number of values is required, B is a value greater than 1. If any number of values is required, B is 0.

2.8.3 Examples

The use of VARARG will become clear with the help of a few examples:

```
local a,b,c = ...
```

Generates:

```
main <(string):0,0> (2 instructions at 00000029D9FA8310)
0+ params, 3 slots, 1 upvalue, 3 locals, 0 constants, 0 functions
      1       [1]     VARARG          0 4
      2       [1]     RETURN          0 1
constants (0) for 00000029D9FA8310:
locals (3) for 00000029D9FA8310:
      0       a       2       3
      1       b       2       3
      2       c       2       3
upvalues (1) for 00000029D9FA8310:
      0       _ENV    1       0
```

Note that the main or top-level chunk is a vararg function. In this example, the left hand side of the assignment statement needs three values (or objects.) So in instruction [1], the operand B of the VARARG instruction is (3+1), or 4. VARARG will copy three values into a, b and c. If there are less than three values available, nils will be used to fill up the empty places.

```
local a = function(...) local a,b,c = ... end
```

This gives:

```
main <(string):0,0> (2 instructions at 00000029D9FA72D0)
0+ params, 2 slots, 1 upvalue, 1 local, 0 constants, 1 function
      1       [1]     CLOSURE         0 0     ; 00000029D9FA86D0
```

```
      2       [1]     RETURN          0 1
constants (0) for 00000029D9FA72D0:
locals (1) for 00000029D9FA72D0:
      0       a       2       3
upvalues (1) for 00000029D9FA72D0:
      0       _ENV    1       0

function <(string):1,1> (2 instructions at 00000029D9FA86D0)
0+ params, 3 slots, 0 upvalues, 3 locals, 0 constants, 0 functions
      1       [1]     VARARG          0 4
      2       [1]     RETURN          0 1
constants (0) for 00000029D9FA86D0:
locals (3) for 00000029D9FA86D0:
      0       a       2       3
      1       b       2       3
      2       c       2       3
upvalues (0) for 00000029D9FA86D0:
```

Here is an alternate version where a function is instantiated and assigned to local a. The old-style arg is retained for compatibility purposes, but is unused in the above example.

```
local a; a(...)
```

Leads to:

```
main <(string):0,0> (5 instructions at 00000029D9FA6D30)
0+ params, 3 slots, 1 upvalue, 1 local, 0 constants, 0 functions
      1       [1]     LOADNIL         0 0
      2       [1]     MOVE            1 0
      3       [1]     VARARG          2 0
      4       [1]     CALL            1 0 1
      5       [1]     RETURN          0 1
constants (0) for 00000029D9FA6D30:
locals (1) for 00000029D9FA6D30:
      0       a       2       6
upvalues (1) for 00000029D9FA6D30:
      0       _ENV    1       0
```

When a function is called with `...` as the argument, the function will accept a variable number of parameters or arguments. On instruction [3], a `VARARG` with a B field of 0 is used. The `VARARG` will copy all the parameters passed on to the main chunk to register 2 onwards, so that the `CALL` in the next line can utilize them as parameters of function `a`. The function call is set to accept a multiple number of parameters and returns zero results.

```
local a = {...}
```

Produces:

```
main <(string):0,0> (4 instructions at 00000029D9FA8130)
0+ params, 2 slots, 1 upvalue, 1 local, 0 constants, 0 functions
      1       [1]     NEWTABLE        0 0 0
      2       [1]     VARARG          1 0
      3       [1]     SETLIST         0 0 1   ; 1
      4       [1]     RETURN          0 1
constants (0) for 00000029D9FA8130:
locals (1) for 00000029D9FA8130:
      0       a       4       5
upvalues (1) for 00000029D9FA8130:
      0       _ENV    1       0
```

And:

```
return ...
```

Produces:

```
main <(string):0,0> (3 instructions at 00000029D9FA8270)
0+ params, 2 slots, 1 upvalue, 0 locals, 0 constants, 0 functions
      1       [1]     VARARG          0 0
      2       [1]     RETURN          0 0
      3       [1]     RETURN          0 1
constants (0) for 00000029D9FA8270:
locals (0) for 00000029D9FA8270:
upvalues (1) for 00000029D9FA8270:
      0       _ENV    1       0
```

Above are two other cases where VARARG needs to copy all passed parameters over to a set of registers in order for the next operation to proceed. Both the above forms of table creation and return accepts a variable number of values or objects.

2.9 OP_LOADBOOL instruction

2.9.1 Syntax

```
LOADBOOL A B C    R(A) := (Bool)B; if (C) pc++
```

2.9.2 Description

Loads a boolean value (true or false) into register R(A). true is usually encoded as an integer 1, false is always 0. If C is non-zero, then the next instruction is skipped (this is used when you have an assignment statement where the expression uses relational operators, e.g. M = K>5.) You can use any non-zero value for the boolean true in field B, but since you cannot use booleans as numbers in Lua, it's best to stick to 1 for true.

LOADBOOL is used for loading a boolean value into a register. It's also used where a boolean result is supposed to be generated, because relational test instructions, for example, do not generate boolean results – they perform conditional jumps instead. The operand C is used to optionally skip the next instruction (by incrementing PC by 1) in order to support such code. For simple assignments of boolean values, C is always 0.

2.9.3 Examples

The following line of code:

```
f=load('local a,b = true,false')
```

generates:

```
main <(string):0,0> (3 instructions at 0000020F274C2610)
0+ params, 2 slots, 1 upvalue, 2 locals, 0 constants, 0 functions
      1       [1]     LOADBOOL        0 1 0
      2       [1]     LOADBOOL        1 0 0
      3       [1]     RETURN          0 1
constants (0) for 0000020F274C2610:
locals (2) for 0000020F274C2610:
      0       a       3       4
      1       b       3       4
```

```
upvalues (1) for 0000020F274C2610:
      0       _ENV    1       0
```

This example is straightforward: Line [1] assigns true to local a (register 0) while line [2] assigns false to local b (register 1). In both cases, field C is 0, so PC is not incremented and the next instruction is not skipped.

Next, look at this line:

```
f=load('local a = 5 > 2')
```

This leads to following bytecode:

```
main <(string):0,0> (5 instructions at 0000020F274BAE00)
0+ params, 2 slots, 1 upvalue, 1 local, 2 constants, 0 functions
      1       [1]     LT              1 -2 -1 ; 2 5
      2       [1]     JMP             0 1     ; to 4
      3       [1]     LOADBOOL        0 0 1
      4       [1]     LOADBOOL        0 1 0
      5       [1]     RETURN          0 1
constants (2) for 0000020F274BAE00:
      1       5
      2       2
locals (1) for 0000020F274BAE00:
      0       a       5       6
upvalues (1) for 0000020F274BAE00:
      0       _ENV    1       0
```

This is an example of an expression that gives a boolean result and is assigned to a variable. Notice that Lua does not optimize the expression into a true value; Lua does not perform compile-time constant evaluation for relational operations, but it can perform simple constant evaluation for arithmetic operations.

Since the relational operator `LT` does not give a boolean result but performs a conditional jump, `LOADBOOL` uses its C operand to perform an unconditional jump in line [3] – this saves one instruction and makes things a little tidier. The reason for all this is that the instruction set is simply optimized for if...then blocks. Essentially, `local a = 5 > 2` is executed in the following way:

```
local a
if 2 < 5 then
  a = true
else
  a = false
end
```

In the disassembly listing, when `LT` tests 2 < 5, it evaluates to true and doesn't perform a conditional jump. Line [2] jumps over the false result path, and in line [4], the local a (register 0) is assigned the boolean true by the instruction `LOADBOOL`. If 2 and 5 were reversed, line [3] will be followed instead, setting a false, and then the true result path (line [4]) will be skipped, since `LOADBOOL` has its field C set to non-zero.

So the true result path goes like this (additional comments in parentheses):

```
1       [1]     LT              1 -2 -1 ; 2 5       (if 2 < 5)
2       [1]     JMP             0 1     ; to 4
4       [1]     LOADBOOL        0 1 0   ;           (a = true)
5       [1]     RETURN          0 1
```

and the false result path (which never executes in this example) goes like this:

```
1       [1]     LT              1 -2 -1 ; 2 5       (if 2 < 5)
3       [1]     LOADBOOL        0 0 1               (a = false)
5       [1]     RETURN          0 1
```

The true result path looks longer, but it isn't, due to the way the virtual machine is implemented. This will be discussed further in the section on relational and logic instructions.

2.10 OP_EQ, OP_LT and OP_LE Instructions

Relational and logic instructions are used in conjunction with other instructions to implement control structures or expressions. Instead of generating boolean results, these instructions conditionally perform a jump over the next instruction; the emphasis is on implementing control blocks. Instructions are arranged so that there are two paths to follow based on the relational test.

```
EQ  A B C if ((RK(B) == RK(C)) ~= A) then PC++
LT  A B C if ((RK(B) <  RK(C)) ~= A) then PC++
LE  A B C if ((RK(B) <= RK(C)) ~= A) then PC++
```

2.10.1 Description

Compares RK(B) and RK(C), which may be registers or constants. If the boolean result is not A, then skip the next instruction. Conversely, if the boolean result equals A, continue with the next instruction.

`EQ` is for equality. `LT` is for "less than" comparison. `LE` is for "less than or equal to" comparison. The boolean A field allows the full set of relational comparison operations to be synthesized from these three instructions. The Lua code generator produces either 0 or 1 for the boolean A.

For the fall-through case, a *OP_JMP instruction* is always expected, in order to optimize execution in the virtual machine. In effect, `EQ`, `LT` and `LE` must always be paired with a following `JMP` instruction.

2.10.2 Examples

By comparing the result of the relational operation with A, the sense of the comparison can be reversed. Obviously the alternative is to reverse the paths taken by the instruction, but that will probably complicate code generation some more. The conditional jump is performed if the comparison result is not A, whereas execution continues normally if the comparison result matches A. Due to the way code is generated and the way the virtual machine works, a `JMP` instruction is always expected to follow an `EQ`, `LT` or `LE`. The following `JMP` is optimized by executing it in conjunction with `EQ`, `LT` or `LE`.

```
local x,y; return x ~= y
```

Generates:

```
main <(string):0,0> (7 instructions at 0000001BC48FD390)
0+ params, 3 slots, 1 upvalue, 2 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     EQ              0 0 1
      3       [1]     JMP             0 1     ; to 5
      4       [1]     LOADBOOL        2 0 1
      5       [1]     LOADBOOL        2 1 0
      6       [1]     RETURN          2 2
      7       [1]     RETURN          0 1
constants (0) for 0000001BC48FD390:
locals (2) for 0000001BC48FD390:
      0       x       2       8
      1       y       2       8
upvalues (1) for 0000001BC48FD390:
      0       _ENV    1       0
```

In the above example, the equality test is performed in instruction [2]. However, since the comparison need to be returned as a result, LOADBOOL instructions are used to set a register with the correct boolean value. This is the usual code pattern generated if the expression requires a boolean value to be generated and stored in a register as an intermediate value or a final result.

It is easier to visualize the disassembled code as:

```
if x ~= y then
  return true
else
  return false
end
```

The true result path (when the comparison result matches A) goes like this:

```
1  [1] LOADNIL    0   1
2  [1] EQ         0   0   1    ; to 4 if true    (x ~= y)
3  [1] JMP        1            ; to 5
5  [1] LOADBOOL   2   1   0    ; true            (true path)
6  [1] RETURN     2   2
```

While the false result path (when the comparison result does not match A) goes like this:

```
1  [1] LOADNIL    0   1
2  [1] EQ         0   0   1    ; to 4 if true    (x ~= y)
4  [1] LOADBOOL   2   0   1    ; false, to 6     (false path)
6  [1] RETURN     2   2
```

Comments following the EQ in line [2] lets the user know when the conditional jump is taken. The jump is taken when "the value in register 0 equals to the value in register 1" (the comparison) is not false (the value of operand A). If the comparison is x == y, everything will be the same except that the A operand in the EQ instruction will be 1, thus reversing the sense of the comparison. Anyway, these are just the Lua code generator's conventions; there are other ways to code x ~= y in terms of Lua virtual machine instructions.

For conditional statements, there is no need to set boolean results. Lua is optimized for coding the more common conditional statements rather than conditional expressions.

```
local x,y; if x ~= y then return "foo" else return "bar" end
```

Results in:

```
main <(string):0,0> (9 instructions at 0000001BC4914D50)
0+ params, 3 slots, 1 upvalue, 2 locals, 2 constants, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     EQ              1 0 1   ; to 4 if false    (x ~= y)
      3       [1]     JMP             0 3     ; to 7
      4       [1]     LOADK           2 -1    ; "foo"           (true block)
      5       [1]     RETURN          2 2
      6       [1]     JMP             0 2     ; to 9
      7       [1]     LOADK           2 -2    ; "bar"           (false block)
      8       [1]     RETURN          2 2
      9       [1]     RETURN          0 1
constants (2) for 0000001BC4914D50:
      1       "foo"
      2       "bar"
locals (2) for 0000001BC4914D50:
      0       x       2       10
      1       y       2       10
upvalues (1) for 0000001BC4914D50:
      0       _ENV    1       0
```

In the above conditional statement, the same inequality operator is used in the source, but the sense of the EQ instruction in line [2] is now reversed. Since the EQ conditional jump can only skip the next instruction, additional JMP instructions are needed to allow large blocks of code to be placed in both true and false paths. In contrast, in the previous example, only a single instruction is needed to set a boolean value. For if statements, the true block comes first followed by the false block in code generated by the code generator. To reverse the positions of the true and false paths, the value of operand A is changed.

The true path (when x ~= y is true) goes from [2] to [4]–[6] and on to [9]. Since there is a RETURN in line [5], the JMP in line [6] and the RETURN in [9] are never executed at all; they are redundant but does not adversely affect performance in any way. The false path is from [2] to [3] to [7]–[9] onwards. So in a disassembly listing, you should see the true and false code blocks in the same order as in the Lua source.

The following is another example, this time with an elseif:

```
if 8 > 9 then return 8 elseif 5 >= 4 then return 5 else return 9 end
```

Generates:

```
main <(string):0,0> (13 instructions at 0000001BC4913770)
0+ params, 2 slots, 1 upvalue, 0 locals, 4 constants, 0 functions
      1       [1]     LT              0 -2 -1 ; 9 8
      2       [1]     JMP             0 3     ; to 6
      3       [1]     LOADK           0 -1    ; 8
      4       [1]     RETURN          0 2
      5       [1]     JMP             0 7     ; to 13
      6       [1]     LE              0 -4 -3 ; 4 5
      7       [1]     JMP             0 3     ; to 11
      8       [1]     LOADK           0 -3    ; 5
      9       [1]     RETURN          0 2
      10      [1]     JMP             0 2     ; to 13
      11      [1]     LOADK           0 -2    ; 9
      12      [1]     RETURN          0 2
      13      [1]     RETURN          0 1
constants (4) for 0000001BC4913770:
      1       8
      2       9
      3       5
      4       4
locals (0) for 0000001BC4913770:
upvalues (1) for 0000001BC4913770:
      0       _ENV    1       0
```

This example is a little more complex, but the blocks are structured in the same order as the Lua source, so interpreting the disassembled code should not be too hard.

2.11 OP_TEST and OP_TESTSET instructions

2.11.1 Syntax

```
TEST        A C     if not (R(A) <=> C) then pc++
TESTSET     A B C   if (R(B) <=> C) then R(A) := R(B) else pc++
```

2.11.2 Description

These two instructions used for performing boolean tests and implementing Lua's logic operators.

Used to implement and and or logical operators, or for testing a single register in a conditional statement.

For TESTSET, register R(B) is coerced into a boolean and compared to the boolean field C. If R(B) matches C, the next instruction is skipped, otherwise R(B) is assigned to R(A) and the VM continues with the next instruction. The and operator uses a C of 0 (false) while or uses a C value of 1 (true).

TEST is a more primitive version of TESTSET. TEST is used when the assignment operation is not needed, otherwise it is the same as TESTSET except that the operand slots are different.

For the fall-through case, a JMP is always expected, in order to optimize execution in the virtual machine. In effect, TEST and TESTSET must always be paired with a following JMP instruction.

2.11.3 Examples

TEST and TESTSET are used in conjunction with a following JMP instruction, while TESTSET has an addditional conditional assignment. Like EQ, LT and LE, the following JMP instruction is compulsory, as the virtual machine will execute the JMP together with TEST or TESTSET. The two instructions are used to implement short-circuit LISP-style logical operators that retains and propagates operand values instead of booleans. First, we'll look at how and and or behaves:

```
f=load('local a,b,c; c = a and b')
```

Generates:

```
main <(string):0,0> (5 instructions at 0000020F274CF1A0)
0+ params, 3 slots, 1 upvalue, 3 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 2
      2       [1]     TESTSET         2 0 0   ; to 4 if true
      3       [1]     JMP             0 1     ; to 5
      4       [1]     MOVE            2 1
      5       [1]     RETURN          0 1
constants (0) for 0000020F274CF1A0:
locals (3) for 0000020F274CF1A0:
      0       a       2       6
      1       b       2       6
      2       c       2       6
upvalues (1) for 0000020F274CF1A0:
      0       _ENV    1       0
```

An and sequence exits on false operands (which can be false or nil) because any false operands in a string of and operations will make the whole boolean expression false. If operands evaluates to true, evaluation continues. When a string of and operations evaluates to true, the result is the last operand value.

In line [2], the first operand (the local a) is set to local c when the test is false (with a field C of 0), while the jump to [4] is made when the test is true, and then in line [4], the expression result is set to the second operand (the local b). This is equivalent to:

```
if a then
  c = b      -- executed by MOVE on line [4]
else
  c = a      -- executed by TESTSET on line [2]
end
```

The c = a portion is done by TESTSET itself, while MOVE performs c = b. Now, if the result is already set with one of the possible values, a TEST instruction is used instead:

```
f=load('local a,b; a = a and b')
```

Generates:

```
main <(string):0,0> (5 instructions at 0000020F274D0A70)
0+ params, 2 slots, 1 upvalue, 2 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     TEST            0 0     ; to 4 if true
      3       [1]     JMP             0 1     ; to 5
      4       [1]     MOVE            0 1
      5       [1]     RETURN          0 1
constants (0) for 0000020F274D0A70:
locals (2) for 0000020F274D0A70:
      0       a       2       6
      1       b       2       6
upvalues (1) for 0000020F274D0A70:
      0       _ENV    1       0
```

The `TEST` instruction does not perform an assignment operation, since `a = a` is redundant. This makes `TEST` a little faster. This is equivalent to:

```
if a then
  a = b
end
```

Next, we will look at the or operator:

```
f=load('local a,b,c; c = a or b')
```

Generates:

```
main <(string):0,0> (5 instructions at 0000020F274D1AB0)
0+ params, 3 slots, 1 upvalue, 3 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 2
      2       [1]     TESTSET         2 0 1   ; to 4 if false
      3       [1]     JMP             0 1     ; to 5
      4       [1]     MOVE            2 1
      5       [1]     RETURN          0 1
constants (0) for 0000020F274D1AB0:
locals (3) for 0000020F274D1AB0:
      0       a       2       6
      1       b       2       6
      2       c       2       6
upvalues (1) for 0000020F274D1AB0:
      0       _ENV    1       0
```

An `or` sequence exits on `true` operands, because any operands evaluating to `true` in a string of or operations will make the whole boolean expression `true`. If operands evaluates to `false`, evaluation continues. When a string of or operations evaluates to `false`, all operands must have evaluated to `false`.

In line [2], the local `a` value is set to local c if it is `true`, while the jump is made if it is `false` (the field C is 1). Thus in line [4], the local `b` value is the result of the expression if local `a` evaluates to `false`. This is equivalent to:

```
if a then
  c = a      -- executed by TESTSET on line [2]
else
  c = b      -- executed by MOVE on line [4]
end
```

Like the case of and, TEST is used when the result already has one of the possible values, saving an assignment operation:

```
f=load('local a,b; a = a or b')
```

Generates:

```
main <(string):0,0> (5 instructions at 0000020F274D1010)
0+ params, 2 slots, 1 upvalue, 2 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     TEST            0 1     ; to 4 if false
      3       [1]     JMP             0 1     ; to 5
      4       [1]     MOVE            0 1
      5       [1]     RETURN          0 1
constants (0) for 0000020F274D1010:
locals (2) for 0000020F274D1010:
      0       a       2       6
      1       b       2       6
upvalues (1) for 0000020F274D1010:
      0       _ENV    1       0
```

Short-circuit logical operators also means that the following Lua code does not require the use of a boolean operation:

```
f=load('local a,b,c; if a > b and a > c then return a end')
```

Leads to:

```
main <(string):0,0> (7 instructions at 0000020F274D1150)
0+ params, 3 slots, 1 upvalue, 3 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 2
      2       [1]     LT              0 1 0   ; to 4 if true
      3       [1]     JMP             0 3     ; to 7
      4       [1]     LT              0 2 0   ; to 6 if true
      5       [1]     JMP             0 1     ; to 7
      6       [1]     RETURN          0 2
      7       [1]     RETURN          0 1
constants (0) for 0000020F274D1150:
locals (3) for 0000020F274D1150:
      0       a       2       8
      1       b       2       8
      2       c       2       8
upvalues (1) for 0000020F274D1150:
      0       _ENV    1       0
```

With short-circuit evaluation, `a > c` is never executed if `a > b` is false, so the logic of the Lua statement can be readily implemented using the normal conditional structure. If both `a > b` and `a > c` are true, the path followed is [2] (the `a > b` test) to [4] (the `a > c` test) and finally to [6], returning the value of `a`. A `TEST` instruction is not required. This is equivalent to:

```
if a > b then
  if a > c then
    return a
  end
end
```

For a single variable used in the expression part of a conditional statement, `TEST` is used to boolean-test the variable:

```
f=load('if Done then return end')
```

Generates:

```
main <(string):0,0> (5 instructions at 0000020F274D13D0)
0+ params, 2 slots, 1 upvalue, 0 locals, 1 constant, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "Done"
      2       [1]     TEST            0 0     ; to 4 if true
```

```
      3       [1]     JMP             0 1     ; to 5
      4       [1]     RETURN          0 1
      5       [1]     RETURN          0 1
constants (1) for 0000020F274D13D0:
      1       "Done"
locals (0) for 0000020F274D13D0:
upvalues (1) for 0000020F274D13D0:
      0       _ENV    1       0
```

In line [2], the `TEST` instruction jumps to the `true` block if the value in temporary register 0 (from the global `Done`) is `true`. The `JMP` at line [3] jumps over the `true` block, which is the code inside the if block (line [4]).

If the test expression of a conditional statement consist of purely boolean operators, then a number of TEST instructions will be used in the usual short-circuit evaluation style:

```
f=load('if Found and Match then return end')
```

Generates:

```
main <(string):0,0> (8 instructions at 0000020F274D1C90)
0+ params, 2 slots, 1 upvalue, 0 locals, 2 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "Found"
      2       [1]     TEST            0 0     ; to 4 if true
      3       [1]     JMP             0 4     ; to 8
      4       [1]     GETTABUP        0 0 -2  ; _ENV "Match"
      5       [1]     TEST            0 0     ; to 7 if true
      6       [1]     JMP             0 1     ; to 8
      7       [1]     RETURN          0 1
      8       [1]     RETURN          0 1
constants (2) for 0000020F274D1C90:
      1       "Found"
      2       "Match"
locals (0) for 0000020F274D1C90:
upvalues (1) for 0000020F274D1C90:
      0       _ENV    1       0
```

In the last example, the true block of the conditional statement is executed only if both `Found` and `Match` evaluate to `true`. The path is from [2] (test for `Found`) to [4] to [5] (test for `Match`) to [7] (the true block, which is an explicit `return` statement.)

If the statement has an `else` section, then the `JMP` on line [6] will jump to the false block (the `else` block) while an additional `JMP` will be added to the true block to jump over this new block of code. If `or` is used instead of `and`, the appropriate C operand will be adjusted accordingly.

Finally, here is how Lua's ternary operator (:? in C) equivalent works:

```
f=load('local a,b,c; a = a and b or c')
```

Generates:

```
main <(string):0,0> (7 instructions at 0000020F274D1A10)
0+ params, 3 slots, 1 upvalue, 3 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 2
      2       [1]     TEST            0 0     ; to 4 if true
      3       [1]     JMP             0 2     ; to 6
      4       [1]     TESTSET         0 1 1   ; to 6 if false
      5       [1]     JMP             0 1     ; to 7
      6       [1]     MOVE            0 2
      7       [1]     RETURN          0 1
constants (0) for 0000020F274D1A10:
```

```
locals (3) for 0000020F274D1A10:
      0       a       2       8
      1       b       2       8
      2       c       2       8
upvalues (1) for 0000020F274D1A10:
      0       _ENV    1       0
```

The `TEST` in line [2] is for the `and` operator. First, local `a` is tested in line [2]. If it is false, then execution continues in [3], jumping to line [6]. Line [6] assigns local `c` to the end result because since if `a` is false, then `a and b` is `false`, and `false or c` is `c`.

If local `a` is `true` in line [2], the `TEST` instruction makes a jump to line [4], where there is a `TESTSET`, for the `or` operator. If `b` evaluates to `true`, then the end result is assigned the value of `b`, because `b or c` is `b` if `b` is `not false`. If `b` is also `false`, the end result will be `c`.

For the instructions in line [2], [4] and [6], the target (in field A) is register 0, or the local `a`, which is the location where the result of the boolean expression is assigned. The equivalent Lua code is:

```
if a then
  if b then
    a = b
  else
    a = c
  end
else
  a = c
end
```

The two `a = c` assignments are actually the same piece of code, but are repeated here to avoid using a `goto` and a label. Normally, if we assume `b` is `not false` and `not nil`, we end up with the more recognizable form:

```
if a then
  a = b     -- assuming b ~= false
else
  a = c
end
```

2.12 OP_FORPREP and OP_FORLOOP instructions

2.12.1 Syntax

```
FORPREP     A sBx   R(A)-=R(A+2); pc+=sBx
FORLOOP     A sBx   R(A)+=R(A+2);
                    if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) }
```

2.12.2 Description

Lua has dedicated instructions to implement the two types of `for` loops, while the other two types of loops uses traditional test-and-jump.

`FORPREP` initializes a numeric for loop, while `FORLOOP` performs an iteration of a numeric for loop.

A numeric for loop requires 4 registers on the stack, and each register must be a number. R(A) holds the initial value and doubles as the internal loop variable (the internal index); R(A+1) is the limit; R(A+2) is the stepping value; R(A+3) is the actual loop variable (the external index) that is local to the for block.

FORPREP sets up a for loop. Since FORLOOP is used for initial testing of the loop condition as well as conditional testing during the loop itself, FORPREP performs a negative step and jumps unconditionally to FORLOOP so that FORLOOP is able to correctly make the initial loop test. After this initial test, FORLOOP performs a loop step as usual, restoring the initial value of the loop index so that the first iteration can start.

In FORLOOP, a jump is made back to the start of the loop body if the limit has not been reached or exceeded. The sense of the comparison depends on whether the stepping is negative or positive, hence the "<?=" operator. Jumps for both instructions are encoded as signed displacements in the sBx field. An empty loop has a FORLOOP sBx value of -1.

FORLOOP also sets R(A+3), the external loop index that is local to the loop block. This is significant if the loop index is used as an upvalue (see below.) R(A), R(A+1) and R(A+2) are not visible to the programmer.

The loop variable ends with the last value before the limit is reached (unlike C) because it is not updated unless the jump is made. However, since loop variables are local to the loop itself, you should not be able to use it unless you cook up an implementation-specific hack.

2.12.3 Examples

For the sake of efficiency, FORLOOP contains a lot of functionality, so when a loop iterates, only one instruction, FORLOOP, is needed. Here is a simple example:

```
f=load('local a = 0; for i = 1,100,5 do a = a + i end')
```

Generates:

```
main <(string):0,0> (8 instructions at 000001E9F0DF52F0)
0+ params, 5 slots, 1 upvalue, 5 locals, 4 constants, 0 functions
      1       [1]     LOADK           0 -1    ; 0
      2       [1]     LOADK           1 -2    ; 1
      3       [1]     LOADK           2 -3    ; 100
      4       [1]     LOADK           3 -4    ; 5
      5       [1]     FORPREP         1 1     ; to 7
      6       [1]     ADD             0 0 4
      7       [1]     FORLOOP         1 -2    ; to 6
      8       [1]     RETURN          0 1
constants (4) for 000001E9F0DF52F0:
      1       0
      2       1
      3       100
      4       5
locals (5) for 000001E9F0DF52F0:
      0       a       2       9
      1       (for index)     5       8
      2       (for limit)     5       8
      3       (for step)      5       8
      4       i       6       7
upvalues (1) for 000001E9F0DF52F0:
      0       _ENV    1       0
```

In the above example, notice that the for loop causes three additional local pseudo-variables (or internal variables) to be defined, apart from the external loop index, i. The three pseudovariables, named (for index), (for limit) and (for step) are required to completely specify the state of the loop, and are not visible to Lua source code. They are arranged in consecutive registers, with the external loop index given by R(A+3) or register 4 in the example.

The loop body is in line [6] while line [7] is the FORLOOP instruction that steps through the loop state. The sBx field of FORLOOP is negative, as it always jumps back to the beginning of the loop body.

Lines [2]–[4] initialize the three register locations where the loop state will be stored. If the loop step is not specified in the Lua source, a constant 1 is added to the constant pool and a `LOADK` instruction is used to initialize the pseudo-variable `(for step)` with the loop step.

`FORPREP` in lines [5] makes a negative loop step and jumps to line [7] for the initial test. In the example, at line [5], the internal loop index (at register 1) will be (1-5) or -4. When the virtual machine arrives at the `FORLOOP` in line [7] for the first time, one loop step is made prior to the first test, so the initial value that is actually tested against the limit is (-4+5) or 1. Since 1 < 100, an iteration will be performed. The external loop index `i` is then set to 1 and a jump is made to line [6], thus starting the first iteration of the loop.

The loop at line [6]–[7] repeats until the internal loop index exceeds the loop limit of 100. The conditional jump is not taken when that occurs and the loop ends. Beyond the scope of the loop body, the loop state (`(for index)`, `(for limit)`, `(for step)` and `i`) is not valid. This is determined by the parser and code generator. The range of PC values for which the loop state variables are valid is located in the locals list.

Here is another example:

```
f=load('for i = 10,1,-1 do if i == 5 then break end end')
```

This leads to:

```
main <(string):0,0> (8 instructions at 000001E9F0DEC110)
0+ params, 4 slots, 1 upvalue, 4 locals, 4 constants, 0 functions
      1       [1]     LOADK           0 -1    ; 10
      2       [1]     LOADK           1 -2    ; 1
      3       [1]     LOADK           2 -3    ; -1
      4       [1]     FORPREP         0 2     ; to 7
      5       [1]     EQ              1 3 -4  ; - 5
      6       [1]     JMP             0 1     ; to 8
      7       [1]     FORLOOP         0 -3    ; to 5
      8       [1]     RETURN          0 1
constants (4) for 000001E9F0DEC110:
      1       10
      2       1
      3       -1
      4       5
locals (4) for 000001E9F0DEC110:
      0       (for index)     4       8
      1       (for limit)     4       8
      2       (for step)      4       8
      3       i       5       7
upvalues (1) for 000001E9F0DEC110:
      0       _ENV    1       0
```

In the second loop example above, except for a negative loop step size, the structure of the loop is identical. The body of the loop is from line [5] to line [7]. Since no additional stacks or states are used, a break translates simply to a `JMP` instruction (line [6]). There is nothing to clean up after a `FORLOOP` ends or after a `JMP` to exit a loop.

2.13 OP_TFORCALL and OP_TFORLOOP instructions

2.13.1 Syntax

```
TFORCALL    A C         R(A+3), ... ,R(A+2+C) := R(A)(R(A+1), R(A+2))
TFORLOOP    A sBx       if R(A+1) ~= nil then { R(A)=R(A+1); pc += sBx }
```

2.13.2 Description

Apart from a numeric `for` loop (implemented by `FORPREP` and `FORLOOP`), Lua has a generic `for` loop, implemented by `TFORCALL` and `TFORLOOP`.

The generic `for` loop keeps 3 items in consecutive register locations to keep track of things. R(A) is the iterator function, which is called once per loop. R(A+1) is the state, and R(A+2) is the control variable. At the start, R(A+2) has an initial value. R(A), R(A+1) and R(A+2) are internal to the loop and cannot be accessed by the programmer.

In addition to these internal loop variables, the programmer specifies one or more loop variables that are external and visible to the programmer. These loop variables reside at locations R(A+3) onwards, and their count is specified in operand C. Operand C must be at least 1. They are also local to the loop body, like the external loop index in a numerical for loop.

Each time `TFORCALL` executes, the iterator function referenced by R(A) is called with two arguments: the state and the control variable (R(A+1) and R(A+2)). The results are returned in the local loop variables, from R(A+3) onwards, up to R(A+2+C).

Next, the `TFORLOOP` instruction tests the first return value, R(A+3). If it is nil, the iterator loop is at an end, and the `for` loop block ends by simply moving to the next instruction.

If R(A+3) is not nil, there is another iteration, and R(A+3) is assigned as the new value of the control variable, R(A+2). Then the `TFORLOOP` instruction sends execution back to the beginning of the loop (the `sBx` operand specifies how many instructions to move to get to the start of the loop body).

2.13.3 Examples

This example has a loop with one additional result (`v`) in addition to the loop enumerator (`i`):

```
f=load('for i,v in pairs(t) do print(i,v) end')
```

This produces:

```
main <(string):0,0> (11 instructions at 0000014DB7FD2610)
0+ params, 8 slots, 1 upvalue, 5 locals, 3 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "pairs"
      2       [1]     GETTABUP        1 0 -2  ; _ENV "t"
      3       [1]     CALL            0 2 4
      4       [1]     JMP             0 4     ; to 9
      5       [1]     GETTABUP        5 0 -3  ; _ENV "print"
      6       [1]     MOVE            6 3
      7       [1]     MOVE            7 4
      8       [1]     CALL            5 3 1
      9       [1]     TFORCALL        0 2
      10      [1]     TFORLOOP        2 -6    ; to 5
      11      [1]     RETURN          0 1
constants (3) for 0000014DB7FD2610:
      1       "pairs"
      2       "t"
      3       "print"
locals (5) for 0000014DB7FD2610:
      0       (for generator) 4       11
      1       (for state)     4       11
      2       (for control)   4       11
      3       i       5       9
      4       v       5       9
upvalues (1) for 0000014DB7FD2610:
      0       _ENV    1       0
```

The iterator function is located in register 0, and is named `(for generator)` for debugging purposes. The state is in register 1, and has the name `(for state)`. The control variable, `(for control)`, is contained in register 2. These correspond to locals R(A), R(A+1) and R(A+2) in the `TFORCALL` description. Results from the iterator function call is placed into register 3 and 4, which are locals `i` and `v`, respectively. On line [9], the operand C of `TFORCALL` is 2, corresponding to two iterator variables (`i` and `v`).

Lines [1]–[3] prepares the iterator state. Note that the call to the `pairs()` standard library function has 1 parameter and 3 results. After the call in line [3], register 0 is the iterator function (which by default is the Lua function `next()` unless `__pairs` meta method has been overriden), register 1 is the loop state, register 2 is the initial value of the control variable (which is `nil` in the default case). The iterator variables `i` and `v` are both invalid at the moment, because we have not entered the loop yet.

Line [4] is a `JMP` to `TFORCALL` on line [9]. The `TFORCALL` instruction calls the iterator function, generating the first set of enumeration results in locals `i` and `v`.

The `TFORLOOP` insruction executes and checks whether `i` is `nil`. If it is not `nil`, then the internal control variable (register 2) is set to the value in `i` and control goes back to to the start of the loop body (lines [5]–[8]).

The body of the generic `for` loop executes (`print(i,v)`) and then `TFORCALL` is encountered again, calling the iterator function to get the next iteration state. Finally, when the `TFORLOOP` finds that the first result from the iterator is `nil`, the loop ends, and execution continues on line [11].

2.14 OP_CLOSURE instruction

2.14.1 Syntax

```
CLOSURE A Bx    R(A) := closure(KPROTO[Bx])
```

2.14.2 Description

Creates an instance (or closure) of a function prototype. The `Bx` parameter identifies the entry in the parent function's table of closure prototypes (the field `p` in the struct `Proto`). The indices start from 0, i.e., a parameter of Bx = 0 references the first closure prototype in the table.

The `OP_CLOSURE` instruction also sets up the `upvalues` for the closure being defined. This is an involved process that is worthy of detailed discussion, and will be described through examples.

2.14.3 Examples

Let's start with a simple example of a Lua function:

```
f=load('function x() end; function y() end')
```

Here we are creating two Lua functions/closures within the main chunk. The bytecodes for the chunk look this:

```
main <(string):0,0> (5 instructions at 0000020E8A352930)
0+ params, 2 slots, 1 upvalue, 0 locals, 2 constants, 2 functions
      1       [1]     CLOSURE         0 0     ; 0000020E8A352A70
      2       [1]     SETTABUP        0 -1 0  ; _ENV "x"
      3       [1]     CLOSURE         0 1     ; 0000020E8A3536A0
      4       [1]     SETTABUP        0 -2 0  ; _ENV "y"
      5       [1]     RETURN          0 1
constants (2) for 0000020E8A352930:
      1       "x"
```

```
      2       "y"
locals (0) for 0000020E8A352930:
upvalues (1) for 0000020E8A352930:
      0       _ENV    1       0

function <(string):1,1> (1 instruction at 0000020E8A352A70)
0 params, 2 slots, 0 upvalues, 0 locals, 0 constants, 0 functions
      1       [1]     RETURN          0 1
constants (0) for 0000020E8A352A70:
locals (0) for 0000020E8A352A70:
upvalues (0) for 0000020E8A352A70:

function <(string):1,1> (1 instruction at 0000020E8A3536A0)
0 params, 2 slots, 0 upvalues, 0 locals, 0 constants, 0 functions
      1       [1]     RETURN          0 1
constants (0) for 0000020E8A3536A0:
locals (0) for 0000020E8A3536A0:
upvalues (0) for 0000020E8A3536A0:
```

What we observe is that the first `CLOSURE` instruction has parameter `Bx` set to 0, and this is the reference to the closure 0000020E8A352A70 which appears at position 0 in the table of closures within the main chunk's `Proto` structure.

Similarly the second `CLOSURE` instruction has parameter `Bx` set to 1, and this references the closure at position 1 in the table, which is 0000020E8A3536A0.

Other things to notice is that the main chunk got an automatic upvalue named `_ENV`:

```
upvalues (1) for 0000020E8A352930:
      0       _ENV    1       0
```

The first `0` is the index of the upvalue in the main chunk. The `1` following the name is a boolean indicating that the upvalue is located on the stack, and the last `0` is identifies the register location on the stack. So the Lua Parser has setup the `upvalue` reference for `_ENV`. However note that there is no actual local in this case; the `_ENV` upvalue is special and is setup by the Lua lua_load() API function.

Now let's look at an example that creates a local up-value:

```
f=load('local u,v; function p() return v end')
```

We get following bytecodes:

```
main <(string):0,0> (4 instructions at 0000022149BBA3B0)
0+ params, 3 slots, 1 upvalue, 2 locals, 1 constant, 1 function
      1       [1]     LOADNIL         0 1
      2       [1]     CLOSURE         2 0     ; 0000022149BBB7B0
      3       [1]     SETTABUP        0 -1 2  ; _ENV "p"
      4       [1]     RETURN          0 1
constants (1) for 0000022149BBA3B0:
      1       "p"
locals (2) for 0000022149BBA3B0:
      0       u       2       5
      1       v       2       5
upvalues (1) for 0000022149BBA3B0:
      0       _ENV    1       0

function <(string):1,1> (3 instructions at 0000022149BBB7B0)
0 params, 2 slots, 1 upvalue, 0 locals, 0 constants, 0 functions
      1       [1]     GETUPVAL        0 0     ; v
      2       [1]     RETURN          0 2
      3       [1]     RETURN          0 1
```

```
constants (0) for 0000022149BBB7B0:
locals (0) for 0000022149BBB7B0:
upvalues (1) for 0000022149BBB7B0:
      0       v       1       1
```

In the function 'p' the upvalue list contains:

```
upvalues (1) for 0000022149BBB7B0:
      0       v       1       1
```

This says that the up-value is in the stack (first '1') and is located at register '1' of the parent function. Access to this upvalue is indirectly obtained via the `GETUPVAL` instruction on line 1.

Now, lets look at what happens when the upvalue is not directly within the parent function:

```
f=load('local u,v; function p() u=1; local function q() return v end end')
```

In this example, we have 1 upvalue reference in function 'p', which is 'u'. Function 'q' has one upvalue reference 'v' but this is not a variable in 'p', but is in the grand-parent. Here are the resulting bytecodes:

```
main <(string):0,0> (4 instructions at 0000022149BBFE40)
0+ params, 3 slots, 1 upvalue, 2 locals, 1 constant, 1 function
      1       [1]     LOADNIL         0 1
      2       [1]     CLOSURE         2 0     ; 0000022149BBFC60
      3       [1]     SETTABUP        0 -1 2  ; _ENV "p"
      4       [1]     RETURN          0 1
constants (1) for 0000022149BBFE40:
      1       "p"
locals (2) for 0000022149BBFE40:
      0       u       2       5
      1       v       2       5
upvalues (1) for 0000022149BBFE40:
      0       _ENV    1       0

function <(string):1,1> (4 instructions at 0000022149BBFC60)
0 params, 2 slots, 2 upvalues, 1 local, 1 constant, 1 function
      1       [1]     LOADK           0 -1    ; 1
      2       [1]     SETUPVAL        0 0     ; u
      3       [1]     CLOSURE         0 0     ; 0000022149BC06B0
      4       [1]     RETURN          0 1
constants (1) for 0000022149BBFC60:
      1       1
locals (1) for 0000022149BBFC60:
      0       q       4       5
upvalues (2) for 0000022149BBFC60:
      0       u       1       0
      1       v       1       1

function <(string):1,1> (3 instructions at 0000022149BC06B0)
0 params, 2 slots, 1 upvalue, 0 locals, 0 constants, 0 functions
      1       [1]     GETUPVAL        0 0     ; v
      2       [1]     RETURN          0 2
      3       [1]     RETURN          0 1
constants (0) for 0000022149BC06B0:
locals (0) for 0000022149BC06B0:
upvalues (1) for 0000022149BC06B0:
      0       v       0       1
```

We see that 'p' got the upvalue 'u' as expected, but it also got the upvalue 'v', and both are marked as 'instack' of the parent function:

```
upvalues (2) for 0000022149BBFC60:
      0       u       1       0
      1       v       1       1
```

The reason for this is that any upvalue references in the inmost nested function will also appear in the parent functions up the chain until the function whose stack contains the variable being referenced. So although the function 'p' does not directly reference 'v', but because its child function 'q' references 'v', 'p' gets the upvalue reference to 'v' as well.

Observe the upvalue list of 'q' now:

```
upvalues (1) for 0000022149BC06B0:
      0       v       0       1
```

'q' has one upvalue reference as expected, but this time the upvalue is not marked 'instack', which means that the reference is to an upvalue and not a local in the parent function (in this case 'p') and the upvalue index is '1' (i.e. the second upvalue in 'p').

2.14.4 Upvalue setup by OP_CLOSURE

When the `CLOSURE` instruction is executed, the up-values referenced by the prototype are resolved. So that means the actual resolution if upvalues occurs at runtime. This is done in the function pushclosure().

2.14.5 Caching of closures

The Lua VM maintains a cache of closures within each function prototype at runtime. If a closure is required that has the same set of upvalues as referenced by an existing closure then the VM reuses the existing closure rather than creating a new one. This is illustrated in this contrived example:

```
f=load('local v; local function q() return function() return v end end; return q(), q()')
```

When the statement `return q(), q()` is executed it will end up returning two closures that are really the same instance, as shown by the result of executing this code:

```
> f()
function: 000001E1E2F007E0      function: 000001E1E2F007E0
```

2.15 OP_GETUPVAL and OP_SETUPVAL instructions

2.15.1 Syntax

```
GETUPVAL  A B     R(A) := UpValue[B]
SETUPVAL  A B     UpValue[B] := R(A)
```

2.15.2 Description

`GETUPVAL` copies the value in upvalue number `B` into register `R(A)`. Each Lua function may have its own upvalue list. This upvalue list is internal to the virtual machine; the list of upvalue name strings in a prototype is not mandatory.

`SETUPVAL` copies the value from register `R(A)` into the upvalue number `B` in the upvalue list for that function.

2.15.3 Examples

GETUPVAL and SETUPVAL instructions use internally-managed upvalue lists. The list of upvalue name strings that are found in a function prototype is for debugging purposes; it is not used by the Lua virtual machine and can be stripped by luac. During execution, upvalues are set up by a CLOSURE, and maintained by the Lua virtual machine. In the following example, function b is declared inside the main chunk, and is shown in the disassembly as a function prototype within a function prototype. The indentation, which is not in the original output, helps to visually separate the two functions.

```
f=load('local a; function b() a = 1 return a end')
```

Leads to:

```
main <(string):0,0> (4 instructions at 000002853D5177F0)
0+ params, 2 slots, 1 upvalue, 1 local, 1 constant, 1 function
      1       [1]     LOADNIL         0 0
      2       [1]     CLOSURE         1 0     ; 000002853D517920
      3       [1]     SETTABUP        0 -1 1  ; _ENV "b"
      4       [1]     RETURN          0 1
constants (1) for 000002853D5177F0:
      1       "b"
locals (1) for 000002853D5177F0:
      0       a       2       5
upvalues (1) for 000002853D5177F0:
      0       _ENV    1       0

  function <(string):1,1> (5 instructions at 000002853D517920)
  0 params, 2 slots, 1 upvalue, 0 locals, 1 constant, 0 functions
        1       [1]     LOADK           0 -1    ; 1
        2       [1]     SETUPVAL        0 0     ; a
        3       [1]     GETUPVAL        0 0     ; a
        4       [1]     RETURN          0 2
        5       [1]     RETURN          0 1
  constants (1) for 000002853D517920:
        1       1
  locals (0) for 000002853D517920:
  upvalues (1) for 000002853D517920:
        0       a       1       0
```

In the main chunk, the local a starts as a nil. The CLOSURE instruction in line [2] then instantiates a function closure with a single upvalue, a. In line [3] the closure is assigned to global b via the SETTABUP instruction.

In function b, there is a single upvalue, *a*. In Pascal, a variable in an outer scope is found by traversing stack frames. However, instantiations of Lua functions are first-class values, and they may be assigned to a variable and referenced elsewhere. Moreover, a single prototype may have multiple instantiations. Managing upvalues thus becomes a little more tricky than traversing stack frames in Pascal. The Lua virtual machine solution is to provide a clean interface to access upvalues via GETUPVAL and SETUPVAL, while the management of upvalues is handled by the virtual machine itself.

Line [2] in function b sets upvalue a (upvalue number 0 in the upvalue table) to a number value of 1 (held in temporary register 0.) In line [3], the value in upvalue a is retrieved and placed into register 0, where the following RETURN instruction will use it as a return value. The RETURN in line [5] is unused.

2.16 OP_NEWTABLE instruction

2.16.1 Syntax

```
NEWTABLE A B C   R(A) := {} (size = B,C)
```

2.16.2 Description

Creates a new empty table at register R(A). B and C are the encoded size information for the array part and the hash part of the table, respectively. Appropriate values for B and C are set in order to avoid rehashing when initially populating the table with array values or hash key-value pairs.

Operand B and C are both encoded as a 'floating point byte' (so named in lobject.c) which is `eeeeexxx` in binary, where x is the mantissa and e is the exponent. The actual value is calculated as `1xxx*2^(eeeee-1)` if `eeeee` is greater than `0` (a range of `8` to `15*2^30`). If `eeeee` is `0`, the actual value is `xxx` (a range of `0` to `7`.)

If an empty table is created, both sizes are zero. If a table is created with a number of objects, the code generator counts the number of array elements and the number of hash elements. Then, each size value is rounded up and encoded in B and C using the floating point byte format.

2.16.3 Examples

Creating an empty table forces both array and hash sizes to be zero:

```
f=load('local q = {}')
```

Leads to:

```
main <(string):0,0> (2 instructions at 0000022C1877A220)
0+ params, 2 slots, 1 upvalue, 1 local, 0 constants, 0 functions
      1       [1]     NEWTABLE        0 0 0
      2       [1]     RETURN          0 1
constants (0) for 0000022C1877A220:
locals (1) for 0000022C1877A220:
      0       q       2       3
upvalues (1) for 0000022C1877A220:
      0       _ENV    1       0
```

More examples are provided in the description of `OP_SETLIST` instruction.

2.17 OP_SETLIST instruction

2.17.1 Syntax

```
SETLIST A B C   R(A)[(C-1)*FPF+i] := R(A+i), 1 <= i <= B
```

2.17.2 Description

Sets the values for a range of array elements in a table referenced by R(A). Field B is the number of elements to set. Field C encodes the block number of the table to be initialized. The values used to initialize the table are located in registers R(A+1), R(A+2), and so on.

The block size is denoted by FPF. FPF is 'fields per flush', defined as `LFIELDS_PER_FLUSH` in the source file lopcodes.h, with a value of 50. For example, for array locations 1 to 20, C will be 1 and B will be 20.

If B is 0, the table is set with a variable number of array elements, from register R(A+1) up to the top of the stack. This happens when the last element in the table constructor is a function call or a vararg operator.

If C is 0, the next instruction is cast as an integer, and used as the C value. This happens only when operand C is unable to encode the block number, i.e. when C > 511, equivalent to an array index greater than 25550.

2.17.3 Examples

We'll start with a simple example:

```
f=load('local q = {1,2,3,4,5,}')
```

This generates:

```
main <(string):0,0> (8 instructions at 0000022C18756E50)
0+ params, 6 slots, 1 upvalue, 1 local, 5 constants, 0 functions
      1       [1]     NEWTABLE        0 5 0
      2       [1]     LOADK           1 -1    ; 1
      3       [1]     LOADK           2 -2    ; 2
      4       [1]     LOADK           3 -3    ; 3
      5       [1]     LOADK           4 -4    ; 4
      6       [1]     LOADK           5 -5    ; 5
      7       [1]     SETLIST         0 5 1   ; 1
      8       [1]     RETURN          0 1
constants (5) for 0000022C18756E50:
      1       1
      2       2
      3       3
      4       4
      5       5
locals (1) for 0000022C18756E50:
      0       q       8       9
upvalues (1) for 0000022C18756E50:
      0       _ENV    1       0
```

A table with the reference in register 0 is created in line [1] by NEWTABLE. Since we are creating a table with no hash elements, the array part of the table has a size of 5, while the hash part has a size of 0.

Constants are then loaded into temporary registers 1 to 5 (lines [2] to [6]) before the SETLIST instruction in line [7] assigns each value to consecutive table elements. The start of the block is encoded as 1 in operand C. The starting index is calculated as (1-1)*50+1 or 1. Since B is 5, the range of the array elements to be set becomes 1 to 5, while the objects used to set the array elements will be R(1) through R(5).

Next is a larger table with 55 array elements. This will require two blocks to initialize. Some lines have been removed and ellipsis (...) added to save space:

```
> f=load('local q = {1,2,3,4,5,6,7,8,9,0,1,2,3,4,5,6,7,8,9,0, \
>> 1,2,3,4,5,6,7,8,9,0,1,2,3,4,5,6,7,8,9,0, \
>> 1,2,3,4,5,6,7,8,9,0,1,2,3,4,5,}')
```

The generated code is:

```
main <(string):0,0> (59 instructions at 0000022C187833C0)
0+ params, 51 slots, 1 upvalue, 1 local, 10 constants, 0 functions
      1       [1]     NEWTABLE        0 30 0
      2       [1]     LOADK           1 -1    ; 1
```

```
      3       [1]     LOADK           2 -2    ; 2
      4       [1]     LOADK           3 -3    ; 3
      ...
      51      [3]     LOADK           50 -10  ; 0
      52      [3]     SETLIST         0 50 1  ; 1
      53      [3]     LOADK           1 -1    ; 1
      54      [3]     LOADK           2 -2    ; 2
      55      [3]     LOADK           3 -3    ; 3
      56      [3]     LOADK           4 -4    ; 4
      57      [3]     LOADK           5 -5    ; 5
      58      [3]     SETLIST         0 5 2   ; 2
      59      [3]     RETURN          0 1
constants (10) for 0000022C187833C0:
      1       1
      2       2
      3       3
      4       4
      5       5
      6       6
      7       7
      8       8
      9       9
      10      0
locals (1) for 0000022C187833C0:
      0       q       59      60
upvalues (1) for 0000022C187833C0:
      0       _ENV    1       0
```

Since FPF is 50, the array will be initialized in two blocks. The first block is for index 1 to 50, while the second block is for index 51 to 55. Each array block to be initialized requires one `SETLIST` instruction. On line [1], `NEWTABLE` has a field B value of 30, or 00011110 in binary. From the description of `NEWTABLE`, `xxx` is `1102`, while `eeeee` is `112`. Thus, the size of the array portion of the table is `(1110)*2^(11-1)` or `(14*2^2)` or `56`.

Lines [2] to [51] sets the values used to initialize the first block. On line [52], `SETLIST` has a B value of 50 and a C value of 1. So the block is from 1 to 50. Source registers are from R(1) to R(50).

Lines [53] to [57] sets the values used to initialize the second block. On line [58], `SETLIST` has a B value of 5 and a C value of 2. So the block is from 51 to 55. The start of the block is calculated as `(2-1)*50+1` or `51`. Source registers are from R(1) to R(5).

Here is a table with hashed elements:

```
> f=load('local q = {a=1,b=2,c=3,d=4,e=5,f=6,g=7,h=8,}')
```

This results in:

```
main <(string):0,0> (10 instructions at 0000022C18783D20)
0+ params, 2 slots, 1 upvalue, 1 local, 16 constants, 0 functions
      1       [1]     NEWTABLE        0 0 8
      2       [1]     SETTABLE        0 -1 -2 ; "a" 1
      3       [1]     SETTABLE        0 -3 -4 ; "b" 2
      4       [1]     SETTABLE        0 -5 -6 ; "c" 3
      5       [1]     SETTABLE        0 -7 -8 ; "d" 4
      6       [1]     SETTABLE        0 -9 -10        ; "e" 5
      7       [1]     SETTABLE        0 -11 -12       ; "f" 6
      8       [1]     SETTABLE        0 -13 -14       ; "g" 7
      9       [1]     SETTABLE        0 -15 -16       ; "h" 8
      10      [1]     RETURN          0 1
constants (16) for 0000022C18783D20:
```

```
      1       "a"
      2       1
      3       "b"
      4       2
      5       "c"
      6       3
      7       "d"
      8       4
      9       "e"
      10      5
      11      "f"
      12      6
      13      "g"
      14      7
      15      "h"
      16      8
locals (1) for 0000022C18783D20:
      0       q       10      11
upvalues (1) for 0000022C18783D20:
      0       _ENV    1       0
```

In line [1], `NEWTABLE` is executed with an array part size of 0 and a hash part size of 8.

On lines [2] to line [9], key-value pairs are set using `SETTABLE`. The `SETLIST` instruction is only for initializing array elements. Using `SETTABLE` to initialize the key-value pairs of a table in the above example is quite efficient as it can reference the constant pool directly.

If there are both array elements and hash elements in a table constructor, both `SETTABLE` and `SETLIST` will be used to initialize the table after the initial `NEWTABLE`. In addition, if the last element of the table constructor is a function call or a vararg operator, then the B operand of `SETLIST` will be 0, to allow objects from R(A+1) up to the top of the stack to be initialized as array elements of the table.

```
> f=load('return {1,2,3,a=1,b=2,c=3,foo()}')
```

Leads to:

```
main <(string):0,0> (12 instructions at 0000022C18788430)
0+ params, 5 slots, 1 upvalue, 0 locals, 7 constants, 0 functions
      1       [1]     NEWTABLE        0 3 3
      2       [1]     LOADK           1 -1    ; 1
      3       [1]     LOADK           2 -2    ; 2
      4       [1]     LOADK           3 -3    ; 3
      5       [1]     SETTABLE        0 -4 -1 ; "a" 1
      6       [1]     SETTABLE        0 -5 -2 ; "b" 2
      7       [1]     SETTABLE        0 -6 -3 ; "c" 3
      8       [1]     GETTABUP        4 0 -7  ; _ENV "foo"
      9       [1]     CALL            4 1 0
      10      [1]     SETLIST         0 0 1   ; 1
      11      [1]     RETURN          0 2
      12      [1]     RETURN          0 1
constants (7) for 0000022C18788430:
      1       1
      2       2
      3       3
      4       "a"
      5       "b"
      6       "c"
      7       "foo"
locals (0) for 0000022C18788430:
```

```
upvalues (1) for 0000022C18788430:
      0       _ENV    1       0
```

In the above example, the table is first created in line [1] with its reference in register 0, and it has both array and hash elements to be set. The size of the array part is 3 while the size of the hash part is also 3.

Lines [2]–[4] loads the values for the first 3 array elements. Lines [5]–[7] set the 3 key-value pairs for the hash part of the table. In lines [8] and [9], the call to function `foo` is made, and then in line [10], the `SETLIST` instruction sets the first 3 array elements (in registers 1 to 3) plus whatever additional results returned by the `foo` function call (from register 4 onwards). This is accomplished by setting operand B in `SETLIST` to 0. For the first block, operand C is 1 as usual. If no results are returned by the function, the top of stack is at register 3 and only the 3 constant array elements in the table are set.

Finally:

```
> f=load('local a; return {a(), a(), a()}')
```

This gives:

```
main <(string):0,0> (11 instructions at 0000022C18787AD0)
0+ params, 5 slots, 1 upvalue, 1 local, 0 constants, 0 functions
      1       [1]     LOADNIL         0 0
      2       [1]     NEWTABLE        1 2 0
      3       [1]     MOVE            2 0
      4       [1]     CALL            2 1 2
      5       [1]     MOVE            3 0
      6       [1]     CALL            3 1 2
      7       [1]     MOVE            4 0
      8       [1]     CALL            4 1 0
      9       [1]     SETLIST         1 0 1   ; 1
      10      [1]     RETURN          1 2
      11      [1]     RETURN          0 1
constants (0) for 0000022C18787AD0:
locals (1) for 0000022C18787AD0:
      0       a       2       12
upvalues (1) for 0000022C18787AD0:
      0       _ENV    1       0
```

Note that only the last function call in a table constructor retains all results. Other function calls in the table constructor keep only one result. This is shown in the above example. For vararg operators in table constructors, please see the discussion for the `VARARG` instruction for an example.

2.18 OP_GETTABLE and OP_SETTABLE instructions

2.18.1 Syntax

```
GETTABLE A B C   R(A) := R(B)[RK(C)]
SETTABLE A B C   R(A)[RK(B)] := RK(C)
```

2.18.2 Description

`OP_GETTABLE` copies the value from a table element into register R(A). The table is referenced by register R(B), while the index to the table is given by RK(C), which may be the value of register R(C) or a constant number.

OP_SETTABLE copies the value from register R(C) or a constant into a table element. The table is referenced by register R(A), while the index to the table is given by RK(B), which may be the value of register R(B) or a constant number.

All 3 operand fields are used, and some of the operands can be constants. A constant is specified by setting the MSB of the operand to 1. If RK(C) need to refer to constant 1, the encoded value will be (256 | 1) or 257, where 256 is the value of bit 8 of the operand. Allowing constants to be used directly reduces considerably the need for temporary registers.

2.18.3 Examples

```
f=load('local p = {}; p[1] = "foo"; return p["bar"]')
```

This compiles to:

```
main <(string):0,0> (5 instructions at 000001FA06FCC3F0)
0+ params, 2 slots, 1 upvalue, 1 local, 3 constants, 0 functions
      1       [1]     NEWTABLE        0 0 0
      2       [1]     SETTABLE        0 -1 -2 ; 1 "foo"
      3       [1]     GETTABLE        1 0 -3  ; "bar"
      4       [1]     RETURN          1 2
      5       [1]     RETURN          0 1
constants (3) for 000001FA06FCC3F0:
      1       1
      2       "foo"
      3       "bar"
locals (1) for 000001FA06FCC3F0:
      0       p       2       6
upvalues (1) for 000001FA06FCC3F0:
      0       _ENV    1       0
```

In line [1], a new empty table is created and the reference placed in local p (register 0). Creating and populating new tables is discussed in detail elsewhere. Table index 1 is set to 'foo' in line [2] by the SETTABLE instruction.

The R(A) value of 0 points to the new table that was defined in line [1]. In line [3], the value of the table element indexed by the string 'bar' is copied into temporary register 1, which is then used by RETURN as a return value.

2.19 OP_SELF instruction

2.19.1 Syntax

```
SELF  A B C   R(A+1) := R(B); R(A) := R(B)[RK(C)]
```

2.19.2 Description

For object-oriented programming using tables. Retrieves a function reference from a table element and places it in register R(A), then a reference to the table itself is placed in the next register, R(A+1). This instruction saves some messy manipulation when setting up a method call.

R(B) is the register holding the reference to the table with the method. The method function itself is found using the table index RK(C), which may be the value of register R(C) or a constant number.

2.19.3 Examples

A SELF instruction saves an extra instruction and speeds up the calling of methods in object oriented programming. It is only generated for method calls that use the colon syntax. In the following example:

```
f=load('foo:bar("baz")')
```

We can see SELF being generated:

```
main <(string):0,0> (5 instructions at 000001FA06FA7830)
0+ params, 3 slots, 1 upvalue, 0 locals, 3 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "foo"
      2       [1]     SELF            0 0 -2  ; "bar"
      3       [1]     LOADK           2 -3    ; "baz"
      4       [1]     CALL            0 3 1
      5       [1]     RETURN          0 1
constants (3) for 000001FA06FA7830:
      1       "foo"
      2       "bar"
      3       "baz"
locals (0) for 000001FA06FA7830:
upvalues (1) for 000001FA06FA7830:
      0       _ENV    1       0
```

The method call is equivalent to: foo.bar(foo, "baz"), except that the global foo is only looked up once. This is significant if metamethods have been set. The SELF in line [2] is equivalent to a GETTABLE lookup (the table is in register 0 and the index is constant 1) and a MOVE (copying the table reference from register 0 to register 1.)

Without SELF, a GETTABLE will write its lookup result to register 0 (which the code generator will normally do) and the table reference will be overwritten before a MOVE can be done. Using SELF saves roughly one instruction and one temporary register slot.

After setting up the method call using SELF, the call is made with the usual CALL instruction in line [4], with two parameters. The equivalent code for a method lookup is compiled in the following manner:

```
f=load('foo.bar(foo, "baz")')
```

And generated code:

```
main <(string):0,0> (6 instructions at 000001FA06FA6960)
0+ params, 3 slots, 1 upvalue, 0 locals, 3 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "foo"
      2       [1]     GETTABLE        0 0 -2  ; "bar"
      3       [1]     GETTABUP        1 0 -1  ; _ENV "foo"
      4       [1]     LOADK           2 -3    ; "baz"
      5       [1]     CALL            0 3 1
      6       [1]     RETURN          0 1
constants (3) for 000001FA06FA6960:
      1       "foo"
      2       "bar"
      3       "baz"
locals (0) for 000001FA06FA6960:
upvalues (1) for 000001FA06FA6960:
      0       _ENV    1       0
```

The alternative form of a method call is one instruction longer, and the user must take note of any metamethods that may affect the call. The SELF in the previous example replaces the GETTABLE on line [2] and the GETTABUP on line [3]. If foo is a local variable, then the equivalent code is a GETTABLE and a MOVE.

2.20 OP_GETTABUP and OP_SETTABUP instructions

2.20.1 Syntax

```
GETTABUP A B C   R(A) := UpValue[B][RK(C)]
SETTABUP A B C   UpValue[A][RK(B)] := RK(C)
```

2.20.2 Description

`OP_GETTABUP` and `OP_SETTABUP` instructions are similar to the `OP_GETTABLE` and `OP_SETTABLE` instructions except that the table is referenced as an upvalue. These instructions are used to access global variables, which since Lua 5.2 are accessed via the upvalue named `_ENV`.

2.20.3 Examples

```
f=load('a = 40; local b = a')
```

Results in:

```
main <(string):0,0> (3 instructions at 0000028D955FEBF0)
0+ params, 2 slots, 1 upvalue, 1 local, 2 constants, 0 functions
      1       [1]     SETTABUP        0 -1 -2 ; _ENV "a" 40
      2       [1]     GETTABUP        0 0 -1  ; _ENV "a"
      3       [1]     RETURN          0 1
constants (2) for 0000028D955FEBF0:
      1       "a"
      2       40
locals (1) for 0000028D955FEBF0:
      0       b       3       4
upvalues (1) for 0000028D955FEBF0:
      0       _ENV    1       0
```

From the example, we can see that 'b' is the name of the local variable while 'a' is the name of the global variable.

Line [1] assigns the number 40 to global 'a'. Line [2] assigns the value in global 'a' to the register 0 which is the local 'b'.

2.21 OP_CONCAT instruction

2.21.1 Syntax

```
CONCAT A B C   R(A) := R(B).. ... ..R(C)
```

2.21.2 Description

Performs concatenation of two or more strings. In a Lua source, this is equivalent to one or more concatenation operators ('..') between two or more expressions. The source registers must be consecutive, and C must always be greater than B. The result is placed in R(A).

2.21.3 Examples

CONCAT accepts a range of registers. Doing more than one string concatenation at a time is faster and more efficient than doing them separately:

```
f=load('local x,y = "foo","bar"; return x..y..x..y')
```

Generates:

```
main <(string):0,0> (9 instructions at 0000028D9560B290)
0+ params, 6 slots, 1 upvalue, 2 locals, 2 constants, 0 functions
      1       [1]     LOADK           0 -1    ; "foo"
      2       [1]     LOADK           1 -2    ; "bar"
      3       [1]     MOVE            2 0
      4       [1]     MOVE            3 1
      5       [1]     MOVE            4 0
      6       [1]     MOVE            5 1
      7       [1]     CONCAT          2 2 5
      8       [1]     RETURN          2 2
      9       [1]     RETURN          0 1
constants (2) for 0000028D9560B290:
      1       "foo"
      2       "bar"
locals (2) for 0000028D9560B290:
      0       x       3       10
      1       y       3       10
upvalues (1) for 0000028D9560B290:
      0       _ENV    1       0
```

In this example, strings are moved into place first (lines [3] to [6]) in the concatenation order before a single CONCAT instruction is executed in line [7]. The result is left in temporary local 2, which is then used as a return value by the RETURN instruction on line [8].

```
f=load('local a = "foo".."bar".."baz"')
```

Compiles to:

```
main <(string):0,0> (5 instructions at 0000028D9560EE40)
0+ params, 3 slots, 1 upvalue, 1 local, 3 constants, 0 functions
      1       [1]     LOADK           0 -1    ; "foo"
      2       [1]     LOADK           1 -2    ; "bar"
      3       [1]     LOADK           2 -3    ; "baz"
      4       [1]     CONCAT          0 0 2
      5       [1]     RETURN          0 1
constants (3) for 0000028D9560EE40:
      1       "foo"
      2       "bar"
      3       "baz"
locals (1) for 0000028D9560EE40:
      0       a       5       6
upvalues (1) for 0000028D9560EE40:
      0       _ENV    1       0
```

In the second example, three strings are concatenated together. Note that there is no string constant folding. Lines [1] through [3] loads the three constants in the correct order for concatenation; the CONCAT on line [4] performs the concatenation itself and assigns the result to local 'a'.

2.22 OP_LEN instruction

2.22.1 Syntax

```
LEN A B     R(A) := length of R(B)
```

2.22.2 Description

Returns the length of the object in R(B). For strings, the string length is returned, while for tables, the table size (as defined in Lua) is returned. For other objects, the metamethod is called. The result, which is a number, is placed in R(A).

2.22.3 Examples

The `LEN` operation implements the # operator. If # operates on a constant, then the constant is loaded in advance using `LOADK`. The `LEN` instruction is currently not optimized away using compile time evaluation, even if it is operating on a constant string or table:

```
f=load('local a,b; a = #b; a= #"foo"')
```

Results in:

```
main <(string):0,0> (5 instructions at 000001DC21778C60)
0+ params, 3 slots, 1 upvalue, 2 locals, 1 constant, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     LEN             0 1
      3       [1]     LOADK           2 -1    ; "foo"
      4       [1]     LEN             0 2
      5       [1]     RETURN          0 1
constants (1) for 000001DC21778C60:
      1       "foo"
locals (2) for 000001DC21778C60:
      0       a       2       6
      1       b       2       6
upvalues (1) for 000001DC21778C60:
      0       _ENV    1       0
```

In the above example, `LEN` operates on local b in line [2], leaving the result in local a. Since `LEN` cannot operate directly on constants, line [3] first loads the constant "foo" into a temporary local, and only then `LEN` is executed.

2.23 OP_MOVE instruction

2.23.1 Syntax

```
MOVE A B     R(A) := R(B)
```

2.23.2 Description

Copies the value of register R(B) into register R(A). If R(B) holds a table, function or userdata, then the reference to that object is copied. `MOVE` is often used for moving values into place for the next operation.

2.23.3 Examples

The most straightforward use of MOVE is for assigning a local to another local:

```
f=load('local a,b = 10; b = a')
```

Produces:

```
main <(string):0,0> (4 instructions at 000001DC217566D0)
0+ params, 2 slots, 1 upvalue, 2 locals, 1 constant, 0 functions
      1       [1]     LOADK           0 -1    ; 10
      2       [1]     LOADNIL         1 0
      3       [1]     MOVE            1 0
      4       [1]     RETURN          0 1
constants (1) for 000001DC217566D0:
      1       10
locals (2) for 000001DC217566D0:
      0       a       3       5
      1       b       3       5
upvalues (1) for 000001DC217566D0:
      0       _ENV    1       0
```

You won't see `MOVE` instructions used in arithmetic expressions because they are not needed by arithmetic operators. All arithmetic operators are in 2- or 3-operand style: the entire local stack frame is already visible to operands R(A), R(B) and R(C) so there is no need for any extra `MOVE` instructions.

Other places where you will see `MOVE` are:

- When moving parameters into place for a function call.
- When moving values into place for certain instructions where stack order is important, e.g. `GETTABLE`, `SETTABLE` and `CONCAT`.
- When copying return values into locals after a function call.

2.24 OP_LOADNIL instruction

2.24.1 Syntax

```
LOADNIL A B     R(A), R(A+1), ..., R(A+B) := nil
```

2.24.2 Description

Sets a range of registers from R(A) to R(B) to nil. If a single register is to be assigned to, then R(A) = R(B). When two or more consecutive locals need to be assigned nil values, only a single `LOADNIL` is needed.

2.24.3 Examples

`LOADNIL` uses the operands A and B to mean a range of register locations. The example for `MOVE` earlier shows `LOADNIL` used to set a single register to `nil`.

```
f=load('local a,b,c,d,e = nil,nil,0')
```

Generates:

```
main <(string):0,0> (4 instructions at 000001DC21780390)
0+ params, 5 slots, 1 upvalue, 5 locals, 1 constant, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     LOADK           2 -1    ; 0
      3       [1]     LOADNIL         3 1
      4       [1]     RETURN          0 1
constants (1) for 000001DC21780390:
      1       0
locals (5) for 000001DC21780390:
      0       a       4       5
      1       b       4       5
      2       c       4       5
      3       d       4       5
      4       e       4       5
upvalues (1) for 000001DC21780390:
      0       _ENV    1       0
```

Line [1] nils locals a and b. Local c is explicitly initialized with the value 0. Line [3] nils d and e.

2.25 OP_LOADK instruction

2.25.1 Syntax

```
LOADK A Bx    R(A) := Kst(Bx)
```

2.25.2 Description

Loads constant number Bx into register R(A). Constants are usually numbers or strings. Each function prototype has its own constant list, or pool.

2.25.3 Examples

`LOADK` loads a constant from the constant list into a register or local. Constants are indexed starting from 0. Some instructions, such as arithmetic instructions, can use the constant list without needing a `LOADK`. Constants are pooled in the list, duplicates are eliminated. The list can hold nils, booleans, numbers or strings.

```
f=load('local a,b,c,d = 3,"foo",3,"foo"')
```

Leads to:

```
main <(string):0,0> (5 instructions at 000001DC21780B50)
0+ params, 4 slots, 1 upvalue, 4 locals, 2 constants, 0 functions
      1       [1]     LOADK           0 -1    ; 3
      2       [1]     LOADK           1 -2    ; "foo"
      3       [1]     LOADK           2 -1    ; 3
      4       [1]     LOADK           3 -2    ; "foo"
      5       [1]     RETURN          0 1
constants (2) for 000001DC21780B50:
      1       3
      2       "foo"
locals (4) for 000001DC21780B50:
      0       a       5       6
      1       b       5       6
```

```
      2       c       5       6
      3       d       5       6
upvalues (1) for 000001DC21780B50:
      0       _ENV    1       0
```

The constant 3 and the constant "foo" are both written twice in the source snippet, but in the constant list, each constant has a single location.

2.26 Binary operators

Lua 5.3 implements a bunch of binary operators for arithmetic and bitwise manipulation of variables. These insructions have a common form.

2.26.1 Syntax

```
ADD   A B C   R(A) := RK(B) + RK(C)
SUB   A B C   R(A) := RK(B) - RK(C)
MUL   A B C   R(A) := RK(B) * RK(C)
MOD   A B C   R(A) := RK(B) % RK(C)
POW   A B C   R(A) := RK(B) ^ RK(C)
DIV   A B C   R(A) := RK(B) / RK(C)
IDIV  A B C   R(A) := RK(B) // RK(C)
BAND  A B C   R(A) := RK(B) & RK(C)
BOR   A B C   R(A) := RK(B) | RK(C)
BXOR  A B C   R(A) := RK(B) ~ RK(C)
SHL   A B C   R(A) := RK(B) << RK(C)
SHR   A B C   R(A) := RK(B) >> RK(C)
```

2.26.2 Description

Binary operators (arithmetic operators and bitwise operators with two inputs.) The result of the operation between RK(B) and RK(C) is placed into R(A). These instructions are in the classic 3-register style.

RK(B) and RK(C) may be either registers or constants in the constant pool.

Opcode	Description
ADD	Addition operator
SUB	Subtraction operator
MUL	Multiplication operator
MOD	Modulus (remainder) operator
POW	Exponentation operator
DIV	Division operator
IDIV	Integer division operator
BAND	Bit-wise AND operator
BOR	Bit-wise OR operator
BXOR	Bit-wise Exclusive OR operator
SHL	Shift bits left
SHR	Shift bits right

The source operands, RK(B) and RK(C), may be constants. If a constant is out of range of field B or field C, then the constant will be loaded into a temporary register in advance.

2.26.3 Examples

```
f=load('local a,b = 2,4; a = a + 4 * b - a / 2 ^ b % 3')
```

Generates:

```
main <(string):0,0> (9 instructions at 000001DC21781DD0)
0+ params, 4 slots, 1 upvalue, 2 locals, 3 constants, 0 functions
      1       [1]     LOADK           0 -1    ; 2
      2       [1]     LOADK           1 -2    ; 4
      3       [1]     MUL             2 -2 1  ; 4 -       (loc2 = 4 * b)
      4       [1]     ADD             2 0 2               (loc2 = A + loc2)
      5       [1]     POW             3 -1 1  ; 2 -       (loc3 = 2 ^ b)
      6       [1]     DIV             3 0 3               (loc3 = a / loc3)
      7       [1]     MOD             3 3 -3              (loc3 = loc3 % 3)
      8       [1]     SUB             0 2 3               (a = loc2 - loc3)
      9       [1]     RETURN          0 1
constants (3) for 000001DC21781DD0:
      1       2
      2       4
      3       3
locals (2) for 000001DC21781DD0:
      0       a       3       10
      1       b       3       10
upvalues (1) for 000001DC21781DD0:
      0       _ENV    1       0
```

In the disassembly shown above, parts of the expression is shown as additional comments in parentheses. Each arithmetic operator translates into a single instruction. This also means that while the statement `count = count + 1` is verbose, it translates into a single instruction if count is a local. If count is a global, then two extra instructions are required to read and write to the global (`GETTABUP` and `SETTABUP`), since arithmetic operations can only be done on registers (locals) only.

The Lua parser and code generator can perform limited constant expression folding or evaluation. Constant folding only works for binary arithmetic operators and the unary minus operator (`UNM`, which will be covered next.) There is no equivalent optimization for relational, boolean or string operators.

The optimization rule is simple: If both terms of a subexpression are numbers, the subexpression will be evaluated at compile time. However, there are exceptions. One, the code generator will not attempt to divide a number by 0 for DIV and MOD, and two, if the result is evaluated as a NaN (Not a Number) then the optimization will not be performed.

Also, constant folding is not done if one term is in the form of a string that need to be coerced. In addition, expression terms are not rearranged, so not all optimization opportunities can be recognized by the code generator. This is intentional; the Lua code generator is not meant to perform heavy duty optimizations, as Lua is a lightweight language. Here are a few examples to illustrate how it works (additional comments in parentheses):

```
f=load('local a = 4 + 7 + b; a = b + 4 * 7; a = b + 4 + 7')
```

Generates:

```
main <(string):0,0> (8 instructions at 000001DC21781650)
0+ params, 2 slots, 1 upvalue, 1 local, 5 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "b"
      2       [1]     ADD             0 -2 0  ; 11 -              (a = 11 + b)
      3       [1]     GETTABUP        1 0 -1  ; _ENV "b"
      4       [1]     ADD             0 1 -3  ; - 28              (a = b + 28)
      5       [1]     GETTABUP        1 0 -1  ; _ENV "b"
      6       [1]     ADD             1 1 -4  ; - 4               (loc1 = b + 4)
```

```
      7       [1]     ADD             0 1 -5  ; - 7             (a = loc1 + 7)
      8       [1]     RETURN          0 1
constants (5) for 000001DC21781650:
      1       "b"
      2       11
      3       28
      4       4
      5       7
locals (1) for 000001DC21781650:
      0       a       3       9
upvalues (1) for 000001DC21781650:
      0       _ENV    1       0
```

For the first assignment statement, `4+7` is evaluated, thus 11 is added to b in line [2]. Next, in line [3] and [4], `b` and `28` are added together and assigned to a because multiplication has a higher precedence and `4*7` is evaluated first. Finally, on lines [5] to [7], there are two addition operations. Since addition is left-associative, code is generated for `b+4` first, and only after that, `7` is added. So in the third example, Lua performs no optimization. This can be fixed using parentheses to explicitly change the precedence of a subexpression:

```
f=load('local a = b + (4 + 7)')
```

And this leads to:

```
main <(string):0,0> (3 instructions at 000001DC21781EC0)
0+ params, 2 slots, 1 upvalue, 1 local, 2 constants, 0 functions
      1       [1]     GETTABUP        0 0 -1  ; _ENV "b"
      2       [1]     ADD             0 0 -2  ; - 11
      3       [1]     RETURN          0 1
constants (2) for 000001DC21781EC0:
      1       "b"
      2       11
locals (1) for 000001DC21781EC0:
      0       a       3       4
upvalues (1) for 000001DC21781EC0:
      0       _ENV    1       0
```

Now, the `4+7` subexpression can be evaluated at compile time. If the statement is written as:

```
local a = 7 + (4 + 7)
```

the code generator will generate a single `LOADK` instruction; Lua first evaluates `4+7`, then `7` is added, giving a total of `18`. The arithmetic expression is completely evaluated in this case, thus no arithmetic instructions are generated.

In order to make full use of constant folding in Lua, the user just need to remember the usual order of evaluation of an expression's elements and apply parentheses where necessary. The following are two expressions which will not be evaluated at compile time:

```
f=load('local a = 1 / 0; local b = 1 + "1"')
```

This produces:

```
main <(string):0,0> (3 instructions at 000001DC21781380)
0+ params, 2 slots, 1 upvalue, 2 locals, 3 constants, 0 functions
      1       [1]     DIV             0 -2 -1 ; 1 0
      2       [1]     ADD             1 -2 -3 ; 1 "1"
      3       [1]     RETURN          0 1
constants (3) for 000001DC21781380:
      1       0
      2       1
      3       "1"
```

```
locals (2) for 000001DC21781380:
      0       a       2       4
      1       b       3       4
upvalues (1) for 000001DC21781380:
      0       _ENV    1       0
```

The first is due to a divide-by-0, while the second is due to a string constant that needs to be coerced into a number. In both cases, constant folding is not performed, so the arithmetic instructions needed to perform the operations at run time are generated instead.

TODO - examples of bitwise operators.

2.27 Unary operators

Lua 5.3 implements following unary operators in addition to `OP_LEN`.

2.27.1 Syntax

```
UNM   A B     R(A) := -R(B)
BNOT  A B     R(A) := ~R(B)
NOT   A B     R(A) := not R(B)
```

2.27.2 Description

The unary operators perform an operation on R(B) and store the result in R(A).

Opcode	Description
UNM	Unary minus
BNOT	Bit-wise NOT operator
NOT	Logical NOT operator

2.27.3 Examples

```
f=load('local p,q = 10,false; q,p = -p,not q')
```

Results in:

```
main <(string):0,0> (6 instructions at 000001DC21781290)
0+ params, 3 slots, 1 upvalue, 2 locals, 1 constant, 0 functions
      1       [1]     LOADK           0 -1    ; 10
      2       [1]     LOADBOOL        1 0 0
      3       [1]     UNM             2 0
      4       [1]     NOT             0 1
      5       [1]     MOVE            1 2
      6       [1]     RETURN          0 1
constants (1) for 000001DC21781290:
      1       10
locals (2) for 000001DC21781290:
      0       p       3       7
      1       q       3       7
upvalues (1) for 000001DC21781290:
      0       _ENV    1       0
```

As UNM and NOT do not accept a constant as a source operand, making the LOADK on line [1] and the LOADBOOL on line [2] necessary. When an unary minus is applied to a constant number, the unary minus is optimized away. Similarly, when a not is applied to true or false, the logical operation is optimized away.

In addition to this, constant folding is performed for unary minus, if the term is a number. So, the expression in the following is completely evaluated at compile time:

```
f=load('local a = - (7 / 4)')
```

Results in:

```
main <(string):0,0> (2 instructions at 000001DC217810B0)
0+ params, 2 slots, 1 upvalue, 1 local, 1 constant, 0 functions
      1       [1]     LOADK           0 -1    ; -1.75
      2       [1]     RETURN          0 1
constants (1) for 000001DC217810B0:
      1       -1.75
locals (1) for 000001DC217810B0:
      0       a       2       3
upvalues (1) for 000001DC217810B0:
      0       _ENV    1       0
```

Constant folding is performed on 7/4 first. Then, since the unary minus operator is applied to the constant 1.75, constant folding can be performed again, and the code generated becomes a simple LOADK (on line [1]).

TODO - example of BNOT.

CHAPTER 3

Lua Parsing and Code Generation Internals

3.1 Stack and Registers

Lua employs two stacks. The `Callinfo` stack tracks activation frames. There is the secondary stack `L->stack` that is an array of `TValue` objects. The `Callinfo` objects index into this array. Registers are basically slots in the `L->stack` array.

When a function is called - the stack is setup as follows:

```
stack
|             function reference
|  base->     parameter 1
|             ...
|             parameter n
|             local 1
|             ...
|             local n
|  top->
|
V
```

So top is just past the registers needed by the function. The number of registers is determined based on locals and temporaries.

The base of the stack is set to just past the function reference - i.e. on the first parameter or register. All register addressing is done as offset from base - so `R(0)` is at `base+0` on the stack.

A description of the stack and registers from Mike Pall on Lua mailing list is reproduced below.

3.1.1 Sliding Register Window - by Mike Pall

Note: this is a reformatted version of a post on Lua mailing list (see MP6 link below).

The Lua 5 VM employs a sliding register window on top of a stack. Frames (named CallInfo aka 'ci' in the source) occupy different (overlapping) ranges on the stack. Successive frames are positioned exactly over the passed arguments (luaD_precall). The compiler ensures that there are no live variables after the arguments for a call. Return values need to be copied down (with truncate/extend) to the slot holding the function object (luaD_poscall). This is because the compiler has no idea how many values another function may return – only how many need to be stored.

Example:

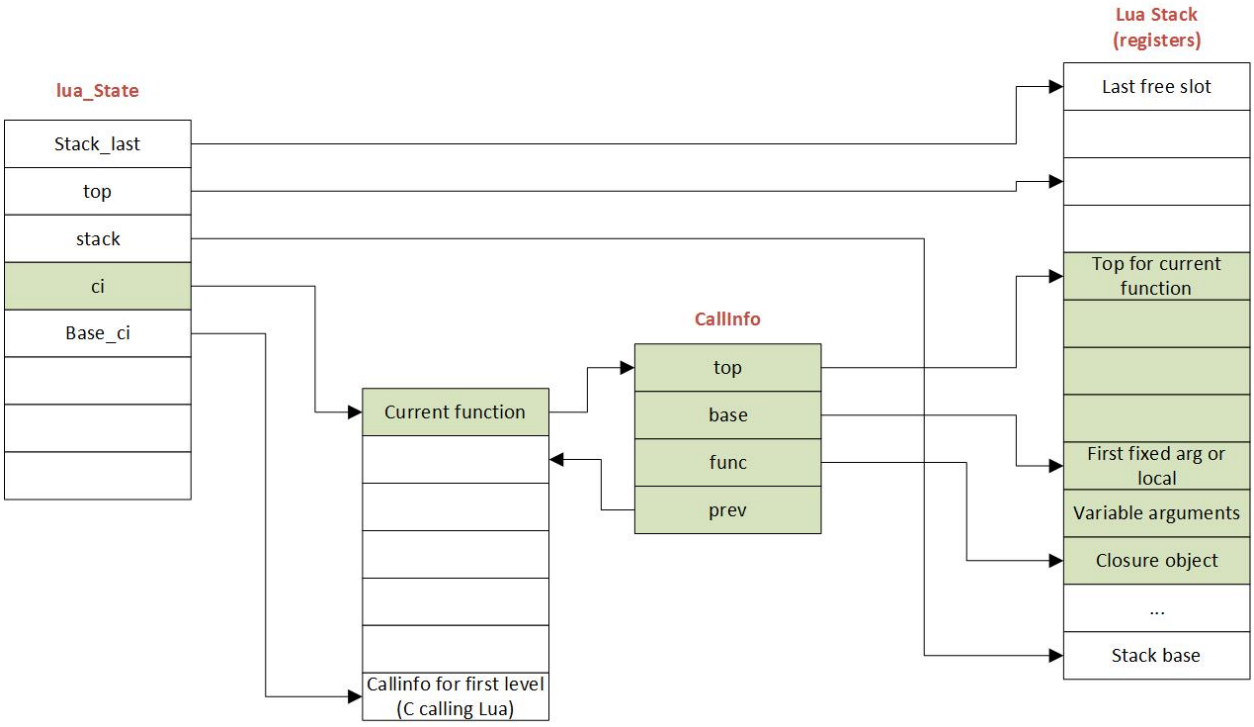

Fig. 3.1: The figure shows how the stack is related to other Lua objects.

```
function f2(arg1, arg2, ..., argN)
  local local1, local2, ...
  ...
  return ret1, ret2, ..., retO
end

function f1(arg1, arg2, ..., argM)
  local local1, local2, ...
  ...
  local ret1, ret2, ..., retP = f2(arg1, arg2, ..., argN)
  ...
end
```

Simplified stack diagram:

```
stack
|
| time: >>>> call >>>>>>>>>>>>>>>>>> call ~~~~~~~~~~~~~~~~~~ return >>>>>
|
|   ciX.func-> f1      f1      f1                                  f1
|   ciX.base-> arg1    arg1    arg1                                arg1
|              arg2    arg2    arg2                                arg2
|              ...     ...     ...                                 ...
|              argM    argM    argM                                argM
|   ciX.topC->         local1  local1                              local11
|                      local2  local2                              local2
|                      local3  local3                              local3
|                      ...     ...                                 ...
|                              f2      ciY.func-> f2      f2       ret1
```

```
|                              arg1    ciY.base-> arg1    arg1    ret2
|                              arg2               arg2    arg2    ...
|                              ...                ...     ...     retP
|                              argN               argN    argN
|   ciX.topL-> ------  ------  ------  ciY.topC-> local1  local1
|                                                 local2  local2
|                                                 ...     ...
|                                                         ret1
|                                                         ret2
|                                                         ...
|                                                         retO
|                                     ciY.topL-> ------  ------
V
```

Note that there is only a single 'top' for each frame:

For Lua functions the top (tagged topL in the diagram) is set to the base plus the maximum number of slots used. The compiler knows this and stores it in the function prototype. The top pointer is used only temporarily for handling variable length argument and return value lists.

For C functions the top (tagged topC in the diagram) is initially set to the base plus the number of passed arguments. C functions can access their part of the stack via Lua API calls which in turn change the stack top. C functions return an integer that indicates the number of return values relative to the stack top.

In reality things are a bit more complex due to overlapped locals, block scopes, varargs, coroutines and a few other things. But this should get you the basic idea.

3.2 Parsing and Code Generation

- The parser is in lparser.c.
- The code generator is in both above and lcode.c.

The parser and code generator are arguably the most complex piece in the whole of Lua. The parser is one-pass - and generates code as it parses. That is, there is no AST build phase. This is primarily for efficiency it seems. The parser uses data structures on the stack - there are no heap allocated structures. Where needed the C stack itself is used to build structures - for example, as the assignment statement is parsed, there is recursion, and a stack based structure is built that links to structures in the call stack.

The main object used by the parser is the `struct expdesc`:

```
typedef struct expdesc {
  expkind k;
  union {
    struct {  /* for indexed variables (VINDEXED) */
      short idx;  /* index (R/K) */
      lu_byte t;  /* table (register or upvalue) */
      lu_byte vt;  /* whether 't' is register (VLOCAL) or upvalue (VUPVAL) */
    } ind;
    int info;  /* for generic use */
    lua_Number nval;  /* for VKFLT */
    lua_Integer ival;    /* for VKINT */
  } u;
  int t;  /* patch list of 'exit when true' */
  int f;  /* patch list of 'exit when false' */
  int ravi_type; /* RAVI change: type of the expression if known, else LUA_TNONE */
} expdesc;
```

The code is somewhat hard to follow as the `expdesc` objects go through various states and are also reused when needed.

As the parser generates code while parsing it needs to go back and patch the generated instructions when it has more information. For example when a function call is parsed the parser assumes that only 1 value is expected to be returned - but later this is patched when more information is available. The most common example is when the register where the value will be stored (operand A) is not known - in this case the parser later on updates this operand in the instruction. I believe jump statements have similar mechanics - however I have not yet gone through the details of these instructions.

3.2.1 Handling of Stack during parsing

Functions have a register window on the stack. The stack is represented in `LexState->dyd.actvar` (Dyndata) structure (see llex.h). The register window of the function starts from `LexState->dyd.actvar.arr[firstlocal]`.

The 'active' local variables of the function extend up to `LexState->dyd.actvar.arr[nactvar-1]`. Note that when parsing a `local` declaration statement the `nactvar` is adjusted at the end of the statement so that during parsing of the statement the `nactvar` covers locals up to the start of the statement. This means that local variables come into scope (become 'active') after the local statement ends. However, if the local statement defines a function then the variable becomes 'active' before the function body is parsed.

A tricky thing to note is that while `nactvar` is adjusted at the end of the statement - the 'stack' as represented by `LexState->dyd.actvar.arr` is extended to the required size as the local variables are created by `new_localvar()`.

When a function is the topmost function being parsed, the registers between `LexState->dyd.actvar.arr[nactvar]` and `LexState->dyd.actvar.arr[freereg-1]` are used by the parser for evaluating expressions - i.e. these are part of the local registers available to the function

Note that function parameters are handled as locals.

Example of what all this mean. Let's say we are parsing following chunk of code:

```
function testfunc()
  -- at this stage 'nactvar' is 0 (no active variables)
  -- 'firstlocal' is set to current top of the variables stack
  -- LexState->dyd.actvar.n (i.e. excluding registers used for expression evaluation)
  -- LexState->dyd.actvar.n = 0 at this stage
  local function tryme()
    -- Since we are inside the local statement and 'tryme' is a local variable,
    -- the LexState->dyd.actvar.n goes to 1. As this is a function definition
    -- the local variable declaration is deemed to end here, so 'nactvar' for testfunc()
    -- is gets set to 1 (making 'tryme' an active variable).
    -- A new FuncState is created for 'tryme' function.
    -- The new tryme() FunState has 'firstlocal' set to value of LexState->dyd.actvar.n, i.e., 1
    local i,j = 5,6
    -- After 'i' is parsed, LexState->dyd.actvar.n = 2, but 'nactvar' = 0 for tryme()
    -- After 'j' is parsed, LexState->dyd.actvar.n = 3, but 'nactvar' = 0 for tryme()
    -- Only after the full statement above is parsed, 'nactvar' for tryme() is set to '2'
    -- This is done by adjustlocalvar().
    return i,j
  end
  -- Here two things happen
  -- Firstly the FuncState for tryme() is popped so that
  -- FuncState for testfunc() is now at top
  -- As part of this popping, leaveblock() calls removevars()
  -- to adjust the LexState->dyd.actvar.n down to 1 where it was
```

```
  -- at before parsing the tryme() function body.
  local i, j = tryme()
  -- After 'i' is parsed, LexState->dyd.actvar.n = 2, but 'nactvar' = 1 still
  -- After 'j' is parsed, LexState->dyd.actvar.n = 3, but 'nactvar' = 1 still
  -- At the end of the statement 'nactvar' is set to 3.
  return i+j
end
-- As before the leaveblock() calls removevars() which resets
-- LexState->dyd.actvar.n to 0 (the value before testfunc() was parsed)
```

A rough debug trace of the above gives:

```
function testfunc()
  -- open_func -> fs->firstlocal set to 0 (ls->dyd->actvar.n), and fs->nactvar reset to 0
  local function tryme()
    -- new_localvar -> registering var tryme fs->f->locvars[0] at ls->dyd->actvar.arr[0]
    -- new_localvar -> ls->dyd->actvar.n set to 1
    -- adjustlocalvars -> set fs->nactvar to 1
    -- open_func -> fs->firstlocal set to 1 (ls->dyd->actvar.n), and fs->nactvar reset to 0
    -- adjustlocalvars -> set fs->nactvar to 0 (no parameters)
    local i,j = 5,6
    -- new_localvar -> registering var i fs->f->locvars[0] at ls->dyd->actvar.arr[1]
    -- new_localvar -> ls->dyd->actvar.n set to 2
    -- new_localvar -> registering var j fs->f->locvars[1] at ls->dyd->actvar.arr[2]
    -- new_localvar -> ls->dyd->actvar.n set to 3
    -- adjustlocalvars -> set fs->nactvar to 2
    return i,j
    -- removevars -> reset fs->nactvar to 0
  end
  local i, j = tryme()
  -- new_localvar -> registering var i fs->f->locvars[1] at ls->dyd->actvar.arr[1]
  -- new_localvar -> ls->dyd->actvar.n set to 2
  -- new_localvar -> registering var j fs->f->locvars[2] at ls->dyd->actvar.arr[2]
  -- new_localvar -> ls->dyd->actvar.n set to 3
  -- adjustlocalvars -> set fs->nactvar to 3
  return i+j
  -- removevars -> reset fs->nactvar to 0
end
```

3.2.2 Notes on Parser by Sven Olsen

"discharging" expressions

"discharging" takes an expression of arbitrary type, and converts it to one having particular properties.

the lowest-level discharge function is `discharge2vars ()`, which converts an expression into one of the two "result" types; either a `VNONRELOC` or a `VRELOCABLE`.

if the variable in question is a `VLOCAL`, `discharge2vars` will simply change the stored type to `VNONRELOC`.

much of lcode.c assumes that the it will be working with discharged expressions. in particular, it assumes that if it encounters a `VNONRELOC` expression, and `e->info < nactvar`, then the register referenced is a local, and therefore shouldn't be implicitly freed after use.

local variables

however, the relationship between `nactvar` and locals is actually somewhat more complex – as each local variable appearing in the code has a collection of data attached to it, data that's being accumulated and changed as the lexer moves through the source.

`fs->nlocvars` stores the total number of named locals inside the function – recall that different local variables are allowed to overlap the same register, depending on which are in-scope at any particular time.

the list of locals that are active at any given time is stored in `ls->dyd` – a vector of stack references that grows or shrinks as locals enter or leave scope.

managing the lifetime of local variables involves several steps. first, new locals are declared using `new_localvar`. this sets their names and creates new references in `dyd`. soon thereafter, the parser is expected to call `adjustlocalvar(ls,nvars)`, with `nvars` set to the number of new locals. `adjustlocalvar` increments `fs->nactvar` by `nvars`, and marks the startpc's of all the locals.

note that neither `new_localvar` or `adjustlocalvar` ensures that anything is actually inside the registers being labeled as locals. failing to initialize said registers is an easy way to write memory access bugs (peter's original table unpack patch includes one such).

after `adjustlocalvar` is called, `luaK_exp2nextreg()` will no longer place new data inside the local's registers – as they're no longer part of the temporary register stack.

when the time comes to deactivate locals, that's done via `removevars(tolevel)`. `tolevel` is assumed to contain `nactvars` as it existed prior to entering the previous block. thus, the number of locals to remove should simply be `fs->nactvar-tolevel`. `removevars(tolevel)` will decrement `nactvars` down to `tolevel`. it also shrinks the `dyd` vector, and marks the endpc's of all the removed locals.

except in between `new_localvar` and `adjustlocalvar` calls, i believe that:

```
fs->ls->dyd->actvar.n - fs->firstlocal == fs->nactvar
```

temporary registers

`freereg` is used to manage the temporary register stack – registers between `[fs->nactvars,fs->freereg)` are assumed to belong to expressions currently being stored by the parser.

`fs->freereg` is incremented explicitly by calls to `luaK_reserveregs`, or implicitly, inside `luaK_exp2nextreg`. it's decremented whenever a `freereg(r)` is called on a register in the temporary stack (i.e., a register for which `r >= fs->nactvar`).

the temporary register stack is cleared when `leaveblock()` is called, by setting `fs->freereg=fs->nactvar`. it's also partially cleared in other places – for example, inside the evaluation of table constructors.

note that `freereg` just pops the top of the stack if r does not appear to be a local – thus it doesn't necessarily, free r. one of the important sanity checks that you'll get by enabling `lua_assert()` checks that the register being freed is also the top of the stack.

when writing parser patches, it's your job to ensure that the registers that you've reserved are freed in an appropriate order.

when a `VINDEXED` expression is discharged, `freereg()` will be called on both the table and the index register. otherwise, `freereg` is only called from `freeexp()` – which gets triggered anytime an expression has been "used up"; typically, anytime it's been transformed into another expression.

3.2.3 State Transitions

The state transitions for `expdesc` structure are as follows:

expkind	Description	State Transitions
VVOID	This is used to indicate the lack of value - e.g. function call with no arguments, the rhs of local variable declaration, and empty table constructor	None
VRELOCABLE	This is used to indicate that the result from expression needs to be set to a register. The operation that created the expression is referenced by the `u.info` parameter which contains an offset into the `code` of the function that is being compiled So you can access this instruction by calling `getcode(FuncState *, expdesc *)` The operations that result in a `VRELOCABLE` object include `OP_CLOSURE` `OP_NEWTABLE` `OP_GETUPVAL` `OP_GETTABUP` `OP_GETTABLE` `OP_NOT` and code for binary and unary expressions that produce values (arithmetic operations, bitwise operations, concat, length). The associated code instruction has operand `A` unset (defaulted to 0) - this the `VRELOCABLE` expression must be later transitioned to `VNONRELOC` state when the register is set.	In terms of transitions the following expression kinds convert to `VRELOCABLE`: `VVARARG` `VUPVAL` (`OP_GETUPVAL` `VINDEXED` (`OP_GETTABUP` or `OP_GETTABLE` And following expression states can result from a `VRELOCABLE` expression: `VNONRELOC` which means that the result register in the instruction operand A has been set.
VNONRELOC	This state indicates that the output or result register has been set. The register is referenced in `u.info` parameter. Once set the register cannot be changed for this expression; subsequent operations involving this expression can refer to the register to obtain the result value.	As for transitions, the `VNONELOC` state results from `VRELOCABLE` after a register is assigned to the operation referenced by `VRELOCABLE`. Also a `VCALL` expression transitions to `VNONRELOC` expression - `u.info` is set to the operand `A` in the call instruction. `VLOCAL` `VNIL` `VTRUE` `VFALSE` `VK` `VKINT` `VKFLT` and `VJMP` expressions transition to `VNONRELOC`.
VLOCAL	This is used when referencing local variables. `u.info` is set to the local variable's register.	The `VLOCAL` expression may transition to `VNONRELOC` although this doesn't change the `u.info` parameter.
VCALL	This results from a function call. The `OP_CALL` instruction is referenced by `u.info` parameter and may be retrieved by calling `getcode(FuncState *, expdesc *)`. The `OP_CALL` instruction gets changed to `OP_TAILCALL` if the function call expression is the value of a `RETURN` statement. The instructions operand `C` gets updated when it is known the number of expected results from the function call.	In terms of transitions, the `VCALL` expression transitions to `VNONRELOC` When this happens the result register in `VNONRELOC` (`u.info` is set to the operand `A` in the `OP_CALL` instruction.
VINDEXED	This expression represents a table access. The `u.ind.t` parameter is set to the register or upvalue? that holds the table, the `u.ind.idx` is set to the register or constant that is the key, and `u.ind.vt` is either `VLOCAL` or `VUPVAL`	The `VINDEXED` expression transitions to `VRELOCABLE` When this happens the `u.info` is set to the offset of the code that contains the opcode `OP_GETTABUP` if `u.ind.vt` was `VUPVAL` or `OP_GETTABLE` if `u.ind.vt` was `VLOCAL`

3.2.4 Examples of Parsing

example 1

We investigate the simple code chunk below:

```
local i,j; j = i*j+i
```

The compiler allocates following local registers, constants and upvalues:

```
constants (0) for 0000007428FED950:
locals (2) for 0000007428FED950:
      0       i       2       5
      1       j       2       5
upvalues (1) for 0000007428FED950:
      0       _ENV    1       0
```

Some of the parse steps are highlighted below.

Reference to variable `i` which is located in register `0`. The `p` here is the pointer address of `expdesc` object so you can see how the same object evolves:

```
{p=0000007428E1F170, k=VLOCAL, register=0}
```

Reference to variable `j` located in register `1`:

```
{p=0000007428E1F078, k=VLOCAL, register=1}
```

Now the MUL operator is applied so we get following. Note that the previously `VLOCAL` expression for `i` is now `VNONRELOC`:

```
{p=0000007428E1F170, k=VNONRELOC, register=0} MUL {p=0000007428E1F078, k=VLOCAL, register=1}
```

Next code gets generated for the `MUL` operator and we can see that first expression is replaced by a `VRELOCABLE` expression. Note also that the `MUL` operator is encoded in the `VRELOCABLE` expression as instruction `1` which is decoded below:

```
{p=0000007428E1F170, k=VRELOCABLE, pc=1, instruction=(MUL A=0 B=0 C=1)}
```

Now a reference to `i` is again required:

```
{p=0000007428E1F078, k=VLOCAL, register=0}
```

And the `ADD` operator must be applied to the result of the `MUL` operator and above. Notice that a temporary register `2` has been allocated to hold the result of the `MUL` operator, and also notice that as a result the `VRELOCABLE` has now changed to `VNONRELOC`:

```
{p=0000007428E1F170, k=VNONRELOC, register=2} ADD {p=0000007428E1F078, k=VLOCAL, register=0}
```

Next the result of the `ADD` expression gets encoded similarly to `MUL` earlier. As this is a `VRELOCABLE` expression it will be later on assigned a result register:

```
{p=0000007428E1F170, k=VRELOCABLE, pc=2, instruction=(ADD A=0 B=2 C=0)}
```

Eventually above gets assigned a result register and becomes `VNONRELOC` (not shown here) - and so the final generated code looks like below:

```
main <(string):0,0> (4 instructions at 0000007428FED950)
0+ params, 3 slots, 1 upvalue, 2 locals, 0 constants, 0 functions
      1       [1]     LOADNIL         0 1
      2       [1]     MUL             2 0 1
```

```
      3        [1]     ADD             1 2 0
      4        [1]     RETURN          0 1
```

3.3 Links

- (MP1) Lua Code Reading Order
- (RL1) Registers allocation and GC
- (MP2) LuaJIT interpreter optmisations
- (MP3) Performance of Switch Based Dispatch
- (MP4) Challenges for static compilation of dynamic langauges
- (MP5) VM Internals (bytecode format)
- (RL2) Upvalues in closures
- (LHF) Lua bytecode dump format
- (MP6) Register VM and sliding stack window
- (SO1) Sven Olsen's notes on registers from Sven Olsen's Lua Users Wiki page
- (KHM) No Frills Introduction to Lua 5.1 VM Instructions
- (MP7) LuaJIT Roadmap 2008
- (MP8) LuaJIT Roadmap 2011

CHAPTER 4

Ravi Parsing and ByteCode Implementation Details

This document covers the enhancements to the Lua parser and byte-code generator. The Ravi JIT implementation is described elsewhere.

4.1 Introduction

Since the reason for introducing optional static typing is to enhance performance primarily - not all types benefit from this capability. In fact it is quite hard to extend this to generic recursive structures such as tables without encurring significant overhead. For instance - even to represent a recursive type in the parser will require dynamic memory allocation and add great overhead to the parser.

From a performance point of view the only types that seem worth specializing are:

- integer (64-bit int)
- number (double)
- array of integers
- array of numbers
- table

4.2 Implementation Strategy

I want to build on existing Lua types rather than introducing completely new types to the Lua system. I quite like the minimalist nature of Lua. However, to make the execution efficient I am adding new type specific opcodes and enhancing the Lua parser/code generator to encode these opcodes only when types are known. The new opcodes will execute more efficiently as they will not need to perform type checks. Morever, type specific instructions will lend themselves to more efficient JIT compilation.

I am adding new opcodes that cover arithmetic operations, array operations, variable assignments, etc..

4.3 Modifications to Lua Bytecode structure

An immediate issue is that the Lua bytecode structure has a 6-bit opcode which is insufficient to hold the various opcodes that I will need. Simply extending the size of this is problematic as then it reduces the space available to the operands A B and C. Furthermore the way Lua bytecodes work means that B and C operands must be 1-bit larger than

A - as the extra bit is used to flag whether the operand refers to a constant or a register. (Thanks to Dirk Laurie for pointing this out).

I am amending the bit mapping in the 32-bit instruction to allow 9-bits for the byte-code, 7-bits for operand A, and 8-bits for operands B and C. This means that some of the Lua limits (maximum number of variables in a function, etc.) have to be revised to be lower than the default.

4.4 New OpCodes

The new instructions are specialised for types, and also for register/versus constant. So for example `OP_RAVI_ADDFI` means add `number` and `integer`. And `OP_RAVI_ADDFF` means add `number` and `number`. The existing Lua opcodes that these are based on define which operands are used.

Example:

```
local i=0; i=i+1
```

Above standard Lua code compiles to:

```
[0] LOADK A=0 Bx=-1
[1] ADD A=0 B=0 C=-2
[2] RETURN A=0 B=1
```

We add type info using Ravi extensions:

```
local i:integer=0; i=i+1
```

Now the code compiles to:

```
[0] LOADK A=0 Bx=-1
[1] ADDII A=0 B=0 C=-2
[2] RETURN A=0 B=1
```

Above uses type specialised opcode `OP_RAVI_ADDII`.

4.5 Type Information

The basic first step is to add type information to Lua.

As the parser progresses it creates a vector of `LocVar` for each function containing a list of local variables. I have enhanced `LocVar` structure in `lobject.h` to hold type information.

```
/* Following are the types we will use
** use in parsing. The rationale for types is
** performance - as of now these are the only types that
** we care about from a performance point of view - if any
** other types appear then they are all treated as ANY
**/
typedef enum {
  RAVI_TANY = -1,       /* Lua dynamic type */
  RAVI_TNUMINT,    /* integer number */
  RAVI_TNUMFLT,    /* floating point number */
  RAVI_TARRAYINT, /* array of ints */
  RAVI_TARRAYFLT,  /* array of doubles */
  RAVI_TFUNCTION,
  RAVI_TTABLE,
```

```
  RAVI_TSTRING,
  RAVI_TNIL,
  RAVI_TBOOLEAN
} ravitype_t;

/*
** Description of a local variable for function prototypes
** (used for debug information)
*/
typedef struct LocVar {
  TString *varname;
  int startpc;  /* first point where variable is active */
  int endpc;    /* first point where variable is dead */
  ravitype_t ravi_type; /* RAVI type of the variable - RAVI_TANY if unknown */
} LocVar;
```

The `expdesc` structure is used by the parser to hold nodes in the expression tree. I have enhanced the `expdesc` structure to hold the type of an expression.

```
typedef struct expdesc {
  expkind k;
  union {
    struct {  /* for indexed variables (VINDEXED) */
      short idx;  /* index (R/K) */
      lu_byte t;  /* table (register or upvalue) */
      lu_byte vt;  /* whether 't' is register (VLOCAL) or upvalue (VUPVAL) */
      ravitype_t key_type; /* key type */
    } ind;
    int info;  /* for generic use */
    lua_Number nval;  /* for VKFLT */
    lua_Integer ival;    /* for VKINT */
  } u;
  int t;  /* patch list of 'exit when true' */
  int f;  /* patch list of 'exit when false' */
  ravitype_t ravi_type; /* RAVI change: type of the expression if known, else RAVI_TANY */
} expdesc;
```

Note the addition of type information in two places. Firstly at the `expdesc` level which identifies the type of the `expdesc`. Secondly in the *ind* structure - the `key_type` is used to track the type of the key that will be used to index into a table.

The table structure has been enhanced to hold additional information for array usage.

```
typedef enum RaviArrayModifer {
  RAVI_ARRAY_SLICE = 1,
  RAVI_ARRAY_FIXEDSIZE = 2
} RaviArrayModifier;

typedef struct RaviArray {
  char *data;
  unsigned int len; /* RAVI len specialization */
  unsigned int size; /* amount of memory allocated */
  lu_byte array_type; /* RAVI specialization */
  lu_byte array_modifier; /* Flags that affect how the array is handled */
} RaviArray;

typedef struct Table {
  CommonHeader;
  lu_byte flags;  /* 1<<p means tagmethod(p) is not present */
```

```
  lu_byte lsizenode;  /* log2 of size of 'node' array */
  unsigned int sizearray;  /* size of 'array' array */
  TValue *array;  /* array part */
  Node *node;
  Node *lastfree;  /* any free position is before this position */
  struct Table *metatable;
  GCObject *gclist;
  RaviArray ravi_array;
} Table;
```

4.6 Parser Enhancements

The parser needs to be enhanced to generate type specific instructions at various points.

4.6.1 Local Variable Declarations

First enhancement needed is when local variable declarations are parsed. We need to allow the type to be defined for each variable and ensure that any assignments are type-checked. This is somewhat complex process, due to the fact that assignments can be expressions involving function calls. The last function call is treated as a variable assignment - i.e. all trailing variables are assumed to be assigned values from the function call - if not the variables are set to nil by default.

The entry point for parsing a local statement is `localstat()` in `lparser.c`. This function has been enhanced to parse the type annotations supported by Ravi. The modified function is shown below.

```
/* Parse
 *   name : type
 *   where type is 'integer', 'integer[]',
 *                 'number', 'number[]'
 */
static ravitype_t declare_localvar(LexState *ls) {
  /* RAVI change - add type */
  TString *name = str_checkname(ls);
  /* assume a dynamic type */
  ravitype_t tt = RAVI_TANY;
  /* if the variable name is followed by a colon then we have a type
   * specifier
   */
  if (testnext(ls, ':')) {
    TString *typename = str_checkname(ls); /* we expect a type name */
    const char *str = getaddrstr(typename);
    /* following is not very nice but easy as
     * the lexer doesn't need to be changed
     */
    if (strcmp(str, "integer") == 0)
      tt = RAVI_TNUMINT;
    else if (strcmp(str, "number") == 0)
      tt = RAVI_TNUMFLT;
    if (tt == RAVI_TNUMFLT || tt == RAVI_TNUMINT) {
      /* if we see [] then it is an array type */
      if (testnext(ls, '[')) {
        checknext(ls, ']');
        tt = (tt == RAVI_TNUMFLT) ? RAVI_TARRAYFLT : RAVI_TARRAYINT;
      }
    }
```

```
  }
  new_localvar(ls, name, tt);
  return tt;
}

/* parse a local variable declaration statement - called from statement() */
static void localstat (LexState *ls) {
  /* stat -> LOCAL NAME {',' NAME} ['=' explist] */
  int nvars = 0;
  int nexps;
  expdesc e;
  e.ravi_type = RAVI_TANY;
  /* RAVI while declaring locals we need to gather the types
   * so that we can check any assignments later on.
   * TODO we may be able to use register_typeinfo() here
   * instead.
   */
  int vars[MAXVARS] = { 0 };
  do {
    /* RAVI changes start */
    /* local name : type = value */
    vars[nvars] = declare_localvar(ls);
    /* RAVI changes end */
    nvars++;
  } while (testnext(ls, ','));
  if (testnext(ls, '='))
    nexps = localvar_explist(ls, &e, vars, nvars);
  else {
    e.k = VVOID;
    nexps = 0;
  }
  localvar_adjust_assign(ls, nvars, nexps, &e);
  adjustlocalvars(ls, nvars);
}
```

The do-while loop is responsible for parsing the variable names and the type annotations. As each variable name is parsed we detect if there is a type annotation, if and if present the type is recorded in the array `vars`.

Parameter lists may have static type annotations as well, so when parsing parameters we again need to invoke `declare_localvar()`.

```
static void parlist (LexState *ls) {
  /* parlist -> [ param { ',' param } ] */
  FuncState *fs = ls->fs;
  Proto *f = fs->f;
  int nparams = 0;
  f->is_vararg = 0;
  if (ls->t.token != ')') {  /* is 'parlist' not empty? */
    do {
      switch (ls->t.token) {
        case TK_NAME: {  /* param -> NAME */
          /* RAVI change - add type */
          declare_localvar(ls);
          nparams++;
          break;
        }
        case TK_DOTS: {  /* param -> '...' */
          luaX_next(ls);
          f->is_vararg = 1;
```

```
          break;
        }
        default: luaX_syntaxerror(ls, "<name> or '...' expected");
      }
    } while (!f->is_vararg && testnext(ls, ','));
  }
  adjustlocalvars(ls, nparams);
  f->numparams = cast_byte(fs->nactvar);
  luaK_reserveregs(fs, fs->nactvar);  /* reserve register for parameters */
  for (int i = 0; i < f->numparams; i++) {
    ravitype_t tt = raviY_get_register_typeinfo(fs, i);
    DEBUG_VARS(raviY_printf(fs, "Parameter [%d] = %v\n", i + 1, getlocvar(fs, i)));
    /* do we need to convert ? */
    if (tt == RAVI_TNUMFLT || tt == RAVI_TNUMINT) {
      /* code an instruction to convert in place */
      luaK_codeABC(ls->fs, tt == RAVI_TNUMFLT ? OP_RAVI_TOFLT : OP_RAVI_TOINT, i, 0, 0);
    }
    else if (tt == RAVI_TARRAYFLT || tt == RAVI_TARRAYINT) {
      /* code an instruction to convert in place */
      luaK_codeABC(ls->fs, tt == RAVI_TARRAYFLT ? OP_RAVI_TOARRAYF : OP_RAVI_TOARRAYI, i, 0, 0);
    }
  }
}
```

Additionally for parameters that are decorated with static types we need to introduce new instructions to coerce the types at run time. That is what is happening in the for loop at the end.

The `declare_localvar()` function passes the type of the variable to `new_localvar()` which records this in the `LocVar` structure associated with the variable.

```
static int registerlocalvar (LexState *ls, TString *varname, int ravi_type) {
  FuncState *fs = ls->fs;
  Proto *f = fs->f;
  int oldsize = f->sizelocvars;
  luaM_growvector(ls->L, f->locvars, fs->nlocvars, f->sizelocvars,
                  LocVar, SHRT_MAX, "local variables");
  while (oldsize < f->sizelocvars) {
    /* RAVI change initialize */
    f->locvars[oldsize].startpc = -1;
    f->locvars[oldsize].endpc = -1;
    f->locvars[oldsize].ravi_type = RAVI_TANY;
    f->locvars[oldsize++].varname = NULL;
  }
  f->locvars[fs->nlocvars].varname = varname;
  f->locvars[fs->nlocvars].ravi_type = ravi_type;
  luaC_objbarrier(ls->L, f, varname);
  return fs->nlocvars++;
}

/* create a new local variable in function scope, and set the
 * variable type (RAVI - added type tt) */
static void new_localvar (LexState *ls, TString *name, ravitype_t tt) {
  FuncState *fs = ls->fs;
  Dyndata *dyd = ls->dyd;
  /* register variable and get its index */
  /* RAVI change - record type info for local variable */
  int i = registerlocalvar(ls, name, tt);
  checklimit(fs, dyd->actvar.n + 1 - fs->firstlocal,
                 MAXVARS, "local variables");
```

```
  luaM_growvector(ls->L, dyd->actvar.arr, dyd->actvar.n + 1,
                  dyd->actvar.size, Vardesc, MAX_INT, "local variables");
  /* variable will be placed at stack position dyd->actvar.n */
  dyd->actvar.arr[dyd->actvar.n].idx = cast(short, i);
  DEBUG_VARS(raviY_printf(fs, "new_localvar -> registering %v fs->f->locvars[%d] at ls->dyd->actvar.a
  dyd->actvar.n++;
  DEBUG_VARS(raviY_printf(fs, "new_localvar -> ls->dyd->actvar.n set to %d\n", dyd->actvar.n));
}
```

The next bit of change is how the expressions are handled following the = symbol. The previously built `vars` array is passed to a modified version of `explist()` called `localvar_explist()`. This handles the parsing of expressions and then ensuring that each expression matches the type of the variable where known. The `localvar_explist()` function is shown next.

```
static int localvar_explist(LexState *ls, expdesc *v, int *vars, int nvars) {
  /* explist -> expr { ',' expr } */
  int n = 1;  /* at least one expression */
  expr(ls, v);
#if RAVI_ENABLED
  ravi_typecheck(ls, v, vars, nvars, 0);
#endif
  while (testnext(ls, ',')) {
    luaK_exp2nextreg(ls->fs, v);
    expr(ls, v);
#if RAVI_ENABLED
    ravi_typecheck(ls, v, vars, nvars, n);
#endif
    n++;
  }
  return n;
}
```

The main changes compared to `explist()` are the calls to `ravi_typecheck()`. Note that the array `vars` is passed to the `ravi_typecheck()` function along with the current variable index in `n`. The `ravi_typecheck()` function is reproduced below.

```
static void ravi_typecheck(LexState *ls, expdesc *v, int *vars, int nvars, int n)
{
  if (n < nvars && vars[n] != RAVI_TANY && v->ravi_type != vars[n]) {
    if (v->ravi_type != vars[n] &&
        (vars[n] == RAVI_TARRAYFLT || vars[n] == RAVI_TARRAYINT) &&
        v->k == VNONRELOC) {
      /* as the bytecode for generating a table is already
       * emitted by this stage we have to amend the generated byte code
       * - not sure if there is a better approach.
       * We look for the last bytecode that is OP_NEWTABLE
       * and that has the same destination
       * register as v->u.info which is our variable
       * local a:integer[] = { 1 }
       *                     ^ We are just past this and
       *                       about to assign to a
       */
      int i = ls->fs->pc - 1;
      for (; i >= 0; i--) {
        Instruction *pc = &ls->fs->f->code[i];
        OpCode op = GET_OPCODE(*pc);
        int reg;
        if (op != OP_NEWTABLE)
```

```
        continue;
      reg = GETARG_A(*pc);
      if (reg != v->u.info)
        continue;
      op = (vars[n] == RAVI_TARRAYINT) ? OP_RAVI_NEWARRAYI : OP_RAVI_NEWARRAYF;
      SET_OPCODE(*pc, op); /* modify opcode */
      DEBUG_CODEGEN(raviY_printf(ls->fs, "[%d]* %o ; modify opcode\n", i, *pc));
      break;
    }
    if (i < 0)
      luaX_syntaxerror(ls, "expecting array initializer");
  }
  /* if we are calling a function then convert return types */
  else if (v->ravi_type != vars[n] &&
          (vars[n] == RAVI_TNUMFLT || vars[n] == RAVI_TNUMINT) &&
          v->k == VCALL) {
    /* For local variable declarations that call functions e.g.
     * local i = func()
     * Lua ensures that the function returns values
     * to register assigned to variable i and above so that no
     * separate OP_MOVE instruction is necessary. So that means that
     * we need to coerce the return values in situ.
     */
    /* Obtain the instruction for OP_CALL */
    Instruction *pc = &getcode(ls->fs, v);
    lua_assert(GET_OPCODE(*pc) == OP_CALL);
    int a = GETARG_A(*pc); /* function return values
                              will be placed from register pointed
                              by A and upwards */
    int nrets = GETARG_C(*pc) - 1; /* operand C contains
                                      number of return values expected  */
    /* Note that at this stage nrets is always 1
     * - as Lua patches in the this value for the last
     * function call in a variable declaration statement
     * in adjust_assign and localvar_adjust_assign */
    /* all return values that are going to be assigned
       to typed local vars must be converted to the correct type */
    int i;
    for (i = n; i < (n+nrets); i++)
      /* do we need to convert ? */
      if ((vars[i] == RAVI_TNUMFLT || vars[i] == RAVI_TNUMINT))
        /* code an instruction to convert in place */
        luaK_codeABC(ls->fs,
                     vars[i] == RAVI_TNUMFLT ?
                               OP_RAVI_TOFLT : OP_RAVI_TOINT,
                     a+(i-n), 0, 0);
      else if ((vars[i] == RAVI_TARRAYFLT || vars[i] == RAVI_TARRAYINT))
        /* code an instruction to convert in place */
        luaK_codeABC(ls->fs,
                     vars[i] == RAVI_TARRAYFLT ?
                               OP_RAVI_TOARRAYF : OP_RAVI_TOARRAYI,
                     a + (i - n), 0, 0);
  }
  else if ((vars[n] == RAVI_TNUMFLT || vars[n] == RAVI_TNUMINT) &&
           v->k == VINDEXED) {
    if (vars[n] == RAVI_TNUMFLT && v->ravi_type != RAVI_TARRAYFLT ||
      vars[n] == RAVI_TNUMINT && v->ravi_type != RAVI_TARRAYINT)
      luaX_syntaxerror(ls, "Invalid local assignment");
```

```
    }
    else
      luaX_syntaxerror(ls, "Invalid local assignment");
  }
}
```

There are several parts to this function.

The simple case is when the type of the expression matches the variable.

Secondly if the expression is a table initializer then we need to generate specialized opcodes if the target variable is supposed to be `integer[]` or `number[]`. The specialized opcode sets up some information in the `Table` structure. The problem is that this requires us to modify `OP_NEWTABLE` instruction which has already been emitted. So we scan the generated instructions to find the last `OP_NEWTABLE` instruction that assigns to the register associated with the target variable.

Next bit of special handling is for function calls. If the assignment makes a function call then we perform type coercion on return values where these values are being assigned to variables with defined types. This means that if the target variable is `integer` or `number` we issue opcodes `TOINT` and `TOFLT` respectively. If the target variable is `integer[]` or `number[]` then we issue `TOARRAYI` and `TOARRAYF` respectively. These opcodes ensure that the values are of required type or can be cast to the required type.

Note that any left over variables that are not assigned values, are set to 0 if they are of integer or number type, else they are set to nil as per Lua's default behavior. This is handled in `localvar_adjust_assign()` which is described later on.

Finally the last case is when the target variable is `integer` or `number` and the expression is a table / array access. In this case we check that the table is of required type.

The `localvar_adjust_assign()` function referred to above is shown below.

```
static void localvar_adjust_assign(LexState *ls, int nvars, int nexps, expdesc *e) {
  FuncState *fs = ls->fs;
  int extra = nvars - nexps;
  if (hasmultret(e->k)) {
    extra++;  /* includes call itself */
    if (extra < 0) extra = 0;
    /* following adjusts the C operand in the OP_CALL instruction */
    luaK_setreturns(fs, e, extra);  /* last exp. provides the difference */
#if RAVI_ENABLED
    /* Since we did not know how many return values to process in localvar_explist() we
    * need to add instructions for type coercions at this stage for any remaining
    * variables
    */
    ravi_coercetype(ls, e, extra);
#endif
    if (extra > 1) luaK_reserveregs(fs, extra - 1);
  }
  else {
    if (e->k != VVOID) luaK_exp2nextreg(fs, e);  /* close last expression */
    if (extra > 0) {
      int reg = fs->freereg;
      luaK_reserveregs(fs, extra);
      /* RAVI TODO for typed variables we should not set to nil? */
      luaK_nil(fs, reg, extra);
#if RAVI_ENABLED
      /* typed variables that are primitives cannot be set to nil so
       * we need to emit instructions to initialise them to default values
       */
      ravi_setzero(fs, reg, extra);
```

```
#endif
    }
  }
}
```

As mentioned before any variables left over in a local declaration that have not been assigned values must be set to default values appropriate for the type. In the case of trailing values returned by a function call we need to coerce the values to the required types. All this is done in the `localvar_adjust_assign()` function above.

Note that local declarations have a complication that until the declaration is complete the variable does not come in scope. So we have to be careful when we wish to map from a register to the local variable declaration as this mapping is only available after the variable is activated. Couple of helper routines are shown below.

```
/* translate from local register to local variable index
 */
static int register_to_locvar_index(FuncState *fs, int reg) {
  int idx;
  lua_assert(reg >= 0 && (fs->firstlocal + reg) < fs->ls->dyd->actvar.n);
  /* Get the LocVar associated with the register */
  idx = fs->ls->dyd->actvar.arr[fs->firstlocal + reg].idx;
  lua_assert(idx < fs->nlocvars);
  return idx;
}

/* get type of a register - if the register is not allocated
 * to an active local variable, then return RAVI_TANY else
 * return the type associated with the variable.
 * This is a RAVI function
 */
ravitype_t raviY_get_register_typeinfo(FuncState *fs, int reg) {
  int idx;
  LocVar *v;
  if (reg < 0 || reg >= fs->nactvar || (fs->firstlocal + reg) >= fs->ls->dyd->actvar.n)
    return RAVI_TANY;
  /* Get the LocVar associated with the register */
  idx = fs->ls->dyd->actvar.arr[fs->firstlocal + reg].idx;
  lua_assert(idx < fs->nlocvars);
  v = &fs->f->locvars[idx];
  /* Variable in scope so return the type if we know it */
  return v->ravi_type;
}
```

Note the use of `register_to_localvar_index()` in functions below.

```
/* Generate instructions for converting types
 * This is needed post a function call to handle
 * variable number of return values
 * n = number of return values to adjust
 */
static void ravi_coercetype(LexState *ls, expdesc *v, int n)
{
  if (v->k != VCALL || n <= 0) return;
  /* For local variable declarations that call functions e.g.
  * local i = func()
  * Lua ensures that the function returns values to register
  * assigned to variable and above so that no separate
  * OP_MOVE instruction is necessary. So that means that
  * we need to coerce the return values in situ.
  */
```

```
  /* Obtain the instruction for OP_CALL */
  Instruction *pc = &getcode(ls->fs, v);
  lua_assert(GET_OPCODE(*pc) == OP_CALL);
  int a = GETARG_A(*pc); /* function return values will be placed
                            from register pointed by A and upwards */
  /* all return values that are going to be assigned
   to typed local vars must be converted to the correct type */
  int i;
  for (i = a + 1; i < a + n; i++) {
    /* Since this is called when parsing local statements the
     * variable may not yet have a register assigned to it
     * so we can't use raviY_get_register_typeinfo()
     * here. Instead we need to check the variable definition - so we
     * first convert from local register to variable index.
     */
    int idx = register_to_locvar_index(ls->fs, i);
    /* get variable's type */
    ravitype_t ravi_type = ls->fs->f->locvars[idx].ravi_type;
    /* do we need to convert ? */
    if (ravi_type == RAVI_TNUMFLT || ravi_type == RAVI_TNUMINT)
      /* code an instruction to convert in place */
      luaK_codeABC(ls->fs, ravi_type == RAVI_TNUMFLT ?
                   OP_RAVI_TOFLT : OP_RAVI_TOINT, i, 0, 0);
    else if (ravi_type == RAVI_TARRAYINT || ravi_type == RAVI_TARRAYFLT)
      luaK_codeABC(ls->fs, ravi_type == RAVI_TARRAYINT ?
                   OP_RAVI_TOARRAYI : OP_RAVI_TOARRAYF, i, 0, 0);
  }
}

static void ravi_setzero(FuncState *fs, int from, int n) {
  int last = from + n - 1;  /* last register to set nil */
  int i;
  for (i = from; i <= last; i++) {
    /* Since this is called when parsing local statements
     * the variable may not yet have a register assigned to
     * it so we can't use raviY_get_register_typeinfo()
     * here. Instead we need to check the variable definition - so we
     * first convert from local register to variable index.
     */
    int idx = register_to_locvar_index(fs, i);
    /* get variable's type */
    ravitype_t ravi_type = fs->f->locvars[idx].ravi_type;
    /* do we need to convert ? */
    if (ravi_type == RAVI_TNUMFLT || ravi_type == RAVI_TNUMINT)
      /* code an instruction to convert in place */
      luaK_codeABC(fs, ravi_type == RAVI_TNUMFLT ?
         OP_RAVI_LOADFZ : OP_RAVI_LOADIZ, i, 0, 0);
  }
}
```

4.6.2 Assignments

Assignment statements have to be enhanced to perform similar type checks as for local declarations. Fortunately he assignment goes through the function `luaK_storevar()` in `lcode.c`. A modified version of this is shown below.

```
void luaK_storevar (FuncState *fs, expdesc *var, expdesc *ex) {
  switch (var->k) {
```

```
    case VLOCAL: {
      check_valid_store(fs, var, ex);
      freeexp(fs, ex);
      exp2reg(fs, ex, var->u.info);
      return;
    }
    case VUPVAL: {
      int e = luaK_exp2anyreg(fs, ex);
      luaK_codeABC(fs, OP_SETUPVAL, e, var->u.info, 0);
      break;
    }
    case VINDEXED: {
      OpCode op = (var->u.ind.vt == VLOCAL) ?
                    OP_SETTABLE : OP_SETTABUP;
      if (op == OP_SETTABLE) {
        /* table value set - if array access then use specialized versions */
        if (var->ravi_type == RAVI_TARRAYFLT &&
            var->u.ind.key_type == RAVI_TNUMINT)
          op = OP_RAVI_SETTABLE_AF;
        else if (var->ravi_type == RAVI_TARRAYINT &&
                 var->u.ind.key_type == RAVI_TNUMINT)
          op = OP_RAVI_SETTABLE_AI;
      }
      int e = luaK_exp2RK(fs, ex);
      luaK_codeABC(fs, op, var->u.ind.t, var->u.ind.idx, e);
      break;
    }
    default: {
      lua_assert(0);  /* invalid var kind to store */
      break;
    }
  }
  freeexp(fs, ex);
}
```

Firstly note the call to `check_valid_store()` for a local variable assignment. The `check_valid_store()` function validates that the assignment is compatible.

Secondly if the assignment is to an indexed variable, i.e., table, then we need to generate special opcodes for arrays.

4.6.3 MOVE opcodes

Any `MOVE` instructions must be modified so that if the target is register that hosts a variable of known type then we need to generate special instructions that do a type conversion during the move. This is handled in `discharge2reg()` function which is reproduced below.

```
static void discharge2reg (FuncState *fs, expdesc *e, int reg) {
  luaK_dischargevars(fs, e);
  switch (e->k) {
    case VNIL: {
      luaK_nil(fs, reg, 1);
      break;
    }
    case VFALSE: case VTRUE: {
      luaK_codeABC(fs, OP_LOADBOOL, reg, e->k == VTRUE, 0);
      break;
    }
```

```
    case VK: {
      luaK_codek(fs, reg, e->u.info);
      break;
    }
    case VKFLT: {
      luaK_codek(fs, reg, luaK_numberK(fs, e->u.nval));
      break;
    }
    case VKINT: {
      luaK_codek(fs, reg, luaK_intK(fs, e->u.ival));
      break;
    }
    case VRELOCABLE: {
      Instruction *pc = &getcode(fs, e);
      SETARG_A(*pc, reg);
      DEBUG_EXPR(raviY_printf(fs, "discharge2reg (VRELOCABLE set arg A) %e\n", e));
      DEBUG_CODEGEN(raviY_printf(fs, "[%d]* %o ; set A to %d\n", e->u.info, *pc, reg));
      break;
    }
    case VNONRELOC: {
      if (reg != e->u.info) {
        /* code a MOVEI or MOVEF if the target register is a local typed variable */
        int ravi_type = raviY_get_register_typeinfo(fs, reg);
        switch (ravi_type) {
        case RAVI_TNUMINT:
          luaK_codeABC(fs, OP_RAVI_MOVEI, reg, e->u.info, 0);
          break;
        case RAVI_TNUMFLT:
          luaK_codeABC(fs, OP_RAVI_MOVEF, reg, e->u.info, 0);
          break;
        case RAVI_TARRAYINT:
          luaK_codeABC(fs, OP_RAVI_MOVEAI, reg, e->u.info, 0);
          break;
        case RAVI_TARRAYFLT:
          luaK_codeABC(fs, OP_RAVI_MOVEAF, reg, e->u.info, 0);
          break;
        default:
          luaK_codeABC(fs, OP_MOVE, reg, e->u.info, 0);
          break;
        }
      }
      break;
    }
    default: {
      lua_assert(e->k == VVOID || e->k == VJMP);
      return;  /* nothing to do... */
    }
  }
  e->u.info = reg;
  e->k = VNONRELOC;
}
```

Note the handling of `VNONRELOC` case.

4.6.4 Expression Parsing

The expression evaluation process must be modified so that type information is retained and flows through as the parser evaluates the expression. This involves ensuring that the type information is passed through as the parser modifies, reuses, creates new `expdesc` objects. Essentially this means keeping the `ravi_type` correct.

Additionally when arithmetic operations take place two things need to happen: a) specialized opcodes need to be emitted and b) the type of the resulting expression needs to be set.

```
static void codeexpval (FuncState *fs, OpCode op,
                        expdesc *e1, expdesc *e2, int line) {
  lua_assert(op >= OP_ADD);
  if (op <= OP_BNOT && constfolding(fs, getarithop(op), e1, e2))
    return;  /* result has been folded */
  else {
    int o1, o2;
    int isbinary = 1;
    /* move operands to registers (if needed) */
    if (op == OP_UNM || op == OP_BNOT || op == OP_LEN) {  /* unary op? */
      o2 = 0;  /* no second expression */
      o1 = luaK_exp2anyreg(fs, e1);  /* cannot operate on constants */
      isbinary = 0;
    }
    else {  /* regular case (binary operators) */
      o2 = luaK_exp2RK(fs, e2);  /* both operands are "RK" */
      o1 = luaK_exp2RK(fs, e1);
    }
    if (o1 > o2) {  /* free registers in proper order */
      freeexp(fs, e1);
      freeexp(fs, e2);
    }
    else {
      freeexp(fs, e2);
      freeexp(fs, e1);
    }
#if RAVI_ENABLED
    if (op == OP_ADD &&
      (e1->ravi_type == RAVI_TNUMFLT || e1->ravi_type == RAVI_TNUMINT) &&
      (e2->ravi_type == RAVI_TNUMFLT || e2->ravi_type == RAVI_TNUMINT))
      generate_binarithop(fs, e1, e2, o1, o2, 0);
    else if (op == OP_MUL &&
      (e1->ravi_type == RAVI_TNUMFLT || e1->ravi_type == RAVI_TNUMINT) &&
      (e2->ravi_type == RAVI_TNUMFLT || e2->ravi_type == RAVI_TNUMINT))
      generate_binarithop(fs, e1, e2, o1, o2, OP_RAVI_MULFF - OP_RAVI_ADDFF);

    /* todo optimize the SUB opcodes when constant is small */
    else if (op == OP_SUB &&
             e1->ravi_type == RAVI_TNUMFLT &&
             e2->ravi_type == RAVI_TNUMFLT) {
      e1->u.info = luaK_codeABC(fs, OP_RAVI_SUBFF, 0, o1, o2);
    }
    else if (op == OP_SUB &&
             e1->ravi_type == RAVI_TNUMFLT &&
             e2->ravi_type == RAVI_TNUMINT) {
      e1->u.info = luaK_codeABC(fs, OP_RAVI_SUBFI, 0, o1, o2);
    }
    /* code omitted here  .... */
    else {
#endif
```

```
      e1->u.info = luaK_codeABC(fs, op, 0, o1, o2);  /* generate opcode */
#if RAVI_ENABLED
    }
#endif
    e1->k = VRELOCABLE;  /* all those operations are relocable */
    if (isbinary) {
      if ((op == OP_ADD || op == OP_SUB || op == OP_MUL || op == OP_DIV)
        && e1->ravi_type == RAVI_TNUMFLT && e2->ravi_type == RAVI_TNUMFLT)
        e1->ravi_type = RAVI_TNUMFLT;
      else if ((op == OP_ADD || op == OP_SUB || op == OP_MUL || op == OP_DIV)
        && e1->ravi_type == RAVI_TNUMFLT && e2->ravi_type == RAVI_TNUMINT)
        e1->ravi_type = RAVI_TNUMFLT;
      else if ((op == OP_ADD || op == OP_SUB || op == OP_MUL || op == OP_DIV)
        && e1->ravi_type == RAVI_TNUMINT && e2->ravi_type == RAVI_TNUMFLT)
        e1->ravi_type = RAVI_TNUMFLT;
      else if ((op == OP_ADD || op == OP_SUB || op == OP_MUL)
        && e1->ravi_type == RAVI_TNUMINT && e2->ravi_type == RAVI_TNUMINT)
        e1->ravi_type = RAVI_TNUMINT;
      else if ((op == OP_DIV)
        && e1->ravi_type == RAVI_TNUMINT && e2->ravi_type == RAVI_TNUMINT)
        e1->ravi_type = RAVI_TNUMFLT;
      else
        e1->ravi_type = RAVI_TANY;
    }
    else {
      if (op == OP_LEN || op == OP_BNOT)
        e1->ravi_type = RAVI_TNUMINT;
    }
    luaK_fixline(fs, line);
  }
}
```

When expression reference indexed variables, i.e., tables, we need to emit specialized opcodes if the table is an array. This is done in `luaK_dischargevars()`.

```
void luaK_dischargevars (FuncState *fs, expdesc *e) {
  switch (e->k) {
    case VLOCAL: {
      e->k = VNONRELOC;
      DEBUG_EXPR(raviY_printf(fs, "luaK_dischargevars (VLOCAL->VNONRELOC) %e\n", e));
      break;
    }
    case VUPVAL: {
      e->u.info = luaK_codeABC(fs, OP_GETUPVAL, 0, e->u.info, 0);
      e->k = VRELOCABLE;
      DEBUG_EXPR(raviY_printf(fs, "luaK_dischargevars (VUPVAL->VRELOCABLE) %e\n", e));
      break;
    }
    case VINDEXED: {
      OpCode op = OP_GETTABUP;  /* assume 't' is in an upvalue */
      freereg(fs, e->u.ind.idx);
      if (e->u.ind.vt == VLOCAL) {  /* 't' is in a register? */
        freereg(fs, e->u.ind.t);
        /* table access - set specialized op codes if array types are detected */
        if (e->ravi_type == RAVI_TARRAYFLT &&
            e->u.ind.key_type == RAVI_TNUMINT)
          op = OP_RAVI_GETTABLE_AF;
        else if (e->ravi_type == RAVI_TARRAYINT &&
```

```
                  e->u.ind.key_type == RAVI_TNUMINT)
          op = OP_RAVI_GETTABLE_AI;
        else
          op = OP_GETTABLE;
        if (e->ravi_type == RAVI_TARRAYFLT || e->ravi_type == RAVI_TARRAYINT)
          /* set the type of resulting expression */
          e->ravi_type = e->ravi_type == RAVI_TARRAYFLT ?
                           RAVI_TNUMFLT : RAVI_TNUMINT;
      }
      e->u.info = luaK_codeABC(fs, op, 0, e->u.ind.t, e->u.ind.idx);
      e->k = VRELOCABLE;
      DEBUG_EXPR(raviY_printf(fs, "luaK_dischargevars (VINDEXED->VRELOCABLE) %e\n", e));
      break;
    }
    case VVARARG:
    case VCALL: {
      luaK_setoneret(fs, e);
      break;
    }
    default: break;  /* there is one value available (somewhere) */
  }
}
```

4.6.5 fornum statements

The Lua fornum statements create special variables. In order to allows the loop variable to be used in expressions within the loop body we need to set the types of these variables. This is handled in `fornum()` as shown below. Additional complexity is due to the fact that Ravi tries to detect when fornum loops use positive integer step and if this step is `1`; specialized bytecodes are generated for these scenarios.

```
typedef struct Fornuminfo {
  ravitype_t type;
  int is_constant;
  int int_value;
} Fornuminfo;

/* parse the single expressions needed in numerical for loops
 * called by fornum()
 */
static int exp1 (LexState *ls, Fornuminfo *info) {
  /* Since the local variable in a fornum loop is local to the loop and does
   * not use any variable in outer scope we don't need to check its
   * type - also the loop is already optimised so no point trying to
   * optimise the iteration variable
   */
  expdesc e;
  int reg;
  e.ravi_type = RAVI_TANY;
  expr(ls, &e);
  DEBUG_EXPR(raviY_printf(ls->fs, "fornum exp -> %e\n", &e));
  info->is_constant = (e.k == VKINT);
  info->int_value = info->is_constant ? e.u.ival : 0;
  luaK_exp2nextreg(ls->fs, &e);
  lua_assert(e.k == VNONRELOC);
  reg = e.u.info;
  info->type = e.ravi_type;
  return reg;
```

```
}

/* parse a for loop body for both versions of the for loop
 * called by fornum(), forlist()
 */
static void forbody (LexState *ls, int base, int line, int nvars, int isnum, Fornuminfo *info) {
  /* forbody -> DO block */
  BlockCnt bl;
  OpCode forprep_inst = OP_FORPREP, forloop_inst = OP_FORLOOP;
  FuncState *fs = ls->fs;
  int prep, endfor;
  adjustlocalvars(ls, 3);  /* control variables */
  checknext(ls, TK_DO);
  if (isnum) {
    ls->fs->f->ravi_jit.jit_flags = 1;
    if (info && info->is_constant && info->int_value > 1) {
      forprep_inst = OP_RAVI_FORPREP_IP;
      forloop_inst = OP_RAVI_FORLOOP_IP;
    }
    else if (info && info->is_constant && info->int_value == 1) {
      forprep_inst = OP_RAVI_FORPREP_I1;
      forloop_inst = OP_RAVI_FORLOOP_I1;
    }
  }
  prep = isnum ? luaK_codeAsBx(fs, forprep_inst, base, NO_JUMP) : luaK_jump(fs);
  enterblock(fs, &bl, 0);  /* scope for declared variables */
  adjustlocalvars(ls, nvars);
  luaK_reserveregs(fs, nvars);
  block(ls);
  leaveblock(fs);  /* end of scope for declared variables */
  luaK_patchtohere(fs, prep);
  if (isnum)  /* numeric for? */
    endfor = luaK_codeAsBx(fs, forloop_inst, base, NO_JUMP);
  else {  /* generic for */
    luaK_codeABC(fs, OP_TFORCALL, base, 0, nvars);
    luaK_fixline(fs, line);
    endfor = luaK_codeAsBx(fs, OP_TFORLOOP, base + 2, NO_JUMP);
  }
  luaK_patchlist(fs, endfor, prep + 1);
  luaK_fixline(fs, line);
}

/* parse a numerical for loop, calls forbody()
 * called from forstat()
 */
static void fornum (LexState *ls, TString *varname, int line) {
  /* fornum -> NAME = exp1,exp1[,exp1] forbody */
  FuncState *fs = ls->fs;
  int base = fs->freereg;
  LocVar *vidx, *vlimit, *vstep, *vvar;
  new_localvarliteral(ls, "(for index)");
  new_localvarliteral(ls, "(for limit)");
  new_localvarliteral(ls, "(for step)");
  new_localvar(ls, varname, RAVI_TANY);
  /* The fornum sets up its own variables as above.
     These are expected to hold numeric values - but from Ravi's
     point of view we need to know if the variable is an integer or
     double. So we need to check if this can be determined from the
```

```
     fornum expressions. If we can then we will set the
     fornum variables to the type we discover.
  */
  vidx = &fs->f->locvars[fs->nlocvars - 4]; /* index variable - not yet active so get it from locvars
  vlimit = &fs->f->locvars[fs->nlocvars - 3]; /* index variable - not yet active so get it from locva
  vstep = &fs->f->locvars[fs->nlocvars - 2]; /* index variable - not yet active so get it from locvar
  vvar = &fs->f->locvars[fs->nlocvars - 1]; /* index variable - not yet active so get it from locvars
  checknext(ls, '=');
  /* get the type of each expression */
  Fornuminfo tidx = { RAVI_TANY,0,0 }, tlimit = { RAVI_TANY,0,0 }, tstep = { RAVI_TNUMINT,0,0 };
  Fornuminfo *info = NULL;
  exp1(ls, &tidx);  /* initial value */
  checknext(ls, ',');
  exp1(ls, &tlimit);  /* limit */
  if (testnext(ls, ','))
    exp1(ls, &tstep);  /* optional step */
  else {  /* default step = 1 */
    tstep.is_constant = 1;
    tstep.int_value = 1;
    luaK_codek(fs, fs->freereg, luaK_intK(fs, 1));
    luaK_reserveregs(fs, 1);
  }
  if (tidx.type == tlimit.type && tlimit.type == tstep.type && (tidx.type == RAVI_TNUMFLT || tidx.typ
    if (tidx.type == RAVI_TNUMINT && tstep.is_constant)
      info = &tstep;
    /* Ok so we have an integer or double */
    vidx->ravi_type = vlimit->ravi_type = vstep->ravi_type = vvar->ravi_type = tidx.type;
    DEBUG_VARS(raviY_printf(fs, "fornum -> setting type for index %v\n", vidx));
    DEBUG_VARS(raviY_printf(fs, "fornum -> setting type for limit %v\n", vlimit));
    DEBUG_VARS(raviY_printf(fs, "fornum -> setting type for step %v\n", vstep));
    DEBUG_VARS(raviY_printf(fs, "fornum -> setting type for variable %v\n", vvar));
  }
  forbody(ls, base, line, 1, 1, info);
}
```

4.7 Handling of Upvalues

Upvalues can be used to update local variables that have static typing specified. So this means that upvalues need to be annotated with types as well and any operation that updates an upvalue must be type checked. To support this the Lua parser has been enhanced to record the type of an upvalue in `Upvaldesc`:

```
/*
** Description of an upvalue for function prototypes
*/
typedef struct Upvaldesc {
  TString *name;  /* upvalue name (for debug information) */
  ravitype_t type; /* RAVI type of upvalue */
  lu_byte instack;  /* whether it is in stack */
  lu_byte idx;  /* index of upvalue (in stack or in outer function's list) */
} Upvaldesc;
```

Whenever a new upvalue is referenced, we assign the type of the the upvalue to the expression in function `singlevaraux()` - relevant code is shown below:

```
static int singlevaraux (FuncState *fs, TString *n, expdesc *var, int base) {
  /* ... omitted code ... */
```

```
    int idx = searchupvalue(fs, n);  /* try existing upvalues */
    if (idx < 0) {  /* not found? */
      if (singlevaraux(fs->prev, n, var, 0) == VVOID) /* try upper levels */
        return VVOID;  /* not found; is a global */
      /* else was LOCAL or UPVAL */
      idx  = newupvalue(fs, n, var);  /* will be a new upvalue */
    }
    init_exp(var, VUPVAL, idx, fs->f->upvalues[idx].type); /* RAVI : set upvalue type */
    return VUPVAL;
    /* ... omitted code ... */
}
```

The function `newupvalue()` sets the type of a new upvalue:

```
/* create a new upvalue */
static int newupvalue (FuncState *fs, TString *name, expdesc *v) {
  Proto *f = fs->f;
  int oldsize = f->sizeupvalues;
  checklimit(fs, fs->nups + 1, MAXUPVAL, "upvalues");
  luaM_growvector(fs->ls->L, f->upvalues, fs->nups, f->sizeupvalues,
                  Upvaldesc, MAXUPVAL, "upvalues");
  while (oldsize < f->sizeupvalues) f->upvalues[oldsize++].name = NULL;

  f->upvalues[fs->nups].instack = (v->k == VLOCAL);
  f->upvalues[fs->nups].idx = cast_byte(v->u.info);
  f->upvalues[fs->nups].name = name;
  f->upvalues[fs->nups].type = v->ravi_type;
  luaC_objbarrier(fs->ls->L, f, name);
  return fs->nups++;
}
```

When we need to generate assignments to an upvalue (OP_SETUPVAL) we need to use more specialized opcodes that do the necessary conversion at runtime. This is handled in `luaK_storevar()` in `lcode.c`:

```
/* Emit store for LHS expression. */
void luaK_storevar (FuncState *fs, expdesc *var, expdesc *ex) {
  switch (var->k) {
    /* ... omitted code .. */
    case VUPVAL: {
      OpCode op = check_valid_setupval(fs, var, ex);
      int e = luaK_exp2anyreg(fs, ex);
      luaK_codeABC(fs, op, e, var->u.info, 0);
      break;
    }
    /* ... omitted code ... */
  }
}

static OpCode check_valid_setupval(FuncState *fs, expdesc *var, expdesc *ex) {
  OpCode op = OP_SETUPVAL;
  if (var->ravi_type != RAVI_TANY && var->ravi_type != ex->ravi_type) {
    if (var->ravi_type == RAVI_TNUMINT)
      op = OP_RAVI_SETUPVALI;
    else if (var->ravi_type == RAVI_TNUMFLT)
      op = OP_RAVI_SETUPVALF;
    else if (var->ravi_type == RAVI_TARRAYINT)
      op = OP_RAVI_SETUPVALAI;
    else if (var->ravi_type == RAVI_TARRAYFLT)
      op = OP_RAVI_SETUPVALAF;
```

```
    else
      luaX_syntaxerror(fs->ls,
                    luaO_pushfstring(fs->ls->L, "Invalid assignment of "
                                                "upvalue: upvalue type "
                                                "%d, expression type %d",
                                     var->ravi_type, ex->ravi_type));
  }
  return op;
}
```

4.8 VM Enhancements

A number of new opcodes are introduced to allow type specific operations.

Currently there are specialized versions of `ADD`, `SUB`, `MUL` and `DIV` operations. This will be extended to cover additional operators such as `IDIV`. The `ADD` and `MUL` operations are implemented in a similar way. Both allow a second operand to be encoded directly in the `C` operand - when the value is a constant in the range [0,127].

One thing to note is that apart from division if an operation involves constants it is folded by Lua. Divisions are treated specially - an expression involving the `0` constant is not folded, even when the `0` is a numerator. Also worth noting is that DIV operator results in a float even when two integers are divided; you have to use `IDIV` to get an integer result - this opcode triggered in Lua 5.3 when the `//` operator is used.

A divide by zero when using integers causes a run time error, whereas for floating point operation the result is NaN.

CHAPTER 5

LLVM Notes

The notes below apply to LLVM 3.5.1 unless noted otherwise. All reflect my understanding - so if anything here is incorrect please log an issue and I will correct.

5.1 Structs and Unions

LLVM does not support defining union types so we need to basically use a `struct` of appropriate size and cast it as we need. The main thing to be careful about is to ensure that the `struct` is of the same size as the `union`.

An example is:

```
///*
//** Information about a call.
//** When a thread yields, 'func' is adjusted to pretend that the
//** top function has only the yielded values in its stack; in that
//** case, the actual 'func' value is saved in field 'extra'.
//** When a function calls another with a continuation, 'extra' keeps
//** the function index so that, in case of errors, the continuation
//** function can be called with the correct top.
//*/
// typedef struct CallInfo {
//  StkId func;  /* function index in the stack */
//  StkId     top;  /* top for this function */
//  struct CallInfo *previous, *next;  /* dynamic call link */
//  union {
//    struct {  /* only for Lua functions */
//      StkId base;  /* base for this function */
//      const Instruction *savedpc;
//    } l;
//    struct {  /* only for C functions */
//      lua_KFunction k;  /* continuation in case of yields */
//      ptrdiff_t old_errfunc;
//      lua_KContext ctx;  /* context info. in case of yields */
//    } c;
//  } u;
//  ptrdiff_t extra;
//  short nresults;  /* expected number of results from this function */
//  lu_byte callstatus;
//} CallInfo;
```

Above the union `u` has two members of unequal size. To handle this I created the following two sub-types of equal size - note the extra dummy field in the first type:

```
elements.clear();
elements.push_back(StkIdT);        /* base */
elements.push_back(pInstructionT); /* savedpc */
elements.push_back(
    C_ptrdiff_t); /* dummy to make this same size as the other member */
CallInfo_lT = llvm::StructType::create(elements);

elements.clear();
elements.push_back(plua_KFunctionT); /* k */
elements.push_back(C_ptrdiff_t);     /* old_errfunc */
elements.push_back(lua_KContextT);   /* ctx */
CallInfo_cT = llvm::StructType::create(elements);
```

Then as I intend to use the `u.l` field more often, I used the following definition for `CallInfo`:

```
CallInfoT = llvm::StructType::create(context, "ravi.CallInfo");
pCallInfoT = llvm::PointerType::get(CallInfoT, 0);
elements.clear();
elements.push_back(StkIdT);     /* func */
elements.push_back(StkIdT);     /* top */
elements.push_back(pCallInfoT); /* previous */
elements.push_back(pCallInfoT); /* next */
elements.push_back(
    CallInfo_lT); /* u.l  - as we will typically access the lua call details
                     */
elements.push_back(C_ptrdiff_t);                    /* extra */
elements.push_back(llvm::Type::getInt16Ty(context)); /* nresults */
elements.push_back(lu_byteT);                       /* callstatus */
CallInfoT->setBody(elements);
```

5.2 JIT Compilation Error on Windows

Note: The latest LLVM version from LLVM source repository appears to support COEFF format. So below applies to version 3.6.x and 3.5.x.

On Windows when we attempt to JIT compile we get an error saying incompatible object format Reading posts on mailing lists I found that the issue is that COEFF format is not supported and therefore we need to set -elf as the object format:

```
#include "llvm/Support/Host.h"

/* some code */

#ifdef _WIN32
  auto triple = llvm::sys::getProcessTriple();
  module->setTargetTriple(triple + "-elf");
#endif
```

5.3 Memory Management

It appears that most things in LLVM are owned by the parent object and when the parent object is deleted the children go too. So in my code the main objects I delete are the `ExecutionEngine` and `Module`. Once a module is associated with an engine then only the engine needs to be explicitly deleted is my understanding.

It doesn't help that the tutorial available does not attempt to delete objects / release memory!

5.4 MCJIT Engines, Modules and Functions

Functions live inside Modules but once a Module is finalized (compiled) then no further functions can be added to it. Although an `MCJIT` instance (engine) can support multiples modules, the recommendation is to ensure each module is assigned its own engine. The rationale for this is not explained.

5.5 Struct Assign

My understanding is that to perform assignment of a struct value, one must call the intrinsic `memcpy` function. Example of code that does this:

```
llvm::Value *src;
llvm::Value *dest;

// First get the declaration for the inttrinsic memcpy
llvm::SmallVector<llvm::Type *, 3> vec;
vec.push_back(def->types->C_pcharT);  /* i8 */
vec.push_back(def->types->C_pcharT);  /* i8 */
vec.push_back(def->types->C_intT);
llvm::Function *f = llvm::Intrinsic::getDeclaration(
    def->raviF->module(), llvm::Intrinsic::memcpy, vec);
lua_assert(f);

// Cast src and dest to i8*
llvm::Value *dest_ptr =
    def->builder->CreateBitCast(dest, def->types->C_pcharT);
llvm::Value *src_ptr = def->builder->CreateBitCast(src, def->types->C_pcharT);

// Create call to intrinsic memcpy
values_.clear();
values_.push_back(dest_ptr);
values_.push_back(src_ptr);
values_.push_back(llvm::ConstantInt::get(def->types->C_intT, sizeof(TValue)));
values_.push_back(
    llvm::ConstantInt::get(def->types->C_intT, sizeof(L_Umaxalign)));
values_.push_back(def->types->kFalse);
def->builder->CreateCall(f, values_);
```

Note that the call to memcpy supply an alignment.

5.6 Accessing `extern` functions from JIT compiled code

If the JITed function needs to access `extern` functions that are statically linked and not exported as dynamic symbols (e.g. in Visual C++) then we need some extra steps. From reading posts on the subject it appears that the way to do this is to add a global mapping in the `ExecutionEngine` by calling the `addGlobalMapping()` method. However this doesn't work with MCJIT due to a bug! So we need to use a workaround. Apparently there are two solutions:

- Create a custom memory manager that resolves the `extern` functions.
- Add the symbol to the global symbols by calling `llvm::sys::DynamicLibrary::AddSymbol()`.

I am using the latter approach for now.

5.7 GEP instruction

The GEP instruction cannot compute addresses of fields in a pointer member - as the pointer needs to be 'loaded' first. This is explained in the GEP FAQ.

5.8 Hooking up Optimization Passes

The LLVM documentation does not provide guidance on how the optimization passes should be hooked up. There are descriptions of what the passes do, but if you are new to LLVM and trying to work out which passes to use and in what order, then there is not much help available. The Kaleidoscope Sample shows a small example of how optimization passes may be hooked up.

Fortunately it seems that there is a PassManagerBuilder component that allows easy setup of the standard passes for a C like language. Unfortunately there isn't much guidance on how to use this either. The best source of information I found was an example toy compiler by David Chisnall.

5.9 Links

- Mapping High Level Constructs to LLVM IR
- IRBuilder Sample
- Using MCJIT with Kaleidoscope
- Object format issue on Windows
- ExecutionEngine::addGlobalMapping() bug in MCJIT
- LLVM Notes
- Implementing Domain-Specific Languages with LLVM.

CHAPTER 6

LLVM First Steps

Note that the discussion below is for LLVM 3.5.1.

Although there appears to be a lot of documentation in the LLVM site surprisingly some basic information is hard to find. The main source of guidance for creating a JIT is in the example toy language Kaleidoscope. But here too there are several versions - so you have to pick the right version that is compatible with the LLVM version you are using.

A Lua JITed function will execute in the context of Lua. So it needs to be able to access the `lua_State` and its various structures. So I wanted a sample that demonstrates passing a pointer to a structure and accessing it within the JITed function.

The initial test program I created is meant to be a "hello world" type test but covering the functionility described above. The test I want to run is:

```
// Creating a function that takes pointer to struct as argument
// The function gets value from one of the fields in the struct
// And returns it
// The equivalent C program is:
//
// extern int printf(const char *, ...);
//
// struct GCObject {
//   struct GCObject *next;
//   unsigned char a;
//   unsigned char b;
// };
//
// int testfunc(struct GCObject *obj) {
//   printf("value = %d\n", obj->a);
//   return obj->a;
// }
```

You can view the test program at test_llvm.cpp. It is also reproduced below.

I used the new `MCJIT` engine in my test. It seems that this engine compiles modules rather than individual functions - and once compiled a module cannot be modified. So in the Lua context we need to create a new module everytime we JIT compile a function - or alternatively we JIT compile a whole Lua source file including all its functions into a single module.

I found the blog post Using MCJIT with Kaleidoscope useful in understanding some finer points about using `MCJIT`.

The Lua GCObject structure in lobject.h we need is:

```
typedef struct RaviGCObject {
  struct RaviGCObject *next;
  unsigned char b1;
```

```
  unsigned char b2;
} RaviGCObject;
```

Our prototype for the JITted function:

```
typedef int (*myfunc_t)(RaviGCObject *);
```

Get global context - not sure what the impact is of sharing:

```
llvm::LLVMContext &context = llvm::getGlobalContext();
```

Module is the translation unit:

```
std::unique_ptr<llvm::Module> theModule =
  std::unique_ptr<llvm::Module>(new llvm::Module("ravi", context));
llvm::Module *module = theModule.get();
llvm::IRBuilder<> builder(context);
```

On Windows we get error saying incompatible object format Reading posts on mailining lists I found that the issue is that COEFF format is not supported and therefore we need to set -elf as the object format:

```
#ifdef _WIN32
  auto triple = llvm::sys::getProcessTriple();
  module->setTargetTriple(triple + "-elf");
#endif
```

create a GCObject structure as defined in lobject.h:

```
llvm::StructType *structType =
    llvm::StructType::create(context, "RaviGCObject");
llvm::PointerType *pstructType =
  llvm::PointerType::get(structType, 0); // pointer to RaviGCObject
std::vector<llvm::Type *> elements;
elements.push_back(pstructType);
elements.push_back(llvm::Type::getInt8Ty(context));
elements.push_back(llvm::Type::getInt8Ty(context));
structType->setBody(elements);
structType->dump();
```

Create printf declaration:

```
std::vector<llvm::Type *> args;
args.push_back(llvm::Type::getInt8PtrTy(context));
// accepts a char*, is vararg, and returns int
llvm::FunctionType *printfType =
    llvm::FunctionType::get(builder.getInt32Ty(), args, true);
llvm::Constant *printfFunc =
    module->getOrInsertFunction("printf", printfType);
```

Create the testfunc():

```
args.clear();
args.push_back(pstructType);
llvm::FunctionType *funcType =
  llvm::FunctionType::get(builder.getInt32Ty(), args, false);
llvm::Function *mainFunc = llvm::Function::Create(
  funcType, llvm::Function::ExternalLinkage, "testfunc", module);
llvm::BasicBlock *entry =
  llvm::BasicBlock::Create(context, "entrypoint", mainFunc);
builder.SetInsertPoint(entry);
```

The printf format string:

```
llvm::Value *formatStr = builder.CreateGlobalStringPtr("value = %d\n");
```

Get the first argument which is RaviGCObject *:

```
auto argiter = mainFunc->arg_begin();
llvm::Value *arg1 = argiter++;
arg1->setName("obj");
```

Now we need a GEP for the second field in RaviGCObject:

```
std::vector<llvm::Value *> values;
llvm::APInt zero(32, 0);
llvm::APInt one(32, 1);
// This is the array offset into RaviGCObject*
values.push_back(
   llvm::Constant::getIntegerValue(llvm::Type::getInt32Ty(context), zero));
// This is the field offset
values.push_back(
  llvm::Constant::getIntegerValue(llvm::Type::getInt32Ty(context), one));
```

Create the GEP value:

```
llvm::Value *arg1_a = builder.CreateGEP(arg1, values, "ptr");
```

Now retrieve the data from the pointer address:

```
llvm::Value *tmp1 = builder.CreateLoad(arg1_a, "a");
```

As the retrieved value is a byte - convert to int i:

```
llvm::Value *tmp2 =
  builder.CreateZExt(tmp1, llvm::Type::getInt32Ty(context), "i");
```

Call the printf function:

```
values.clear();
values.push_back(formatStr);
values.push_back(tmp2);
builder.CreateCall(printfFunc, values);
```

return i:

```
builder.CreateRet(tmp2);
module->dump();
```

Lets create the MCJIT engine:

```
std::string errStr;
auto engine = llvm::EngineBuilder(module)
                .setErrorStr(&errStr)
                .setEngineKind(llvm::EngineKind::JIT)
                .setUseMCJIT(true)
                .create();
if (!engine) {
  llvm::errs() << "Failed to construct MCJIT ExecutionEngine: " << errStr
               << "\n";
  return 1;
}
```

Now lets compile our function into machine code:

```
std::string funcname = "testfunc";
myfunc_t funcptr = (myfunc_t)engine->getFunctionAddress(funcname);
if (funcptr == nullptr) {
  llvm::errs() << "Failed to obtain compiled function\n";
  return 1;
}
```

Run the function and test results:

```
RaviGCObject obj = {NULL, 42, 65};
int ans = funcptr(&obj);
printf("The answer is %d\n", ans);
return ans == 42 ? 0 : 1;
```

6.1 Accessing `extern` functions from JIT compiled code

The JITed function needs to access `extern` Lua functions. We need a way to map these to make these visible to the JITed code. Simply declaring the functions `extern` only appears to work if the functios are available as exported symbols in dynamic libraries, e.g. the call to `printf` above.

From reading posts on the subject it appears that the way to do this is to add a global mapping in the `ExecutionEngine` by calling the `addGlobalMapping()` method. However this doesn't work with MCJIT due to a bug! So we need to use a workaround. Apparently there are two solutions:

- Create a custom memory manager that resolves the `extern` functions.
- Add the symbol to the global symbols by calling `llvm::sys::DynamicLibrary::AddSymbol()`.

I am using the latter approach for now.

6.2 Memory Management in LLVM

Curiously LLVM docs do not say much about how memory should be managed. I am still trying to figure this out, but in general it seems that there is hierarchy of ownership. Example: `ExecutionEngine` owns the `Module`. By deleting the parent the 'owned' objects are automatically deleted.

6.3 Links

- Object format issue on Windows
- ExecutionEngine::addGlobalMapping() bug in MCJIT
- LLVM Notes

CHAPTER 7

Ravi LLVM JIT Infrastructure

Ravi is using the LLVM MCJIT infrastrucure for JIT compilation. This poses some challenges as we have to worry about how modules are related to MCJIT engines. It seems from the official samples and from feedback on mailing lists that due to limitations in MCJIT it is advisable to ensure that each module has its own MCJIT engine. Furthermore it appears that once a module is compiled to machine code i.e. JITed then no further changes can be made to the module, so it is not possible to keep adding functions to the same module.

The Kaleidoscope MCJIT samples try to optimise the module to function mapping by ensuring that a module is reused for each new function until a function is compiled. Once any function in the module is compiled, then a new module is allocated. Each module of course gets its own MCJIT instance.

For Ravi, I want to have an abstraction layer that hides all this detail and allows us to change the implementation strategy without having to modify the rest of the system. In order to allow that I have created the following two interfaces.

7.1 RaviJITState interface

This is where the LLVM state is held for any Lua instance. An instance of this will be stored in the Lua State. The interface looks like this:

```
class RAVI_API RaviJITState {
public:
  RaviJITState();
  ~RaviJITState();

  // Create a function of specified type and linkage
  RaviJITFunction *createFunction(llvm::FunctionType *type,
                                  llvm::GlobalValue::LinkageTypes linkage,
                                  const std::string &name);


  // Get the LLVM Context
  llvm::LLVMContext &context();
};
```

7.2 RaviJITFunction interface

This interface is for each JITed function. It looks like this:

```
class RAVI_API RaviJITFunction {

public:
  // Destroy the function releasing any resources used
  // by the function - this can be called from Lua's
  // garbage collector when a function is finalized
  ~RaviJITFunction();

  // Compile the function if not already compiled and
  // return pointer to function
  void *compile();

  // Add declaration for an extern function that is not
  // loaded dynamically - i.e., is part of the the executable
  // and therefore not visible at runtime by name
  llvm::Constant *addExternFunction(llvm::FunctionType *type, void *address,
                                    const std::string &name);

  // Get the function name
  const std::string &name() const;

  // Get the LLVM Function object
  llvm::Function *function() const;

  // Get the Module that owns this function
  llvm::Module *module() const;

  // Get the MCJIT instance that owns this function
  llvm::ExecutionEngine *engine() const;
};
```

7.3 Example Usage

What above does is abstracts away all the implementation details. Here is a test program that uses the above two interfaces:

```
// This mirrors the Lua GCObject structure in lobject.h
typedef struct RaviGCObject {
  struct RaviGCObject *next;
  unsigned char b1;
  unsigned char b2;
} RaviGCObject;

// Our prototype for the JITted function
typedef int (*myfunc_t)(RaviGCObject *);

extern "C" int mytest(RaviGCObject *obj) {
  printf("value = %d\n", obj->b1);
  return obj->b1;
}

// This version of the test calls mytest() rather than
// printf() as in test1(). Also we use RaviJITState and related
// infrastructure
int test2() {
```

```
RaviJITState jitState;

llvm::LLVMContext &context = jitState.context();
llvm::IRBuilder<> builder(context);

// create a GCObject structure as defined in lobject.h
llvm::StructType *structType =
  llvm::StructType::create(context, "RaviGCObject");
llvm::PointerType *pstructType =
  llvm::PointerType::get(structType, 0); // pointer to RaviGCObject
std::vector<llvm::Type *> elements;
elements.push_back(pstructType);
elements.push_back(llvm::Type::getInt8Ty(context));
elements.push_back(llvm::Type::getInt8Ty(context));
structType->setBody(elements);
structType->dump();

// Create declaration for mytest
// int mytest(RaviGCObject *obj)
std::vector<llvm::Type *> args;
args.push_back(pstructType);
llvm::FunctionType *mytestFuncType =
  llvm::FunctionType::get(builder.getInt32Ty(), args, false);

// Create the testfunc()
args.clear();
args.push_back(pstructType);
llvm::FunctionType *funcType =
  llvm::FunctionType::get(builder.getInt32Ty(), args, false);
RaviJITFunction *func = jitState.createFunction(
  funcType, llvm::Function::ExternalLinkage, "testfunc");

llvm::Function *mainFunc = func->function();
llvm::BasicBlock *entry =
  llvm::BasicBlock::Create(context, "entrypoint", mainFunc);
builder.SetInsertPoint(entry);

// Get the first argument which is RaviGCObject *
auto argiter = mainFunc->arg_begin();
llvm::Value *arg1 = argiter++;
arg1->setName("obj");

// Add an extern int mytest(RaviGCObject *obj) and link this
// to mytest()
llvm::Constant *mytestFunc =
  func->addExternFunction(mytestFuncType, &mytest, "mytest");

// Call the mytest() function
std::vector<llvm::Value *> values;
values.push_back(arg1);
llvm::Value *tmp2 = builder.CreateCall(mytestFunc, values, "i");

// return i
builder.CreateRet(tmp2);
func->dump();

// Now lets compile our function into machine code
myfunc_t funcptr = (myfunc_t)func->compile();
```

```
  if (funcptr == nullptr) {
    llvm::errs() << "Failed to obtain compiled function\n";
    return 1;
  }

  // Run the function and test results.
  RaviGCObject obj = {NULL, 42, 65};
  int ans = funcptr(&obj);
  printf("The answer is %d\n", ans);
  return ans == 42 ? 0 : 1;
}
```

CHAPTER 8

LLVM Compilation hooks in Ravi

The current approach is Ravi is that a Lua function can be compiled at the function level. (Note that this is the plan - I am working on the implementation).

In terms of changes to support this - we essentially have following. First we have a bunch of C functions - think of these are the compiler API:

```
#ifdef __cplusplus
extern "C" {
#endif

struct lua_State;
struct Proto;

/* Initialise the JIT engine */
int raviV_initjit(struct lua_State *L);

/* Shutdown the JIT engine */
void raviV_close(struct lua_State *L);

/* Compile the given function if possible */
int raviV_compile(struct lua_State *L, struct Proto *p);

/* Free the JIT structures associated with the prototype */
void raviV_freeproto(struct lua_State *L, struct Proto *p);

#ifdef __cplusplus
}
#endif
```

Next the `Proto` struct definition has some extra fields:

```
typedef struct RaviJITProto {
  lu_byte jit_status; // 0=not compiled, 1=can't compile, 2=compiled, 3=freed
  void *jit_data;
  lua_CFunction jit_function;
} RaviJITProto;

/*
** Function Prototypes
*/
typedef struct Proto {
  CommonHeader;
  lu_byte numparams;  /* number of fixed parameters */
```

```
  lu_byte is_vararg;
  lu_byte maxstacksize;  /* maximum stack used by this function */
  int sizeupvalues;  /* size of 'upvalues' */
  int sizek;  /* size of 'k' */
  int sizecode;
  int sizelineinfo;
  int sizep;  /* size of 'p' */
  int sizelocvars;
  int linedefined;
  int lastlinedefined;
  TValue *k;  /* constants used by the function */
  Instruction *code;
  struct Proto **p;  /* functions defined inside the function */
  int *lineinfo;  /* map from opcodes to source lines (debug information) */
  LocVar *locvars;  /* information about local variables (debug information) */
  Upvaldesc *upvalues;  /* upvalue information */
  struct LClosure *cache;  /* last created closure with this prototype */
  TString  *source;  /* used for debug information */
  GCObject *gclist;
  /* RAVI */
  RaviJITProto ravi_jit;
} Proto;
```

The `ravi_jit` member is initialized in `lfunc.c`:

```
Proto *luaF_newproto (lua_State *L) {
  GCObject *o = luaC_newobj(L, LUA_TPROTO, sizeof(Proto));
  Proto *f = gco2p(o);
  f->k = NULL;
  /* code ommitted */
  f->ravi_jit.jit_data = NULL;
  f->ravi_jit.jit_function = NULL;
  f->ravi_jit.jit_status = 0; /* not compiled */
  return f;
}
```

The corresponding function to free is:

```
void luaF_freeproto (lua_State *L, Proto *f) {
  raviV_freeproto(L, f);
  luaM_freearray(L, f->code, f->sizecode);
  luaM_freearray(L, f->p, f->sizep);
  luaM_freearray(L, f->k, f->sizek);
  luaM_freearray(L, f->lineinfo, f->sizelineinfo);
  luaM_freearray(L, f->locvars, f->sizelocvars);
  luaM_freearray(L, f->upvalues, f->sizeupvalues);
  luaM_free(L, f);
}
```

When a Lua Function is called it goes through `luaD_precall()` in `ldo.c`. This has been modified to invoke the compiler / use compiled version:

```
/*
** returns true if function has been executed (C function)
*/
int luaD_precall (lua_State *L, StkId func, int nresults) {
  lua_CFunction f;
  CallInfo *ci;
  int n;  /* number of arguments (Lua) or returns (C) */
```

```
ptrdiff_t funcr = savestack(L, func);
switch (ttype(func)) {

    /* omitted */

case LUA_TLCL: {  /* Lua function: prepare its call */
  CallInfo *prevci = L->ci; /* RAVI - for validation */
  StkId base;
  Proto *p = clLvalue(func)->p;
  n = cast_int(L->top - func) - 1;  /* number of real arguments */
  luaD_checkstack(L, p->maxstacksize);
  for (; n < p->numparams; n++)
    setnilvalue(L->top++);  /* complete missing arguments */
  if (!p->is_vararg) {
    func = restorestack(L, funcr);
    base = func + 1;
  }
  else {
    base = adjust_varargs(L, p, n);
    func = restorestack(L, funcr);  /* previous call can change stack */
  }
  ci = next_ci(L);  /* now 'enter' new function */
  ci->nresults = nresults;
  ci->func = func;
  ci->u.l.base = base;
  ci->top = base + p->maxstacksize;
  lua_assert(ci->top <= L->stack_last);
  ci->u.l.savedpc = p->code;  /* starting point */
  ci->callstatus = CIST_LUA;
  ci->jitstatus = 0;
  L->top = ci->top;
  luaC_checkGC(L);  /* stack grow uses memory */
  if (L->hookmask & LUA_MASKCALL)
    callhook(L, ci);
  if (compile) {
    if (p->ravi_jit.jit_status == 0) {
      /* not compiled */
      raviV_compile(L, p, 0);
    }
    if (p->ravi_jit.jit_status == 2) {
      /* compiled */
      lua_assert(p->ravi_jit.jit_function != NULL);
      ci->jitstatus = 1;
      /* As JITed function is like a C function
       * employ the same restrictions on recursive
       * calls as for C functions
       */
      if (++L->nCcalls >= LUAI_MAXCCALLS) {
        if (L->nCcalls == LUAI_MAXCCALLS)
          luaG_runerror(L, "C stack overflow");
        else if (L->nCcalls >= (LUAI_MAXCCALLS + (LUAI_MAXCCALLS >> 3)))
          luaD_throw(L, LUA_ERRERR);  /* error while handing stack error */
      }
      /* Disable YIELDs - so JITed functions cannot
       * yield
       */
      L->nny++;
      (*p->ravi_jit.jit_function)(L);
```

```
        L->nny--;
        L->nCcalls--;
        lua_assert(L->ci == prevci);
        /* Return a different value from 1 to
         * allow luaV_execute() to distinguish between
         * JITed function and true C function
         */
        return 2;
      }
    }
    return 0;
  }
  default: {  /* not a function */

        /* omitted */
  }
  }
}
```

Note that the above returns 2 if compiled Lua function is called. The behaviour in `lvm.c` is similar to that when a C function is called.

CHAPTER 9

Lua Types in LLVM

We need to map Lua types to equivalent type definitions in LLVM. In Ravi we do hold all the type definitions in a struct as shown below:

```
struct LuaLLVMTypes {

llvm::Type *C_intptr_t;
llvm::Type *C_size_t;
llvm::Type *C_ptrdiff_t;

llvm::Type *lua_NumberT;
llvm::Type *lua_IntegerT;
llvm::Type *lua_UnsignedT;
llvm::Type *lua_KContextT;

llvm::FunctionType *lua_CFunctionT;
llvm::PointerType *plua_CFunctionT;

llvm::FunctionType *lua_KFunctionT;
llvm::PointerType *plua_KFunctionT;

llvm::FunctionType *lua_HookT;
llvm::PointerType *plua_HookT;

llvm::FunctionType *lua_AllocT;
llvm::PointerType *plua_AllocT;

llvm::Type *l_memT;
llvm::Type *lu_memT;

llvm::Type *lu_byteT;
llvm::Type *L_UmaxalignT;
llvm::Type *C_pcharT;

llvm::Type *C_intT;

llvm::StructType *lua_StateT;
llvm::PointerType *plua_StateT;

llvm::StructType *global_StateT;
llvm::PointerType *pglobal_StateT;

llvm::StructType *ravi_StateT;
llvm::PointerType *pravi_StateT;
```

```
llvm::StructType *GCObjectT;
llvm::PointerType *pGCObjectT;

llvm::StructType *ValueT;
llvm::StructType *TValueT;
llvm::PointerType *pTValueT;

llvm::StructType *TStringT;
llvm::PointerType *pTStringT;
llvm::PointerType *ppTStringT;

llvm::StructType *UdataT;
llvm::StructType *TableT;
llvm::PointerType *pTableT;

llvm::StructType *UpvaldescT;
llvm::PointerType *pUpvaldescT;

llvm::Type *ravitype_tT;
llvm::StructType *LocVarT;
llvm::PointerType *pLocVarT;

llvm::Type *InstructionT;
llvm::PointerType *pInstructionT;
llvm::StructType *LClosureT;
llvm::PointerType *pLClosureT;
llvm::PointerType *ppLClosureT;
llvm::PointerType *pppLClosureT;

llvm::StructType *RaviJITProtoT;
llvm::PointerType *pRaviJITProtoT;

llvm::StructType *ProtoT;
llvm::PointerType *pProtoT;
llvm::PointerType *ppProtoT;

llvm::StructType *UpValT;
llvm::PointerType *pUpValT;

llvm::StructType *CClosureT;
llvm::PointerType *pCClosureT;

llvm::StructType *TKeyT;
llvm::PointerType *pTKeyT;

llvm::StructType *NodeT;
llvm::PointerType *pNodeT;

llvm::StructType *lua_DebugT;
llvm::PointerType *plua_DebugT;

llvm::StructType *lua_longjumpT;
llvm::PointerType *plua_longjumpT;

llvm::StructType *MbufferT;
llvm::StructType *stringtableT;

llvm::PointerType *StkIdT;
```

```
llvm::StructType *CallInfoT;
llvm::StructType *CallInfo_cT;
llvm::StructType *CallInfo_lT;
llvm::PointerType *pCallInfoT;

llvm::FunctionType *jitFunctionT;

llvm::FunctionType *luaD_poscallT;

};
```

The actual definition of the types above is shown below:

```
static_assert(std::is_floating_point<lua_Number>::value &&
                  sizeof(lua_Number) == sizeof(double),
              "lua_Number is not a double");
lua_NumberT = llvm::Type::getDoubleTy(context);

static_assert(std::is_integral<lua_Integer>::value,
              "lua_Integer is not an integer type");
lua_IntegerT = llvm::Type::getIntNTy(context, sizeof(lua_Integer) * 8);

static_assert(sizeof(lua_Integer) == sizeof(lua_Unsigned),
              "lua_Integer and lua_Unsigned are of different size");
lua_UnsignedT = lua_IntegerT;

C_intptr_t = llvm::Type::getIntNTy(context, sizeof(intptr_t) * 8);
C_size_t = llvm::Type::getIntNTy(context, sizeof(size_t) * 8);
C_ptrdiff_t = llvm::Type::getIntNTy(context, sizeof(ptrdiff_t) * 8);
C_intT = llvm::Type::getIntNTy(context, sizeof(int) * 8);

static_assert(sizeof(size_t) == sizeof(lu_mem),
              "lu_mem size is not same as size_t");
lu_memT = C_size_t;

static_assert(sizeof(ptrdiff_t) == sizeof(l_mem),
              "l_mem size is not same as ptrdiff_t");
l_memT = C_ptrdiff_t;

static_assert(sizeof(L_Umaxalign) == sizeof(double),
              "L_Umaxalign is not same size as double");
L_UmaxalignT = llvm::Type::getDoubleTy(context);

lu_byteT = llvm::Type::getInt8Ty(context);
C_pcharT = llvm::Type::getInt8PtrTy(context);

InstructionT = C_intT;
pInstructionT = llvm::PointerType::get(InstructionT, 0);

lua_StateT = llvm::StructType::create(context, "ravi.lua_State");
plua_StateT = llvm::PointerType::get(lua_StateT, 0);

lua_KContextT = C_ptrdiff_t;

std::vector<llvm::Type *> elements;
elements.push_back(plua_StateT);
lua_CFunctionT = llvm::FunctionType::get(C_intT, elements, false);
plua_CFunctionT = llvm::PointerType::get(lua_CFunctionT, 0);
```

```
jitFunctionT = lua_CFunctionT;

elements.clear();
elements.push_back(plua_StateT);
elements.push_back(C_intT);
elements.push_back(lua_KContextT);
lua_KFunctionT = llvm::FunctionType::get(C_intT, elements, false);
plua_KFunctionT = llvm::PointerType::get(lua_KFunctionT, 0);

elements.clear();
elements.push_back(llvm::Type::getInt8PtrTy(context));
elements.push_back(llvm::Type::getInt8PtrTy(context));
elements.push_back(C_size_t);
elements.push_back(C_size_t);
lua_AllocT = llvm::FunctionType::get(llvm::Type::getInt8PtrTy(context),
                                     elements, false);
plua_AllocT = llvm::PointerType::get(lua_AllocT, 0);

lua_DebugT = llvm::StructType::create(context, "ravi.lua_Debug");
plua_DebugT = llvm::PointerType::get(lua_DebugT, 0);

elements.clear();
elements.push_back(plua_StateT);
elements.push_back(plua_DebugT);
lua_HookT = llvm::FunctionType::get(llvm::Type::getInt8PtrTy(context),
                                    elements, false);
plua_HookT = llvm::PointerType::get(lua_HookT, 0);

// struct GCObject {
//   GCObject *next;
//   lu_byte tt;
//   lu_byte marked
// };
GCObjectT = llvm::StructType::create(context, "ravi.GCObject");
pGCObjectT = llvm::PointerType::get(GCObjectT, 0);
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
GCObjectT->setBody(elements);

static_assert(sizeof(Value) == sizeof(lua_Number),
              "Value type is larger than lua_Number");
// In LLVM unions should be set to the largest member
// So in the case of a Value this is the double type
// union Value {
//   GCObject *gc;    /* collectable objects */
//   void *p;         /* light userdata */
//   int b;           /* booleans */
//   lua_CFunction f; /* light C functions */
//   lua_Integer i;   /* integer numbers */
//   lua_Number n;    /* float numbers */
// };
ValueT = llvm::StructType::create(context, "ravi.Value");
elements.clear();
elements.push_back(lua_NumberT);
ValueT->setBody(elements);
```

```
// struct TValue {
//   union Value value_;
//   int tt_;
// };
TValueT = llvm::StructType::create(context, "ravi.TValue");
elements.clear();
elements.push_back(ValueT);
elements.push_back(C_intT);
TValueT->setBody(elements);
pTValueT = llvm::PointerType::get(TValueT, 0);

StkIdT = pTValueT;

///*
//** Header for string value; string bytes follow the end of this structure
//** (aligned according to 'UTString'; see next).
//*/
// typedef struct TString {
//   GCObject *next;
//   lu_byte tt;
//   lu_byte marked
//   lu_byte extra;  /* reserved words for short strings; "has hash" for longs
//   */
//   unsigned int hash;
//   size_t len;  /* number of characters in string */
//   struct TString *hnext;  /* linked list for hash table */
// } TString;

///*
//** Ensures that address after this type is always fully aligned.
//*/
// typedef union UTString {
//  L_Umaxalign dummy;  /* ensures maximum alignment for strings */
//  TString tsv;
//} UTString;
TStringT = llvm::StructType::create(context, "ravi.TString");
pTStringT = llvm::PointerType::get(TStringT, 0);
ppTStringT = llvm::PointerType::get(pTStringT, 0);
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);  /* extra */
elements.push_back(C_intT);    /* hash */
elements.push_back(C_size_t);  /* len */
elements.push_back(pTStringT); /* hnext */
TStringT->setBody(elements);

// Table
TableT = llvm::StructType::create(context, "ravi.Table");
pTableT = llvm::PointerType::get(TableT, 0);

///*
//** Header for userdata; memory area follows the end of this structure
//** (aligned according to 'UUdata'; see next).
//*/
// typedef struct Udata {
//  GCObject *next;
```

```
//  lu_byte tt;
//  lu_byte marked
//  lu_byte ttuv_;  /* user value's tag */
//  struct Table *metatable;
//  size_t len;  /* number of bytes */
//  union Value user_;  /* user value */
//} Udata;
UdataT = llvm::StructType::create(context, "ravi.Udata");
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT); /* ttuv_ */
elements.push_back(pTableT);  /* metatable */
elements.push_back(C_size_t); /* len */
elements.push_back(ValueT);   /* user_ */
UdataT->setBody(elements);

///*
//** Description of an upvalue for function prototypes
//*/
// typedef struct Upvaldesc {
//  TString *name;  /* upvalue name (for debug information) */
//  lu_byte instack;  /* whether it is in stack */
//  lu_byte idx;  /* index of upvalue (in stack or in outer function's list)
//  */
//}Upvaldesc;
UpvaldescT = llvm::StructType::create(context, "ravi.Upvaldesc");
elements.clear();
elements.push_back(pTStringT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
UpvaldescT->setBody(elements);
pUpvaldescT = llvm::PointerType::get(UpvaldescT, 0);

///*
//** Description of a local variable for function prototypes
//** (used for debug information)
//*/
// typedef struct LocVar {
//  TString *varname;
//  int startpc;  /* first point where variable is active */
//  int endpc;    /* first point where variable is dead */
//  ravitype_t ravi_type; /* RAVI type of the variable - RAVI_TANY if unknown
//  */
//} LocVar;
ravitype_tT = llvm::Type::getIntNTy(context, sizeof(ravitype_t) * 8);
LocVarT = llvm::StructType::create(context, "ravi.LocVar");
elements.clear();
elements.push_back(pTStringT);   /* varname */
elements.push_back(C_intT);      /* startpc */
elements.push_back(C_intT);      /* endpc */
elements.push_back(ravitype_tT); /* ravi_type */
LocVarT->setBody(elements);
pLocVarT = llvm::PointerType::get(LocVarT, 0);

LClosureT = llvm::StructType::create(context, "ravi.LClosure");
pLClosureT = llvm::PointerType::get(LClosureT, 0);
```

```
ppLClosureT = llvm::PointerType::get(pLClosureT, 0);
pppLClosureT = llvm::PointerType::get(ppLClosureT, 0);

RaviJITProtoT = llvm::StructType::create(context, "ravi.RaviJITProto");
pRaviJITProtoT = llvm::PointerType::get(RaviJITProtoT, 0);

///*
//** Function Prototypes
//*/
// typedef struct Proto {
//  CommonHeader;
//  lu_byte numparams;  /* number of fixed parameters */
//  lu_byte is_vararg;
//  lu_byte maxstacksize;  /* maximum stack used by this function */
//  int sizeupvalues;  /* size of 'upvalues' */
//  int sizek;  /* size of 'k' */
//  int sizecode;
//  int sizelineinfo;
//  int sizep;  /* size of 'p' */
//  int sizelocvars;
//  int linedefined;
//  int lastlinedefined;
//  TValue *k;  /* constants used by the function */
//  Instruction *code;
//  struct Proto **p;  /* functions defined inside the function */
//  int *lineinfo;  /* map from opcodes to source lines (debug information) */
//  LocVar *locvars;  /* information about local variables (debug information)
//  */
//  Upvaldesc *upvalues;  /* upvalue information */
//  struct LClosure *cache;  /* last created closure with this prototype */
//  TString  *source;  /* used for debug information */
//  GCObject *gclist;
//  /* RAVI */
//  RaviJITProto *ravi_jit;
//} Proto;

ProtoT = llvm::StructType::create(context, "ravi.Proto");
pProtoT = llvm::PointerType::get(ProtoT, 0);
ppProtoT = llvm::PointerType::get(pProtoT, 0);
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);                           /* numparams */
elements.push_back(lu_byteT);                           /* is_vararg */
elements.push_back(lu_byteT);                           /* maxstacksize */
elements.push_back(C_intT);                             /* sizeupvalues */
elements.push_back(C_intT);                             /* sizek */
elements.push_back(C_intT);                             /* sizecode */
elements.push_back(C_intT);                             /* sizelineinfo */
elements.push_back(C_intT);                             /* sizep */
elements.push_back(C_intT);                             /* sizelocvars */
elements.push_back(C_intT);                             /* linedefined */
elements.push_back(C_intT);                             /* lastlinedefined */
elements.push_back(pTValueT);                           /* k */
elements.push_back(pInstructionT);                      /* code */
elements.push_back(ppProtoT);                           /* p */
elements.push_back(llvm::PointerType::get(C_intT, 0)); /* lineinfo */
```

```
elements.push_back(pLocVarT);                           /* locvars */
elements.push_back(pUpvaldescT);                        /* upvalues */
elements.push_back(pLClosureT);                         /* cache */
elements.push_back(pTStringT);                          /* source */
elements.push_back(pGCObjectT);                         /* gclist */
elements.push_back(pRaviJITProtoT);                     /* ravi_jit */
ProtoT->setBody(elements);

///*
//** Lua Upvalues
//*/
// typedef struct UpVal UpVal;
UpValT = llvm::StructType::create(context, "ravi.UpVal");
pUpValT = llvm::PointerType::get(UpValT, 0);

///*
//** Closures
//*/

//#define ClosureHeader \
//CommonHeader; lu_byte nupvalues; GCObject *gclist

// typedef struct CClosure {
//  ClosureHeader;
//  lua_CFunction f;
//  TValue upvalue[1];  /* list of upvalues */
//} CClosure;

CClosureT = llvm::StructType::create(context, "ravi.CClosure");
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);        /* nupvalues */
elements.push_back(pGCObjectT);      /* gclist */
elements.push_back(plua_CFunctionT); /* f */
elements.push_back(llvm::ArrayType::get(TValueT, 1));
CClosureT->setBody(elements);
pCClosureT = llvm::PointerType::get(CClosureT, 0);

// typedef struct LClosure {
//  ClosureHeader;
//  struct Proto *p;
//  UpVal *upvals[1];  /* list of upvalues */
//} LClosure;
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);   /* nupvalues */
elements.push_back(pGCObjectT); /* gclist */
elements.push_back(pProtoT);    /* p */
elements.push_back(llvm::ArrayType::get(pUpValT, 1));
LClosureT->setBody(elements);

///*
//** Tables
//*/
```

```
// typedef union TKey {
//  struct {
//    TValuefields;
//    int next;  /* for chaining (offset for next node) */
//  } nk;
//  TValue tvk;
//} TKey;
TKeyT = llvm::StructType::create(context, "ravi.TKey");
elements.clear();
elements.push_back(ValueT);
elements.push_back(C_intT);
elements.push_back(C_intT); /* next */
TKeyT->setBody(elements);
pTKeyT = llvm::PointerType::get(TKeyT, 0);

// typedef struct Node {
// TValue i_val;
// TKey i_key;
//} Node;
NodeT = llvm::StructType::create(context, "ravi.Node");
elements.clear();
elements.push_back(TValueT); /* i_val */
elements.push_back(TKeyT);   /* i_key */
NodeT->setBody(elements);
pNodeT = llvm::PointerType::get(NodeT, 0);

// typedef struct Table {
//  CommonHeader;
//  lu_byte flags;  /* 1<<p means tagmethod(p) is not present */
//  lu_byte lsizenode;  /* log2 of size of 'node' array */
//  unsigned int sizearray;  /* size of 'array' array */
//  TValue *array;  /* array part */
//  Node *node;
//  Node *lastfree;  /* any free position is before this position */
//  struct Table *metatable;
//  GCObject *gclist;
//  ravitype_t ravi_array_type; /* RAVI specialization */
//  unsigned int ravi_array_len; /* RAVI len specialization */
//} Table;
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);    /* flags */
elements.push_back(lu_byteT);    /* lsizenode */
elements.push_back(C_intT);      /* sizearray */
elements.push_back(pTValueT);    /* array part */
elements.push_back(pNodeT);      /* node */
elements.push_back(pNodeT);      /* lastfree */
elements.push_back(pTableT);     /* metatable */
elements.push_back(pGCObjectT);  /* gclist */
elements.push_back(ravitype_tT); /* ravi_array_type */
elements.push_back(C_intT);      /* ravi_array_len */
TableT->setBody(elements);

// struct lua_longjmp;  /* defined in ldo.c */
lua_longjumpT = llvm::StructType::create(context, "ravi.lua_longjmp");
plua_longjumpT = llvm::PointerType::get(lua_longjumpT, 0);
```

```
// lzio.h
// typedef struct Mbuffer {
//  char *buffer;
//  size_t n;
//  size_t buffsize;
//} Mbuffer;
MbufferT = llvm::StructType::create(context, "ravi.Mbuffer");
elements.clear();
elements.push_back(llvm::Type::getInt8PtrTy(context)); /* buffer */
elements.push_back(C_size_t);                          /* n */
elements.push_back(C_size_t);                          /* buffsize */
MbufferT->setBody(elements);

// typedef struct stringtable {
//  TString **hash;
//  int nuse;  /* number of elements */
//  int size;
//} stringtable;
stringtableT = llvm::StructType::create(context, "ravi.stringtable");
elements.clear();
elements.push_back(ppTStringT); /* hash */
elements.push_back(C_intT);     /* nuse */
elements.push_back(C_intT);     /* size */
stringtableT->setBody(elements);

///*
//** Information about a call.
//** When a thread yields, 'func' is adjusted to pretend that the
//** top function has only the yielded values in its stack; in that
//** case, the actual 'func' value is saved in field 'extra'.
//** When a function calls another with a continuation, 'extra' keeps
//** the function index so that, in case of errors, the continuation
//** function can be called with the correct top.
//*/
// typedef struct CallInfo {
//  StkId func;  /* function index in the stack */
//  StkId     top;  /* top for this function */
//  struct CallInfo *previous, *next;  /* dynamic call link */
//  union {
//    struct {  /* only for Lua functions */
//      StkId base;  /* base for this function */
//      const Instruction *savedpc;
//    } l;
//    struct {  /* only for C functions */
//      lua_KFunction k;  /* continuation in case of yields */
//      ptrdiff_t old_errfunc;
//      lua_KContext ctx;  /* context info. in case of yields */
//    } c;
//  } u;
//  ptrdiff_t extra;
//  short nresults;  /* expected number of results from this function */
//  lu_byte callstatus;
//} CallInfo;

elements.clear();
elements.push_back(StkIdT);        /* base */
elements.push_back(pInstructionT); /* savedpc */
elements.push_back(
```

```
    C_ptrdiff_t); /* dummy to make this same size as the other member */
CallInfo_lT = llvm::StructType::create(elements);

elements.clear();
elements.push_back(plua_KFunctionT); /* k */
elements.push_back(C_ptrdiff_t);     /* old_errfunc */
elements.push_back(lua_KContextT);   /* ctx */
CallInfo_cT = llvm::StructType::create(elements);

CallInfoT = llvm::StructType::create(context, "ravi.CallInfo");
pCallInfoT = llvm::PointerType::get(CallInfoT, 0);
elements.clear();
elements.push_back(StkIdT);     /* func */
elements.push_back(StkIdT);     /* top */
elements.push_back(pCallInfoT); /* previous */
elements.push_back(pCallInfoT); /* next */
elements.push_back(
    CallInfo_lT); /* u.l  - as we will typically access the lua call details
                     */
elements.push_back(C_ptrdiff_t);                     /* extra */
elements.push_back(llvm::Type::getInt16Ty(context)); /* nresults */
elements.push_back(lu_byteT);                        /* callstatus */
CallInfoT->setBody(elements);

// typedef struct ravi_State ravi_State;

ravi_StateT = llvm::StructType::create(context, "ravi.ravi_State");
pravi_StateT = llvm::PointerType::get(ravi_StateT, 0);

///*
//** 'global state', shared by all threads of this state
//*/
// typedef struct global_State {
//  lua_Alloc frealloc;  /* function to reallocate memory */
//  void *ud;         /* auxiliary data to 'frealloc' */
//  lu_mem totalbytes;  /* number of bytes currently allocated - GCdebt */
//  l_mem GCdebt;  /* bytes allocated not yet compensated by the collector */
//  lu_mem GCmemtrav;  /* memory traversed by the GC */
//  lu_mem GCestimate;  /* an estimate of the non-garbage memory in use */
//  stringtable strt;  /* hash table for strings */
//  TValue l_registry;
//  unsigned int seed;  /* randomized seed for hashes */
//  lu_byte currentwhite;
//  lu_byte gcstate;  /* state of garbage collector */
//  lu_byte gckind;  /* kind of GC running */
//  lu_byte gcrunning;  /* true if GC is running */
//  GCObject *allgc;  /* list of all collectable objects */
//  GCObject **sweepgc;  /* current position of sweep in list */
//  GCObject *finobj;  /* list of collectable objects with finalizers */
//  GCObject *gray;  /* list of gray objects */
//  GCObject *grayagain;  /* list of objects to be traversed atomically */
//  GCObject *weak;  /* list of tables with weak values */
//  GCObject *ephemeron;  /* list of ephemeron tables (weak keys) */
//  GCObject *allweak;  /* list of all-weak tables */
//  GCObject *tobefnz;  /* list of userdata to be GC */
//  GCObject *fixedgc;  /* list of objects not to be collected */
//  struct lua_State *twups;  /* list of threads with open upvalues */
//  Mbuffer buff;  /* temporary buffer for string concatenation */
```

```
//  unsigned int gcfinnum;  /* number of finalizers to call in each GC step */
//  int gcpause;  /* size of pause between successive GCs */
//  int gcstepmul;  /* GC 'granularity' */
//  lua_CFunction panic;  /* to be called in unprotected errors */
//  struct lua_State *mainthread;
//  const lua_Number *version;  /* pointer to version number */
//  TString *memerrmsg;  /* memory-error message */
//  TString *tmname[TM_N];  /* array with tag-method names */
//  struct Table *mt[LUA_NUMTAGS];  /* metatables for basic types */
//  /* RAVI */
//  ravi_State *ravi_state;
//} global_State;

global_StateT = llvm::StructType::create(context, "ravi.global_State");
pglobal_StateT = llvm::PointerType::get(global_StateT, 0);

///*
//** 'per thread' state
//*/
// struct lua_State {
//  CommonHeader;
//  lu_byte status;
//  StkId top;  /* first free slot in the stack */
//  global_State *l_G;
//  CallInfo *ci;  /* call info for current function */
//  const Instruction *oldpc;  /* last pc traced */
//  StkId stack_last;  /* last free slot in the stack */
//  StkId stack;  /* stack base */
//  UpVal *openupval;  /* list of open upvalues in this stack */
//  GCObject *gclist;
//  struct lua_State *twups;  /* list of threads with open upvalues */
//  struct lua_longjmp *errorJmp;  /* current error recover point */
//  CallInfo base_ci;  /* CallInfo for first level (C calling Lua) */
//  lua_Hook hook;
//  ptrdiff_t errfunc;  /* current error handling function (stack index) */
//  int stacksize;
//  int basehookcount;
//  int hookcount;
//  unsigned short nny;  /* number of non-yieldable calls in stack */
//  unsigned short nCcalls;  /* number of nested C calls */
//  lu_byte hookmask;
//  lu_byte allowhook;
//};
elements.clear();
elements.push_back(pGCObjectT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);
elements.push_back(lu_byteT);                       /* status */
elements.push_back(StkIdT);                         /* top */
elements.push_back(pglobal_StateT);                 /* l_G */
elements.push_back(pCallInfoT);                     /* ci */
elements.push_back(pInstructionT);                  /* oldpc */
elements.push_back(StkIdT);                         /* stack_last */
elements.push_back(StkIdT);                         /* stack */
elements.push_back(pUpValT);                        /* openupval */
elements.push_back(pGCObjectT);                     /* gclist */
elements.push_back(plua_StateT);                    /* twups */
elements.push_back(plua_longjumpT);                 /* errorJmp */
```

```
elements.push_back(CallInfoT);                          /* base_ci */
elements.push_back(plua_HookT);                         /* hook */
elements.push_back(C_ptrdiff_t);                        /* errfunc */
elements.push_back(C_intT);                             /* stacksize */
elements.push_back(C_intT);                             /* basehookcount */
elements.push_back(C_intT);                             /* hookcount */
elements.push_back(llvm::Type::getInt16Ty(context)); /* nny */
elements.push_back(llvm::Type::getInt16Ty(context)); /* nCcalls */
elements.push_back(lu_byteT);                           /* hookmask */
elements.push_back(lu_byteT);                           /* allowhook */
lua_StateT->setBody(elements);

// int luaD_poscall (lua_State *L, StkId firstResult)
elements.clear();
elements.push_back(plua_StateT);
elements.push_back(StkIdT);
luaD_poscallT = llvm::FunctionType::get(C_intT, elements, false);
```

CHAPTER 10

LLVM Type Based Alias Analysis

When a Lua opcode involves a call to a Lua function, the Lua stack may be reallocated. So then the base pointer which points to the function's base stack position must be refreshed.

To keep compilation simple I coded the compiler so that at the beginning of each opcode the base pointer is reloaded. My assumption was that the LLVM optimizer will realise that the base pointer hasn't changed and so the loads are redundant and can be removed. However to my surprise I found that this is not the case.

The main difference between the IR I was generating and that produced by Clang was that Clang generated IR appeared to be decorated by tbaa metadata. Example:

```
%base2 = getelementptr inbounds %struct.CallInfoLua* %0, i32 0, i32 4, i32 0
%1 = load %struct.TValue** %base2, align 4, !tbaa !12
```

Here the `!tbaa !12` refers to a tbaa metadata entry.

I won't show the Clang generated tbaa metadata here, but here is how I added similar support in Ravi. The required steps are:

1. Create tbaa metadata mappings for the types in the system.
2. Annotate Load and Store instructions with tbaa references.

10.1 Creating TBAA Metadata

Firstly you need an MDBuilder instance. So you need to include following headers:

```
#include "llvm/IR/MDBuilder.h"
#include "llvm/IR/Metadata.h"
```

We can create an MDBuilder instance like this:

```
llvm::MDBuilder mdbuilder(llvm::getGlobalContext());
```

The TBAA nodes hang off a root node. So we create that next:

```
llvm::MDNode *tbaa_root;
// Do what Clang does
tbaa_root = mdbuilder.createTBAARoot("Simple C / C++ TBAA");
```

Next we need to create some simple scalar types. We only need one type per size, so that means we don't need `long long` and `double` - either one will do. We create these scalar types as follows:

```
llvm::MDNode *tbaa_charT;
llvm::MDNode *tbaa_shortT;
llvm::MDNode *tbaa_intT;
llvm::MDNode *tbaa_longlongT;
llvm::MDNode *tbaa_pointerT;


//!4 = metadata !{metadata !"omnipotent char", metadata !5, i64 0}
tbaa_charT = mdbuilder.createTBAAScalarTypeNode("omnipotent char", tbaa_root, 0);
//!3 = metadata !{metadata !"any pointer", metadata !4, i64 0}
tbaa_pointerT = mdbuilder.createTBAAScalarTypeNode("any pointer", tbaa_charT, 0);
//!10 = metadata !{metadata !"short", metadata !4, i64 0}
tbaa_shortT = mdbuilder.createTBAAScalarTypeNode("short", tbaa_charT, 0);
//!11 = metadata !{metadata !"int", metadata !4, i64 0}
tbaa_intT = mdbuilder.createTBAAScalarTypeNode("int", tbaa_charT, 0);
//!9 = metadata !{metadata !"long long", metadata !4, i64 0}
tbaa_longlongT = mdbuilder.createTBAAScalarTypeNode("long long", tbaa_charT, 0);
```

The second argument to `createTBAAScalarTypeNode()` is the parent node. Note the hierarchy here:

```
+ root
|
+--+ char
   |
   +--+-- any pointer
      |
      +-- short
      |
      +-- int
      |
      +-- long long
```

This is how Clang has it defined.

Next we need to define aggregate (struct) types. The API we need for this is `createTBAAStructTypeNode()`. This method accepts a vector of `std::pair<llvm::MDNode *, uint64_t>` objects - each element in the vector defines a field in the struct. The integer parameter needs to be the offset of the field within the struct. Interestingly Clang generates offsets that indicate pointers are being treated as 32-bit quantities - even though I ran this on a 64-bit machine. So I guess that as long as we consistently use the size then this doesn't matter. The sizes used by Clang are:

- char - 1 byte
- short - 2 bytes
- int - 4 bytes
- pointer - 4 bytes
- long long - 8 bytes

Another interesting thing is that padding needs to be accounted for.

So now lets look at how to map following struct:

```
struct CallInfoL {     /* only for Lua functions */
  struct TValue *base; /* base for this function */
  const unsigned int *savedpc;
  ptrdiff_t dummy;
};
```

We map this as:

```
llvm::MDNode *tbaa_CallInfo_lT;

//!14 = metadata !{metadata !"CallInfoL", metadata !3, i64 0, metadata !3, i64 4, metadata !9, i64 8
std::vector<std::pair<llvm::MDNode *, uint64_t> > nodes;
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 0));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 4));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_longlongT, 8));
tbaa_CallInfo_lT = mdbuilder.createTBAAStructTypeNode("CallInfo_l", nodes);
```

To illustrate how a structure is referenced as a field in another lets also look at:

```
struct CallInfo {
  struct TValue *func;              /* function index in the stack */
  struct TValue *top;               /* top for this function */
  struct CallInfo *previous, *next; /* dynamic call link */
  struct CallInfoL l;
  ptrdiff_t extra;
  short nresults; /* expected number of results from this function */
  unsigned char callstatus;
};
```

We have a `CallInfoL` as the type of a field within the struct. Therefore:

```
llvm::MDNode *tbaa_CallInfoT;

//!13 = metadata !{metadata !"CallInfo",
//                 metadata !3, i64 0, metadata !3, i64 4, metadata !3, i64 8,
//                 metadata !3, i64 12, metadata !14, i64 16, metadata !9, i64 32,
//                 metadata !10, i64 40, metadata !4, i64 42}
nodes.clear();
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 0));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 4));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 8));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_pointerT, 12));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_CallInfo_lT, 16));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_longlongT, 32));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_shortT, 40));
nodes.push_back(std::pair<llvm::MDNode*, uint64_t>(tbaa_charT, 42));
tbaa_CallInfoT = mdbuilder.createTBAAStructTypeNode("CallInfo", nodes);
```

10.2 Decorating Load and Store instructions

So now we have created TBAA metadata for two struct types. Next we need to see how we use these in Load and Store instructions. Lets assume we need to load the pointer stored in `Callinfo.top`. In order to decorate the Load instruction with tbaa we need to create a Struct Tag Node - which is like a path node. Here it is:

```
llvm::MDNode *tbaa_CallInfo_topT;
tbaa_CallInfo_topT = mdbuilder.createTBAAStructTagNode(tbaa_CallInfoT, tbaa_pointerT, 4);
```

Above is saying that the field `top` in struct `CallInfo` is a pointer at offset 4.

Armed with this we can code:

```
llvm::Value *callinfo_top = /* GEP instruction */
llvm::Instruction *top = Builder.CreateLoad(callinfo_top);
top->setMetadata(llvm::LLVMContext::MD_tbaa, tbaa_CallInfo_topT);
```

10.3 Links

- TypeBasedAliasAnalysis code.
- IR documentation on tbaa metadata.
- Embedded metadata.

CHAPTER 11

LLVM Bindings for Lua/Ravi

As part of the Ravi Programming Language, it is my intention to provide a Lua 5.3 compatible LLVM binding. This will allow Lua programmers to write their own JIT compilers in Lua!

Right now this is in early development so there is no documentation. But the Lua programs here demonstrate the features available to date.

11.1 LLVM Modules and Execution Engines

One of the complexities of LLVM is the handling of modules and execution engines in a JIT environment. In Ravi I made the simple decision that each Lua function would get its own module and EE. This allows the function to be garbage collected as normal and release the associated module and EE. One of the things that is possible but not yet implemented is releasing the module and EE early; this requires implementing a custom memory manager (issue #48).

To mimic the Ravi model, the LLVM bindings provide a shortcut to setup an LLVM module and execution engine for a Lua C function. The following example illustrates:

```
-- Get the LLVM context - right now this is the
-- global context
local context = llvm.context()

-- Create a lua_CFunction instance
-- At this stage the function will get a module and
-- execution engine but no body
local mainfunc = context:lua_CFunction("demo")
```

Above creates an `llvm::Function` instance within a new module. An EE is automatically attached. You can get hold of the module as shown below:

```
-- Get hold of the module
local module = mainfunc:module()
```

Other native functions may be created within the same module as normal. However note that once the Lua function is compiled then no further updates to the module are possible.

The model I recommend when using this feature is to create one exported Lua C function in the module, with several private 'internal' supporting functions within the module.

11.2 Creating Modules and Execution Engines

The LLVM api for these functions are not exposed yet.

11.3 Examples

For examples that illustrate the bindings please visit the llvmbindings folder in the repository.

11.4 Type Hierarchy

The bindings provide a number of Lua types:

```
+ LLVMcontext
+ LLVMfunction
  + LLVMmainfunction
+ LLVMmodule
+ LLVMtype
  + LLVMstructtype
  + LLVMpointertype
  + LLVMfunctiontype
+ LLVMvalue
  + LLVMinstruction
  + LLVMconstant
  + LLVMphinode
+ LLVMirbuilder
+ LLVMbasicblock
```

11.5 Available Bindings

The following table lists the Lua LLVM api functions available.

Lua LLVM API

`llvm.context() -> LLVMcontext` Returns global llvm::Context

LLVMcontext methods

`lua_CFunction(name) -> LLVMmainfunction` Creates an llvm::Function within a new llvm::Module; and associates an llvm::ExecutionEngine with the module
`types() -> table of predefined type bindings` Returns a table of predefined LLVM type bindings
`structtype(name) -> LLVMstructtype` Opaque struct type; body can be added
`pointertype(type) -> LLVMpointertype` Given a type returns a pointertype
`functiontype(return_type, {argtypes}, {options}) -> LLVMfunctiontype` Creates a function type with specified return type, argument types. Takes the option 'vararg' which is false by default.
`basicblock(name) -> LLVMbasicblock` Create a basic block
`intconstant(intgervalue) -> LLVMvalue` Returns an integer constant value
`nullconstant(pointertype) -> LLVMvalue` Returns a NULL constant of specified pointertype

LLVMstructtype methods

`setbody({types})` Adds members to the struct type

LLVMmainfunction methods

`appendblock(LLVMbasicblock)` Adds a basic block to the end
`compile()` Compiles the module and returns a reference to the C Closure
`arg(position) -> LLVMvalue` Returns the argument at position; position >= 1; returns `nil` if argument not available
`module() -> LLVMmodule` Returns the module associated with the function
`extern(name[, functiontype]) -> LLVMconstant` Returns an extern declaration; A number of Lua Api functions are predefined.

LLVMmodule methods

`newfunction(name, functiontype) -> LLVMfunction` Returns an internal linkage function within the module
`dump()` Dumps the module

LLVMfunction methods

`appendblock(LLVMbasicblock)` Adds a basic block to the end
`arg(position) -> LLVMvalue` Returns the argument at position; position >= 1; returns `nil` if argument not available
`alloca(type[, name [,arraysize]]) -> LLVMinstruction` Creates a variable in the first block of the function

LLVMirbuilder methods

`setinsertpoint(basicblock)` Set current basicblock
`ret([value])` Emit return instruction
`stringconstant(string) -> LLVMvalue` Create a global string constant
`call({args}, {options}) -> LLVMinstruction` Emit call instruction; 'tailcall' option is false by default
`br(basicblock) -> LLVMinstruction` Emit a branch instruction
`condbr(value, true_block, false_block) -> LLVMinstruction` Emit a conditional branch
`phi(type, num_values[, name]) -> LLVMphinode` Generate a PHINode
GEP Operators
`gep(value, {offsets}) -> LLVMvalue` getelementptr to obtain ptr to an array or struct element
`inboundsgep(value, {offsets}) -> LLVMvalue` inbounds version of getelementptr
Memory Operators
`load(ptr) -> LLVMinstruction` Loads the value at ptr

CHAPTER 12

Ravi Performance Benchmarks

Ravi's reason for existence is to achieve greater performance than standard Lua 5.3. Hence performance benchmarks are of interest.

The programs used in the performance testing can be found at Ravi Tests folder.

Program	Lua5.3.2	Ravi Int	Ravi(LLVM)	LuaJIT2.1 Int	LuaJIT2.1
fornum_test1.lua	8.94	8.587	0.309	3.516	0.312
fornum_test2.lua	9.195	9.243	4.446	3.75	0.922
fornum_test3.lua	52.494	48.223	4.748	16.74	7.75
mandel1.lua(4000)	20.324	19.835	8.056	8.469	1.594
mandel1.ravi(4000)	n/a	16.192	1.571	n/a	n/a
fannkuchen.lua(11)	46.203	48.654	28.422	20.6	4.672
fannkuchen.ravi(11)	n/a	34.411	4.634	n/a	n/a
matmul1.lua(1000)	26.672	26.51	16.83	12.594	1.078
matmul1_ravi.lua(1000)	n/a	20.123	1.137	n/a	n/a
matmul1.ravi(1000)	n/a	25.387	1.039	n/a	n/a

Following points are worth bearing in mind when looking at above benchmarks.

1. For Ravi the timings above do not include the LLVM compilation time.
2. The benchmarks were run on Windows 10 64-bit. LLVM version 3.9 was used. Ravi and Lua 5.3.2 were compiled using Visual C++ 2015.
3. Some of the Ravi benchmarks are based on code that uses optional static types; additionally for the *matmul* benchmark a setting was used to disable array bounds checks for array read operations.
4. Above benchmarks are primarily numerical. In real life scenarios there are other factors that affect performance. For instance, via FFI LuaJIT is able to make efficient calls to external C functions, but Ravi does not have a similar FFI interface. LuaJIT can also inline Lua function calls but Ravi does not have this ability and hence function calls go via the Lua infrastructure and are therefore expensive. Ravi's code generation is best when types are annotated as otherwise the dynamic type checks degrade performance as above benchmarks show. Finally LLVM is a slow compiler relative to LuaJIT's JIT compiler which is extremely fast.
5. Performance of Lua 5.3.2 is better than 5.3.0 or 5.3.1, thanks to the table optimizations in this version.

In general to obtain the best performanc with Ravi, following steps are necessary.

1. Annotate types as much as possible.
2. Use fornum loops with integer counters.
3. Avoid function calls inside loop bodies.
4. Do not assume that JIT compilation is beneficial - benchmark code with and without JIT compilation.

5. Try to compile a set of functions (in a table) preferably at program startup. This way you pay for the JIT compilation cost only once.
6. Dump the generated Lua bytecode to see if specialised Ravi bytecodes are being generated or not. If not you may be missing type annotations.
7. Avoid using globals.
8. Note that only functions executing in the main Lua thread are run in JIT mode. Coroutines in particular are always interpreted.
9. Also note that tail calls are expensive in JIT mode as they are treated as normal function calls; so it is better to avoid JIT compilation of code that relies upon tail calls.

CHAPTER 13

Ravi JIT Compilation Status

13.1 Introduction

Ravi uses LLVM for JIT compilation.

13.2 Benefits of using LLVM

- LLVM has a well documented intermediate representation called LLVM IR.
- The LLVM `IRBuilder` implements type checks so that when LLVM code is being generated, basic type errors are caught by the builder.
- LLVM provides a verifier to check that the generated IR is valid. This allows the IR to be validated prior to machine code generation.
- All of the LLVM optimization passes can be used.
- The Clang compiler supports generating LLVM IR so that if you want to know what the LLVM IR should look like for a particular piece of code, you can write a small C snippet and have Clang generate the IR for you.
- There is great momentum behind LLVM.
- The LLVM license is not based on GPL, so it is not viral.
- LLVM is much better documented than other products that aim to cover similar ground.
- LLVM's API is well designed and has a layered architecture.

13.3 Drawbacks of LLVM

- LLVM is huge in size. Lua on its own is tiny - but when linked to LLVM the resulting binary is a monster.
- There is a cost to compiling in LLVM so the benefit of compilation accrues only when a Lua function will be used again and again.
- LLVM cannot be linked as a shared library on Windows and a shared library configuration is not recommended on other platforms as well.
- LLVM's API keeps changing so that with every release of LLVM one has to revise the way it is used.

13.4 The Architecture of Ravi's JIT Compilation

- The unit of compilation is a Lua function
- Each Lua function is compiled to a Module/Function in LLVM parlance
- The compiled code is attached to the Lua function prototype
- The compiled code is garbage collected as normal by Lua
- The Lua runtime coordinates function calls - so anytime a Lua function is called it goes via the Lua infrastructure.
- The decision to call a JIT compiled version is made in the Lua Infrastructure (specifically in `luaD_precall()` function in `ldo.c`)
- The JIT compiler translates Lua/Ravi bytecode to LLVM IR - i.e. it does not translate Lua source code.
- There is no inlining of Lua functions.
- Generally the JIT compiler implements the same instructions as in `lvm.c` - however for some bytecodes the code calls a C function rather than generating inline IR. These opcodes are OP_LOADNIL, OP_NEWTABLE, OP_RAVI_NEWARRAYINT, OP_RAVI_NEWARRAYFLT, OP_SETLIST, OP_CONCAT, OP_CLOSURE, OP_VARARG, OP_RAVI_SHL_II, OP_RAVI_SHR_II.
- Ravi represents Lua values as done by Lua 5.3 - i.e. in a 16 byte structure.
- Ravi compiler generates type specifc opcodes which result in simpler and higher performance LLVM IR.

13.5 Limitations of JIT compilation

- Coroutines are not supported - JITed functions cannot yield
- The Debug API relies upon a field called `savedpc` which tracks the current instruction being executed by Lua interpreter. As this is not updated by the JIT code the Debug API can only provide a subset of normal functionality. The Debug API is not yet fully tested.
- The Lua VM supports infinite tail recursion. The JIT compiler treats OP_TAILCALL as normal OP_CALL so that recursion is limited to about 110 levels.
- The Lua C API has not yet been tested against the Ravi extensions - especially static typing and array types. Do not use the C API for now - as you could break the type system of Ravi.
- Bit-wise operators are JIT compiled only when the variables are known to be integers (specialized byte codes are used).

13.6 JIT Status of Lua/Ravi Bytecodes

The JIT compilation status of the Lua and Ravi bytecodes are given below.

This information was last updated on 25th July 2015. As new bytecodes are being added to the JIT compiler on a regular basis the status information below may be slightly out of date.

Note that if a Lua functions contains a bytecode that cannot be be JITed then the function cannot be JITed.

name	JITed?	description
OP_MOVE	YES	R(A) := R(B)
OP_LOADK	YES	R(A) := Kst(Bx)

Table 13.1 – continued from previous page

name	JITed?	description
OP_LOADKX	YES	R(A) := Kst(extra arg)
OP_LOADBOOL	YES	R(A) := (Bool)B; if (C) pc++
OP_LOADNIL	YES (1)	R(A), R(A+1), ..., R(A+B) := nil
OP_GETUPVAL	YES	R(A) := UpValue[B]
OP_GETTABUP	YES	R(A) := UpValue[B][RK(C)]
OP_GETTABLE	YES	R(A) := R(B)[RK(C)]
OP_SETTABUP	YES	UpValue[A][RK(B)] := RK(C)
OP_SETUPVAL	YES	UpValue[B] := R(A)
OP_SETTABLE	YES	R(A)[RK(B)] := RK(C)
OP_NEWTABLE	YES (1)	R(A) := {} (size = B,C)
OP_SELF	YES (1)	R(A+1) := R(B); R(A) := R(B)[RK(C)]
OP_ADD	YES	R(A) := RK(B) + RK(C)
OP_SUB	YES	R(A) := RK(B) - RK(C)
OP_MUL	YES	R(A) := RK(B) * RK(C)
OP_MOD	YES	R(A) := RK(B) % RK(C)
OP_POW	YES	R(A) := RK(B) ^ RK(C)
OP_DIV	YES	R(A) := RK(B) / RK(C)
OP_IDIV	YES	R(A) := RK(B) // RK(C)
OP_BAND	YES (1)	R(A) := RK(B) & RK(C)
OP_BOR	YES (1)	R(A) := RK(B) | RK(C)
OP_BXOR	YES (1)	R(A) := RK(B) ~ RK(C)
OP_SHL	YES (1)	R(A) := RK(B) << RK(C)
OP_SHR	YES (1)	R(A) := RK(B) >> RK(C)
OP_UNM	YES	R(A) := -R(B)
OP_BNOT	YES (1)	R(A) := ~R(B)
OP_NOT	YES	R(A) := not R(B)
OP_LEN	YES (1)	R(A) := length of R(B)
OP_CONCAT	YES (1)	R(A) := R(B).. ... ..R(C)
OP_JMP	YES	c+=sBx; if (A) close all upvalues >= R(A - 1)
OP_EQ	YES (1)	if ((RK(B) == RK(C)) ~= A) then pc++
OP_LT	YES (1)	if ((RK(B) < RK(C)) ~= A) then pc++
OP_LE	YES (1)	if ((RK(B) <= RK(C)) ~= A) then pc++
OP_TEST	YES	if not (R(A) <=> C) then pc++
OP_TESTSET	YES	if (R(B) <=> C) then R(A) := R(B) else pc++
OP_CALL	YES	R(A), .. ,R(A+C-2) := R(A)(R(A+1), .. ,R(A+B-1))
OP_TAILCALL	YES (2)	return R(A)(R(A+1), ... ,R(A+B-1)) Compiled as OP_CALL so no tail call optimization
OP_RETURN	YES	return R(A), ... ,R(A+B-2) (see note)
OP_FORLOOP	YES	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) }
OP_FORPREP	YES	R(A)-=R(A+2); pc+=sBx
OP_TFORCALL	YES	R(A+3), ... ,R(A+2+C) := R(A)(R(A+1), R(A+2));
OP_TFORLOOP	YES	if R(A+1) ~= nil then { R(A)=R(A+1); pc += sBx }
OP_SETLIST	YES (1)	R(A)[(C-1)*FPF+i] := R(A+i), 1 <= i <= B
OP_CLOSURE	YES (1)	R(A) := closure(KPROTO[Bx])
OP_VARARG	YES (1)	R(A), R(A+1), ..., R(A+B-2) = vararg
OP_EXTRAARG	N/A	extra (larger) argument for previous opcode
OP_RAVI_NEWARRAYI	YES	R(A) := array of int
OP_RAVI_NEWARRAYF	YES	R(A) := array of float
OP_RAVI_LOADIZ	YES	R(A) := tointeger(0)
OP_RAVI_LOADFZ	YES	R(A) := tonumber(0)

Table 13.1 – continued from previous page

name	JITed?	description
OP_RAVI_ADDFF	YES	R(A) := RK(B) + RK(C)
OP_RAVI_ADDFI	YES	R(A) := RK(B) + RK(C)
OP_RAVI_ADDII	YES	R(A) := RK(B) + RK(C)
OP_RAVI_SUBFF	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBFI	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBIF	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBII	YES	R(A) := RK(B) - RK(C)
OP_RAVI_MULFF	YES	R(A) := RK(B) * RK(C)
OP_RAVI_MULFI	YES	R(A) := RK(B) * RK(C)
OP_RAVI_MULII	YES	R(A) := RK(B) * RK(C)
OP_RAVI_DIVFF	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVFI	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVIF	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVII	YES	R(A) := RK(B) / RK(C)
OP_RAVI_TOINT	YES	R(A) := toint(R(A))
OP_RAVI_TOFLT	YES	R(A) := tofloat(R(A))
OP_RAVI_TOARRAYI	YES	R(A) := to_arrayi(R(A))
OP_RAVI_TOARRAYF	YES	R(A) := to_arrayf(R(A))
OP_RAVI_MOVEI	YES	R(A) := R(B), check R(B) is integer
OP_RAVI_MOVEF	YES	R(A) := R(B), check R(B) is number
OP_RAVI_MOVEAI	YES	R(A) := R(B), check R(B) is array of integer
OP_RAVI_MOVEAF	YES	R(A) := R(B), check R(B) is array of numbers
OP_RAVI_GETTABLE_AI	YES	R(A) := R(B)[RK(C)] where R(B) is array of integers and RK(C) is integer
OP_RAVI_GETTABLE_AF	YES	R(A) := R(B)[RK(C)] where R(B) is array of numbers and RK(C) is integer
OP_RAVI_SETTABLE_AI	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of integers, and RK(C) is
OP_RAVI_SETTABLE_AF	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of numbers, and RK(C) i
OP_RAVI_FORLOOP_IP	YES	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) } Specialization for
OP_RAVI_FORPREP_IP	YES	R(A)-=R(A+2); pc+=sBx Specialization for integer step > 1
OP_RAVI_FORLOOP_I1	YES	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) } Specialization for
OP_RAVI_FORPREP_I1	YES	R(A)-=R(A+2); pc+=sBx Specialization for integer step == 1
OP_RAVI_SETUPVALI	YES (1)	UpValue[B] := tointeger(R(A))
OP_RAVI_SETUPVALF	YES (1)	UpValue[B] := tonumber(R(A))
OP_RAVI_SETUPVALAI	YES (1)	UpValue[B] := toarrayint(R(A))
OP_RAVI_SETUPVALAF	YES (1)	UpValue[B] := toarrayflt(R(A))
OP_RAVI_SETTABLE_AII	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of integers, and RK(C) is
OP_RAVI_SETTABLE_AFF	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of numbers, and RK(C) i
OP_RAVI_BAND_II	YES	R(A) := RK(B) & RK(C), operands are int
OP_RAVI_BOR_II	YES	R(A) := RK(B) | RK(C), operands are int
OP_RAVI_BXOR_II	YES	R(A) := RK(B) ~ RK(C), operands are int
OP_RAVI_SHL_II	YES (5)	R(A) := RK(B) << RK(C), operands are int
OP_RAVI_SHR_II	YES (5)	R(A) := RK(B) >> RK(C), operands are int
OP_RAVI_BNOT_I	YES	R(A) := ~R(B), int operand
OP_RAVI_EQ_II	YES	if ((RK(B) == RK(C)) ~= A) then pc++
OP_RAVI_EQ_FF	YES	if ((RK(B) == RK(C)) ~= A) then pc++
OP_RAVI_LT_II	YES	if ((RK(B) < RK(C)) ~= A) then pc++
OP_RAVI_LT_FF	YES	if ((RK(B) < RK(C)) ~= A) then pc++
OP_RAVI_LE_II	YES	if ((RK(B) <= RK(C)) ~= A) then pc++
OP_RAVI_LE_FF	YES	if ((RK(B) <= RK(C)) ~= A) then pc++
OP_RAVI_GETTABLE_I	YES	R(A) := R(B)[RK(C)], integer key

Table 13.1 – continued from previous page

name	JITed?	description
OP_RAVI_GETTABLE_S	YES	R(A) := R(B)[RK(C)], string key
OP_RAVI_GETTABLE_SK	YES	R(A) := R(B)[RK(C)], string key
OP_RAVI_SETTABLE_I	YES (4)	R(A)[RK(B)] := RK(C), integer key
OP_RAVI_SETTABLE_S	YES (3)	R(A)[RK(B)] := RK(C), string key
OP_RAVI_SETTABLE_SK	YES	R(A)[RK(B)] := RK(C), string key
OP_RAVI_TOTAB	YES	R(A) := to_table(R(A))
OP_RAVI_MOVETAB	YES	R(A) := R(B), check R(B) is a table
OP_RAVI_SETUPVALT	YES (1)	UpValue[B] := to_table(R(A))
OP_RAVI_SELF_SK	YES	R(A+1) := R(B); R(A) := R(B)[RK(C)]
OP_RAVI_SELF_S	YES	R(A+1) := R(B); R(A) := R(B)[RK(C)]
OP_RAVI_GETTABUP_SK	YES	R(A) := UpValue[B][RK(C)]

1. These bytecoes are handled via function calls rather than inline code generation
2. Tail calls are the same as ordinary calls.
3. The _SK variant is generated
4. Generates generic SETTABLE
5. Inline code is generated only when operand is a constant integer

13.7 Ravi's LLVM JIT compiler source

The LLVM JIT implementation is in following sources:

- ravillvm.h - includes LLVM headers and defines the generic JIT State and Function interfaces
- ravijit.h - defines the JIT API
- ravi_llvmcodegen.h - defines the types used by the code generator
- ravijit.cpp - Non implementation specific JIT API functions
- ravi_llvmjit.cpp - basic LLVM infrastructure and Ravi API definition
- ravi_llvmtypes.cpp - contains LLVM type definitions for Lua objects
- ravi_llvmcodegen.cpp - LLVM JIT compiler - main driver for compiling Lua bytecodes into LLVM IR
- ravi_llvmload.cpp - implements OP_LOADK and OP_MOVE, and related operations, also OP_LOADBOOL
- ravi_llvmcomp.cpp - implements OP_EQ, OP_LT, OP_LE, OP_TEST and OP_TESTSET.
- ravi_llvmreturn.cpp - implements OP_RETURN
- ravi_llvmforprep.cpp - implements OP_FORPREP
- ravi_llvmforloop.cpp - implements OP_FORLOOP
- ravi_llvmtforcall.cpp - implements OP_TFORCALL and OP_TFORLOOP
- ravi_llvmarith1.cpp - implements various type specialized arithmetic operations - these are Ravi extensions
- ravi_llvmarith2.cpp - implements Lua opcodes such as OP_ADD, OP_SUB, OP_MUL, OP_DIV, OP_POW, OP_IDIV, OP_MOD, OP_UNM
- ravi_llvmcall.cpp - implements OP_CALL, OP_JMP

- ravi_llvmtable.cpp - implements OP_GETTABLE, OP_SETTABLE and various other table operations, OP_SELF, and also upvalue operations
- ravi_llvmrest.cpp - OP_CLOSURE, OP_VARARG, OP_CONCAT

CHAPTER 14

JIT Compilation for Ravi using `libgccjit`

14.1 Introduction

The latest gcc 5.2 release contains a new component called `libgccjit`. This basically exposes an API via a shared library to the compilation functions within gcc.

I am keen to provide support for this in Ravi. From initial look it seems to contain all the features I need to implement a JIT compiler for Ravi. Obviously having implemented the LLVM version it is going to be a little easier as I can mostly do a port of the LLVM version.

14.2 License

Ravi itself is licensed under MIT license (including the code that implements the JIT compiler) - however I think that when linked to `libgccjit` the effective license will be GPLv3.

14.3 Why another JIT engine?

Well partly as I feel I have a moral obligation to support gcc, given it has been instrumental in bringing about the OpenSource / Free Software ecosystem.

Secondly I am always looking for alternatives that will let me reduce the footprint of Ravi. The `libgccjit` is offered as a shared library - this is a great thing. I hate to have to statically link LLVM.

LLVM implementation and `libgccjit` implementation will both be kept in sync so that user can choose either option. Right now the LLVM implementation is more advanced and new features are implemented there first and then ported to the `libgccjit` implementation.

14.4 Building GCC

I am running Ubuntu 14.04 LTS on VMWare virtual machine.

I built gcc 5.2 from source as follows.

1. Extracted gcc-5.2 source to `~/gcc-5.2.0`.
2. Created a build folder `~/buildgcc`.
3. Installed various pre-requisites for gcc.

4. Then ran following from inside the build folder:

```
../gcc-5.2.0/configure --prefix=~/local --enable-host-shared --enable-languages=jit,c++ --disabl
```

5. Next performed the build as follows:

```
make
make install
```

14.5 On Mac OSX Yosemite

It appears that the HomeBrew project supports creating the `libgccjit 5.2` library. However the default formula doesn't quite work and needs to be patched for libgccjit to work properly. A patched formula can be found at here. To use the patched version edit the gcc formula and copy the patched version. After that following should build and install gcc 5.2 including the JIT library:

```
brew install gcc --with-jit --without-multilib
```

14.6 Current Status

Many bytecodes are now compiled - see below for detailed status. The current version of Ravi passes the Lua test cases using `libgccjit`.

14.7 Building Ravi with `libgccjit` on Linux

Warning: Note that right now the Ravi's `libgccjit` based JIT implementation is work in progress - please expect bugs.

You can build Ravi with `libgccjit` linked in as follows:

```
mkdir build
cd build
export LD_LIBRARY_PATH=~/local/lib64:$LD_LIBRARY_PATH
cmake -DCMAKE_BUILD_TYPE=Debug -DCMAKE_C_COMPILER=~/local/bin/gcc -DCMAKE_CXX_COMPILER=~/local/bin/g
make
```

Above assumes that gccjit is installed under `~/local` as described in Building GCC section above.

A helloworld test program is built. To run it though you need to ensure that your `PATH` and `LD_LIBRARY_PATH` variables include `~/local/bin` and `~/local/lib` respectively.

14.8 Initial Observations

In terms of packaging `libgccjit` consists of a C header file, a C++ header file and one shared library. That is pretty neat as it simplifies the usage.

Setting up of the Lua types is proving easier in `libgccjit` due to the fact that Lua uses unions extensively and `libgccjit` supports defining union types. This means that most of the Lua types can be translated more naturally. LLVM on the other hand does not support unions so I had to carefully define structs that would match the size of the union, and in the JIT compilation use casts where needed.

14.9 JIT Status of Lua/Ravi Bytecodes

Following is the status as of 4 July 2015.

name	JITed?	description
OP_MOVE	YES	R(A) := R(B)
OP_LOADK	YES	R(A) := Kst(Bx)
OP_LOADKX	NO	R(A) := Kst(extra arg)
OP_LOADBOOL	YES	R(A) := (Bool)B; if (C) pc++
OP_LOADNIL	YES	R(A), R(A+1), ..., R(A+B) := nil
OP_GETUPVAL	YES	R(A) := UpValue[B]
OP_GETTABUP	YES	R(A) := UpValue[B][RK(C)]
OP_GETTABLE	YES	R(A) := R(B)[RK(C)]
OP_SETTABUP	YES	UpValue[A][RK(B)] := RK(C)
OP_SETUPVAL	YES	UpValue[B] := R(A)
OP_SETTABLE	YES	R(A)[RK(B)] := RK(C)
OP_NEWTABLE	YES	R(A) := { } (size = B,C)
OP_SELF	YES	R(A+1) := R(B); R(A) := R(B)[RK(C)]
OP_ADD	YES	R(A) := RK(B) + RK(C)
OP_SUB	YES	R(A) := RK(B) - RK(C)
OP_MUL	YES	R(A) := RK(B) * RK(C)
OP_MOD	NO	R(A) := RK(B) % RK(C)
OP_POW	NO	R(A) := RK(B) ^ RK(C)
OP_DIV	YES	R(A) := RK(B) / RK(C)
OP_IDIV	NO	R(A) := RK(B) // RK(C)
OP_BAND	NO	R(A) := RK(B) & RK(C)
OP_BOR	NO	R(A) := RK(B) | RK(C)
OP_BXOR	NO	R(A) := RK(B) ~ RK(C)
OP_SHL	NO	R(A) := RK(B) << RK(C)
OP_SHR	NO	R(A) := RK(B) >> RK(C)
OP_UNM	NO	R(A) := -R(B)
OP_BNOT	NO	R(A) := ~R(B)
OP_NOT	YES	R(A) := not R(B)
OP_LEN	YES	R(A) := length of R(B)
OP_CONCAT	YES	R(A) := R(B).. ... ..R(C)
OP_JMP	YES	c+=sBx; if (A) close all upvalues >= R(A - 1)
OP_EQ	YES	if ((RK(B) == RK(C)) ~= A) then pc++
OP_LT	YES	if ((RK(B) < RK(C)) ~= A) then pc++
OP_LE	YES	if ((RK(B) <= RK(C)) ~= A) then pc++
OP_TEST	YES	if not (R(A) <=> C) then pc++
OP_TESTSET	YES	if (R(B) <=> C) then R(A) := R(B) else pc++
OP_CALL	YES	R(A), .. ,R(A+C-2) := R(A)(R(A+1), .. ,R(A+B-1))
OP_TAILCALL	YES	return R(A)(R(A+1), ... ,R(A+B-1)) Compiled as OP_CALL so no tail call optimization
OP_RETURN	YES	return R(A), ... ,R(A+B-2) (see note)
OP_FORLOOP	NO	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) }
OP_FORPREP	NO	R(A)-=R(A+2); pc+=sBx
OP_TFORCALL	YES	R(A+3), ... ,R(A+2+C) := R(A)(R(A+1), R(A+2));
OP_TFORLOOP	YES	if R(A+1) ~= nil then { R(A)=R(A+1); pc += sBx }
OP_SETLIST	YES	R(A)[(C-1)*FPF+i] := R(A+i), 1 <= i <= B
OP_CLOSURE	YES	R(A) := closure(KPROTO[Bx])
OP_VARARG	YES	R(A), R(A+1), ..., R(A+B-2) = vararg

Table 14.1 – continued from previous page

name	JITed?	description
OP_EXTRAARG	N/A	extra (larger) argument for previous opcode
OP_RAVI_NEWARRAYI	YES	R(A) := array of int
OP_RAVI_NEWARRAYF	YES	R(A) := array of float
OP_RAVI_LOADIZ	YES	R(A) := tointeger(0)
OP_RAVI_LOADFZ	YES	R(A) := tonumber(0)
OP_RAVI_ADDFF	YES	R(A) := RK(B) + RK(C)
OP_RAVI_ADDFI	YES	R(A) := RK(B) + RK(C)
OP_RAVI_ADDII	YES	R(A) := RK(B) + RK(C)
OP_RAVI_SUBFF	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBFI	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBIF	YES	R(A) := RK(B) - RK(C)
OP_RAVI_SUBII	YES	R(A) := RK(B) - RK(C)
OP_RAVI_MULFF	YES	R(A) := RK(B) * RK(C)
OP_RAVI_MULFI	YES	R(A) := RK(B) * RK(C)
OP_RAVI_MULII	YES	R(A) := RK(B) * RK(C)
OP_RAVI_DIVFF	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVFI	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVIF	YES	R(A) := RK(B) / RK(C)
OP_RAVI_DIVII	YES	R(A) := RK(B) / RK(C)
OP_RAVI_TOINT	YES	R(A) := toint(R(A))
OP_RAVI_TOFLT	YES	R(A) := tofloat(R(A))
OP_RAVI_TOARRAYI	YES	R(A) := to_arrayi(R(A))
OP_RAVI_TOARRAYF	YES	R(A) := to_arrayf(R(A))
OP_RAVI_MOVEI	YES	R(A) := R(B), check R(B) is integer
OP_RAVI_MOVEF	YES	R(A) := R(B), check R(B) is number
OP_RAVI_MOVEAI	YES	R(A) := R(B), check R(B) is array of integer
OP_RAVI_MOVEAF	YES	R(A) := R(B), check R(B) is array of numbers
OP_RAVI_GETTABLE_AI	YES	R(A) := R(B)[RK(C)] where R(B) is array of integers and RK(C) is integer
OP_RAVI_GETTABLE_AF	YES	R(A) := R(B)[RK(C)] where R(B) is array of numbers and RK(C) is integer
OP_RAVI_SETTABLE_AI	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of integers, and RK(C) is
OP_RAVI_SETTABLE_AF	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of numbers, and RK(C) i
OP_RAVI_FORLOOP_IP	YES	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) } Specialization for
OP_RAVI_FORPREP_IP	YES	R(A)-=R(A+2); pc+=sBx Specialization for integer step > 1
OP_RAVI_FORLOOP_I1	YES	R(A)+=R(A+2); if R(A) <?= R(A+1) then { pc+=sBx; R(A+3)=R(A) } Specialization for
OP_RAVI_FORPREP_I1	YES	R(A)-=R(A+2); pc+=sBx Specialization for integer step == 1
OP_RAVI_SETUPVALI	NO	UpValue[B] := tointeger(R(A))
OP_RAVI_SETUPVALF	NO	UpValue[B] := tonumber(R(A))
OP_RAVI_SETUPVALAI	NO	UpValue[B] := toarrayint(R(A))
OP_RAVI_SETUPVALAF	NO	UpValue[B] := toarrayflt(R(A))
OP_RAVI_SETTABLE_AII	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of integers, and RK(C) is
OP_RAVI_SETTABLE_AFF	YES	R(A)[RK(B)] := RK(C) where RK(B) is an integer R(A) is array of numbers, and RK(C) i
OP_RAVI_BAND_II	NO	R(A) := RK(B) & RK(C), operands are int
OP_RAVI_BOR_II	NO	R(A) := RK(B) | RK(C), operands are int
OP_RAVI_BXOR_II	NO	R(A) := RK(B) ~ RK(C), operands are int
OP_RAVI_SHL_II	NO	R(A) := RK(B) << RK(C), operands are int
OP_RAVI_SHR_II	NO	R(A) := RK(B) >> RK(C), operands are int
OP_RAVI_BNOT_I	NO	R(A) := ~R(B), int operand
OP_RAVI_EQ_II	YES	if ((RK(B) == RK(C)) ~= A) then pc++
OP_RAVI_EQ_FF	YES	if ((RK(B) == RK(C)) ~= A) then pc++

Table 14.1 – continued from previous page

name	JITed?	description
OP_RAVI_LT_II	YES	if ((RK(B) < RK(C)) ~= A) then pc++
OP_RAVI_LT_FF	YES	if ((RK(B) < RK(C)) ~= A) then pc++
OP_RAVI_LE_II	YES	if ((RK(B) <= RK(C)) ~= A) then pc++
OP_RAVI_LE_FF	YES	if ((RK(B) <= RK(C)) ~= A) then pc++

14.10 Ravi's libgccjit JIT compiler source

The libgccjit JIT implementation is in following sources:

- ravijit.h - defines the JIT API
- ravi_gccjit.h - defines the types used by the code generator, and declares prototypes
- ravijit.cpp - basic JIT infrastructure and Ravi API definition
- ravi_gcctypes.c - contains JIT type definitions for Lua objects
- ravi_gcccodegen.c - JIT compiler - main driver for compiling Lua bytecodes
- ravi_gccload.c - implements OP_LOADK and OP_MOVE, and related operations, also OP_LOADBOOL
- ravi_gcccomp.c - implements OP_EQ, OP_LT, OP_LE, OP_TEST and OP_TESTSET.
- ravi_gccreturn.c - implements OP_RETURN
- ravi_gccforprep.c - implements OP_RAVI_FORPREP_I1 and OP_RAVI_FORPREP_IP
- ravi_gccforloop.c - implements OP_RAVI_FORLOOP_I1 and OP_RAVI_FORLOOP_IP
- ravi_gcctforcall.c - implements OP_TFORCALL and OP_TFORLOOP
- ravi_gccarith1.c - implements various type specialized arithmetic operations - these are Ravi extensions
- ravi_gccarith2.c - implements Lua opcodes such as OP_ADD, OP_SUB, OP_MUL, OP_DIV, OP_UNM
- ravi_gcccall.c - implements OP_CALL, OP_JMP
- ravi_gcctable.c - implements OP_GETTABLE, OP_SETTABLE and various other table operations, OP_SELF, and also upvalue operations
- ravi_gccrest.c - OP_CLOSURE, OP_VARARG, OP_CONCAT

CHAPTER 15

Indices and tables

- genindex
- modindex
- search