Lab 4
Syntax-Directed Code Generation

Objective

During the previous lab, you have written your own interpreter of the MiniC language. In this lab the objective is to generate *valid* RISCV codes from MiniC programs:

- Generate 3-address code for the MiniC language.
- Generate executable "dummy" RISCV from programs in MiniC via two simple allocation algorithms.
- **Please follow instructions and COMMENT YOUR CODE!**

Student files are in the Git repository.

You may have to install some additional Python libs:

```
python3 -m pip install --user networkx graphviz
```

And on your personal machines:

```
sudo apt-get install graphviz-dev
```

Make sure your Git repository is up-to-date, using `git pull`.

4.1 Preliminaries

This section must be read **carefully**.

Important remark From now on, we add the following restriction to the MiniC language: Values (variables, argument of `println_int`) are of type (signed) `int` or `bool` only (no float, no string, no char). Thus all values can be stored in regular registers or in one cell (64 bits) in memory. You can let your program crash if another type of variable is provided.

Note that real compilers would perform the code generation from a decorated AST (with type annotations attached to nodes). For simplicity, we will work on the non-decorated AST: our language is simple enough to generate code without decorations.

Structure of the compiler's code

- In `TP04/APIRiscV.py` we provide you with utility functions to encode 3-address RISCV instructions. Instruction classes are in `TP04/Instruction3A.py` and `TP04/Operands.py`. An Instruction is either a Comment, a Label, or a `Instru3A`; it has arguments which can be immediate numbers (of type `Immediate`), temporaries (of type `Temporary`), regular registers (`Register` [1]), offsets in memory (`Offset`).

- A RISCV program contains a list of instructions, and also a temporary pool (temporary variables).

- In Section 4.2, you will use an instance of the `RiscVFunction` class in order to construct a list of such instructions via calls to `addInstructionXXX` methods. A call to the `printCode` method will dump this code into a text file.

- File `TP05/Allocations.py` is responsible for the allocation part. From a `RiscVFunction` with temporaries (instructions formed with temporaries), producing an actual RISCV program (instructions with regular registers or memory accesses) is done by the two following steps:

[1] in the library, Python constants representing registers are in capital letters, but in lowercase when they are printed.

  - First, compute an allocation for each temporary (in the current `RiscVFunction` instance). In Section 4.3, we provide you with `NaiveAllocator.run()` in `Allocations.py` which computes such a (naive) allocation, you will have to design your own allocation function in Section 4.5.
  - For each instruction of the program, if the instruction contains a read or write access to a temporary, replace operands with the corresponding actual registers/memory location (and possibly add some instructions before and after). This is done by the use of the `RiscVFunction.iter_instructions` iterator on instructions and `Allocations.replace_reg` methods. In Section 4.5 you will have to write such a "replacement" function.

- The file `TP03/MiniCTypingVisitor.py` is the same that was provided for lab3.
- The file `MiniCC.py` launches the chain: production of 3-address code with temporaries, allocation, replacement, print.
- The script `test_codegen.py` will help you test your code. We will use it in Section 4.3.
- A `README-codegen.md` file to be completed progressively during the lab.

EXERCISE #1 ► RISCV **Simulator - test**
Re-test the command-line version of the RISCV simulator, for example with code from TP01:

```
cd ../TP01/code/
riscv64-unknown-elf-gcc libprint.s test_print.s -o test_print.riscv
spike pk test_print.riscv
cd ../../MiniC/
```

4.1.1 Conventions used in the assembly code

- All data items are stored on 64 bits (double-words, 8 bytes)
- Registers `s1`, `s2`, and `s3` are reserved for temporary computations (e.g. to compute an address before a `sd` or a `ld`, or to store a value between a memory access and an arithmetic operation). Note that `s0` is an alias for `fp`, hence `s0` must not be used as a general purpose register either.
- Registers `s4`, ..., `s11`, `t0`, ..., `t6` are general purpose registers, that can be used freely by the code generator. In your Python code, you can access the list of general-purpose registers with `Operands.GP_REGS`. `si` and `ti` registers will behave differently in presence of function calls, but are considered equivalent for now.
- To store properly in memory, it is mandatory to compute offsets from the "reserved" register `fp`. To be compatible with the RISCV ecosystem, we will use a stack **growing with decreasing addresses**. Thus data in the stack is accessed by adding a **negative offset** (multiple of 8) to `fp`. The `sp` register points to the first data contained in the stack. It is always 16-byte (2 double-words) aligned.

4.2 First step: three-address code generation

In this section you have to implement the course rules in order to produce RISCV code with temporaries. These rules are given in Figure 4.2 on page 9 and Figure 4.3 on page 10.

Here is an example of the expected output of this part. From the following MiniC program:

```
#include "printlib.h"

int main() {
    int a,n;
    n = 1;
    a = 7;
    while (n < a) {
      n = n+1;
```

```
    }
    println_int(n);
    return 0;
}
```

the following code is supposed to be generated:

```
##Automatically generated RISCV code, MIF08 & CAP 2019
##non executable 3-Address instructions version

##prelude
# [...] Some automatically generated code that will be explained in a future lab

##Generated Code
# [...] Some automatically generated code that will be explained in a future lab
        # (stat (assignment n = (expr (atom 1))) ;)
        li temp_2, 1
        mv temp_0, temp_2
        # (stat (assignment a = (expr (atom 7))) ;)
        li temp_3, 7
        mv temp_1, temp_3
        # (stat (while_stat while ( (expr (expr (atom n)) < (expr (atom a))) ) (
    stat_block { (block (stat (assignment n = (expr (expr (atom n)) + (expr (atom 1)))
    ) ;)) })))
lbl_begin_while_1_main:
        li temp_4, 0
        bge temp_0, temp_1, lbl_end_relational_3_main
        li temp_4, 1
lbl_end_relational_3_main:
        beq temp_4, zero, lbl_end_while_2_main
        # (stat (assignment n = (expr (expr (atom n)) + (expr (atom 1)))) ;)
        li temp_5, 1
        add temp_6, temp_0, temp_5
        mv temp_0, temp_6
        j lbl_begin_while_1_main
lbl_end_while_2_main:
        # (stat (print_stat println_int ( (expr (atom n)) ) ;))
        mv a0, temp_0
        call println_int
# [...] Some automatically generated code that will be explained in a future lab

##postlude

# [...] Some automatically generated code that will be explained in a future lab
```

EXERCISE #2 ► **3-address code generation**

In the archive, we provide you a main and an incomplete `MiniCCodeGen3AVisitor.py`. To test it, type

```
python3 MiniCC.py TP04/tests/provided/step1/test00.c --reg-alloc=naive
```

Don't forget to run `make` if you need to regenerate the lexer and parser with ANTLR (i.e. if `python3` complains with `No module named 'MiniCLexer'`). Observe the generated code in `<samepath>/test00.s`[2]. You now have to implement the 3-address code generation rules seen in the course. Code and test incrementally [3]:

[2] We generated RISCV comments with MiniC statements for debug.

[3] Using files in the `TP04/tests/*` directories. All the test files you use will have to be in your archive.

- We give you the code generation for the `println_int` instruction. It basically produces a call to the proper function in the library.
- numerical expressions without variables (constants are expected to hold on 64 bits, no boolean expression for the moment!).
- then check that (numerical) expressions with variables work (assignment and usage of variables in expressions are given); we advise you to postpone the implementation of MultiplicativeExpr, and first finish this Lab without them.

At this step, the code generation is not finished, but we will do some allocation to be able to test properly. All examples in `tests/provided/step1` directory should generate code without any error at this point:

```
for i in TP04/tests/provided/step1/*.c; do
  echo "file="$i; python3 MiniCC.py --reg-alloc=none $i >/dev/null;
done
```

4.3 Testing with the trivial allocator (and real RISCV instructions)

The former code is not executable since it uses temporaries. We provide you with an allocation method which allocates temporaries in registers as long as possible, and fails if there is no available registers. The process takes as input the former 3-address code and transforms each instruction according to the allocation function.

EXERCISE #3 ▸ **Testing the trivial allocator**
Open, read, understand the `NaiveAllocator` implementation in `Allocations.py` and how it is used to perform the actual RISCV code generation [4]. Then, intensively test your former code generation with this allocator [5]:

1. Have a look at the `test_codegen.py` script: comment or uncomment files to test, and what to test.
2. Test with:
   `make TEST_FILES="TP04/tests/provided/step1/*.c" tests-naive`
   This script tests all files specified in `TEST_FILES` (or, if not specified, all files in the `*/tests/*` directories except those whose name start with a special character):
   - if the pragma `// EXPECTED` is present in the file, it compares the actual output after assembling and simulating with the list of expected values. For instance:

     ```
     int main(){
       int x, y;
       x = 42;
       println_int(x);
       y = x + 8;
       println_int(y);
       return 0;
     }
     // EXPECTED
     // 42
     // 50
     ```

     is a great test case to test assignments.
   - If the `AllocationError` exception is raised by the naive allocator, the test is skipped.
   - If the compilation succeeded, it compares the actual output after assembling and simulating to the `// EXPECTED` statements given in the file (which are themselves compared to the output given by `riscv64-unknown-elf-gcc`).
   - For debugging, you can obviously launch your compiler manually with e.g.

     ```
     python3 MiniCC.py --reg-alloc naive --stdout TP04/tests/provided/step1/test00.c
     ```

[4] All available registers are in a list named `GP_REGS`
[5] Be careful, this allocator crashes if there is more than a certain number of temporaries!

Run `python3 MiniCC.py --help` or see `MiniCC.py` for more options. The `--debug` option allows getting some debug output. Alternatively, you can run the testsuite on a single testfile with:

```
make TEST_FILES=TP04/tests/provided/step1/test00.c tests-naive
```

At this step, the tests should be OK or SKIPPED for all files given in directory `tests/step1/`:

```
make tests
[...]
=========================== xx passed, xx skipped in xx seconds ========
```

"skipped" here means that we cannot compare the output to the ideal output since some of our 3 adress-codes cannot be allocated with registers only. That's life !

Now that we have a way to test our code generation for tiny MiniC codes, we can come back to it.

4.4 Finish 3 address code generation

Now that you know how to test your code using the naive allocator, go back to code generation and finish it.

EXERCISE #4 ► **A few corner-cases**
Some points may require extra care (implementation and test):

- Don't forget the automatic initialization (in MiniC and not real C). Unlike the interpreter, initialization cannot be done by initializing a Python dictionary, you need to generate the initialization code.
- Don't forget the explicit errors for division by zero. We provide you a piece of assembly code raising the error (see `RiscVFunction.printCode()`), you need to generate the instruction to jump to this label (get the label with `RiscVFunction.get_label_div_by_zero()`) when the right operand of a division or modulo is 0.
- `float` and `string` are unsupported. Raise `MiniCUnsupportedError` when you encounter any of them. Tests are provided for this, so if you're not sure what to do: make the tests pass!

The specifications are the same as for the previous (semantics, typing, and exit code).

EXERCISE #5 ► **End of 3-address code generation for MiniC**
Implement the 3-address code generation rules:

- for boolean expressions and numerical comparison: compute 1 (true) or 0 (false) in the destination register; be careful the `not` boolean instruction is not what you want.
- while loops;
- if then else. **Be careful with nested ifs and their labels!**.

At this point all the tests should be ok for all files in directory `TP04/tests/provided/step2/` (You should modify the test script pathes). However these tests are not sufficient, you should add some other ones (in the directory `TP04/tests/students/`). Run the testsuite with `make tests-naive` to use all the test files.

About `if` and `while` For tests (and boolean expressions), make sure you generate "conditional jumps" with:

```
self._prog.addInstructionCondJUMP(label, op1, cond, op2)
```

where `op1` (resp `op2`) is the left operand (resp right operand or the numerical constant 0, nothing else), ie a register or a value of the boolean condition (`Condition('eq')` for equality, for instance) [6], and `label` is a label to jump to if the condition evaluates to true.

4.5 RISCV code with "all-in-mem" allocation of temporaries

Tests Up to now, you used `make tests-naive` to test your code, and at this point all tests should pass, or be skipped. From now, you should use the more complete `make tests-notsmart` command, that tests everything except the smart allocator (that we'll write during the next lab).

Check that `make tests-notsmart` does fail.

[6] We suggest to use `grep` and find this class definition and this method somewhere in the code we provide.

Implementation As the number of registers for allocation is bounded by N [7], the naive allocator cannot deal with more than N temporaries: we have to find a way to store the results elsewhere. In this particular lab, we will use the following solution:

- the generated code will use memory locations in the stack, and will not use registers a_1 to a_7 at all for the moment.
- but all values that are propagated from one rule to another (sub-expressions, ...) must be stored in the stack, whose address will be stored in FP (as defined in RiscVFunction.printCode).
- $s1, s2, s3$ will be used to compute the value to store or as a destination register for the value(s) to read. **Technically, only 2 of these registers are mandatory, but you should be precautionous if you try a 2-registers-only solution.**
- In order to know if a given (temporary) operand should be read and/or written, use the `is_read_ony` method of the `Instruction3A` class.

Figure 4.1 depicts the stack implementation for the RISCV machine, that follows the RISC-V calling convention (stack growing downwards, stack-pointer always 16-bytes aligned).

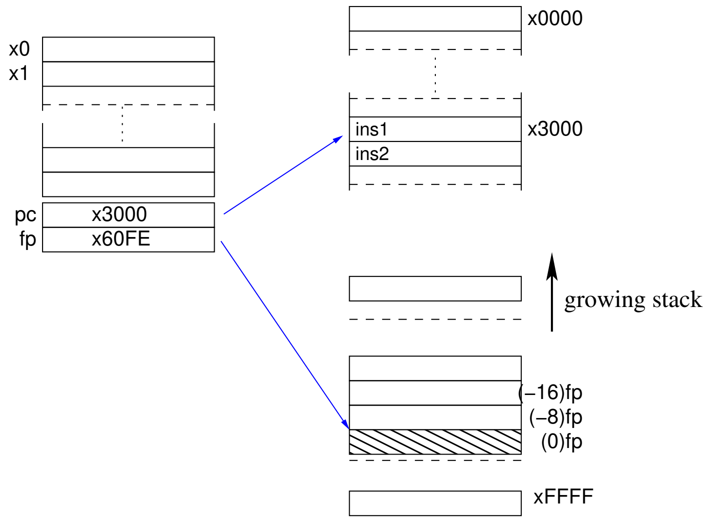

Figure 4.1: Memory model for RISCV

Following the convention that `fp` always stores the "begining of stack address", pushing the content of register $s3$ in the stack will be done following the steps:

- compute a new offset (call to the `new_offset` method of the class `RiscVFunction`).
- generate the following instruction:

```
sd s3, -offset*8(fp)
# sd = store double = 64-bits store
# -offset*8(fp) = memory location at address fp-offset*8
```

Getting back the value is similar.

EXERCISE #6 ► **Manual translation**

Complete the expected output for the following two statements (13/15 lines of RISCV code). `temp_3` is located at −32(fp) and `temp_4` is located at −40(fp):

```
int x, y;
x=4;
y=12+x
```

[7] The size of the `GP_REGS` list in the `Operands.py` file, i.e. `len(Operands.GP_REGS)`

Listing 4.1: 'all in mem alloc for test00b.c'

```
##Generated code without prelude and postlude
        # (stat (assignment x = (expr (atom 4))) ;)
        # li temp_2, 4
        li s2, 4
        sd s2, -24(fp)
        # end li temp_2, 4
        # mv temp_1, temp_2
        ld s1, -24(fp)
        mv s2, s1
        sd s2, -16(fp)
        # end mv temp_1, temp_2
        # (stat (assignment y = (expr (expr (atom 12)) + (expr (atom x)))) ;)
        # li temp_3, 12
        # TODO 2 lines


        # end li temp_3, 12
        # add temp_4, temp_3, temp_1
        # TODO 4 lines




        # end add temp_4, temp_3, temp_1
        # mv temp_0, temp_4
        # NOT TODO
```

EXERCISE #7 ► **Implement**

Now you are on your own to implement this code generation. Here are the main steps (less than 50 locs of PYTHON):

1. We have implemented for you an `AllInMemAllocator.run()` method in `Allocations.py`. This method only maps each temporary ("temporary") to a new offset in memory (in a PYTHON `dict`), then iterates the `replace_mem` function on all instructions of the three adress program to perform the actual allocation.

2. In `Allocations.py`, implement a `replace_mem(old_i)` that takes as input a "3-address with temporaries" RISCV code and outputs a list of instructions as a replacement. For instance, each time we access a source operand, we have to load it from memory before, thus the `replace_mem` should contain something like

```
# regxxx is the register used to hold the value between the load and
# the operation itself (one of t0, t1, t2).
# operand is the place in memory where the temporary is allocated (of
# the form Offset(..., fp), obtained with get_alloced_loc().
before.append(Instru3A('ld', regxxx , operand))
```

The files you generate have to be tested with the RISCV simulator with the same script as before. **Of course, with "all-in-mem" allocation, there should not be any "skipped" test any more.**

More tests Now that your compiler can deal with a large number of temporaries, make sure all features are heavily tested (the testsuite we provide is in no way sufficient).

4.6 Multiplicative Expressions (multiplication, division, modulo)

If not already done, extend your work to multiplicative expressions. Conventions for division and multiplication should be the same as in C: division is truncated toward zero, and modulo is such that $(a/b)*b+a\%b=a$.

$$
\begin{array}{rclcrcl}
4/3 & = & 1 & \quad & 4\%3 & = & 1 \\
(-4)/3 & = & -1 & & (-4)\%3 & = & -1 \\
4/(-3) & = & -1 & & 4\%(-3) & = & 1 \\
(-4)/(-3) & = & 1 & & (-4)\%(-3) & = & -1
\end{array}
$$

c	`dest <- new_tmp()`
	`code.add("li dest, c")`
	`return dest`

x	`# get the temporary associated to x.`
	`reg <- symbol_table[x]`
	`return reg`

e_1+e_2	`t1 <- GenCodeExpr(e_1)`
	`t2 <- GenCodeExpr(e_2)`
	`dest <- new_tmp()`
	`code.add("add dest, t1, t2")`
	`return dest`

e_1-e_2	`t1 <- GenCodeExpr(e_1)`
	`t2 <- GenCodeExpr(e_2)`
	`dest <- new_tmp()`
	`code.add("sub dest, t1, t2")`
	`return dest`

true	`dest <-new_tmp()`
	`code.add("li dest, 1")`
	`return dest`

$e_1 < e_2$	`dest <- new_tmp()`
	`t1 <- GenCodeExpr(e1)`
	`t2 <- GenCodeExpr(e2)`
	`endrel <- new_label()`
	`code.add("li dest, 0")`
	`# if t1>=t2 jump to endrel`
	`code.add("bge endrel, t1, t2")`
	`code.add("li dest, 1")`
	`code.addLabel(endrel)`
	`return dest`

Figure 4.2: 3@ Code generation for numerical or Boolean expressions

x = e	<pre>dest <- GenCodeExpr(e)
	loc <- symbol_table[x]
	code.add("mv *loc*, *dest*")</pre>

S1; S2	<pre># Just concatenate codes
	GenCodeSmt(S1)
	GenCodeSmt(S2)</pre>

`if` b `then` $S1$ `else` $S2$	<pre>lelse <- new_label()
	lendif <- new_label()
	t1 <- GenCodeExpr(b)
	#if the condition is false, jump to else
	code.add("beq *lelse*, *t1*, 0")
	GenCodeSmt(S1) # then
	code.add("j *lendif*")
	code.addLabel(lelse)
	GenCodeSmt(S2) # else
	code.addLabel(lendif)</pre>

`while` b `do` S `done`	<pre>ltest <- new_label()
	lendwhile <- new_label()
	code.addLabel(ltest)
	t1 <- GenCodeExpr(b)
	code.add("beq *lendwhile*, *t1*, 0")
	GenCodeSmt(S) # execute S
	code.add("j *ltest*") # and jump to the test
	code.addLabel(lendwhile) # else it is done.</pre>

Figure 4.3: 3@ Code generation for Statements

Appendix A

RISCV Assembly Documentation (ISA), rv64g

About

- RISCV is an open instruction set initially developed by Berkeley University, used among others by Western Digital, Alibaba and Nvidia.
- We are using the rv64g instruction set: **R**isc-**V**, 64 bits, **G**eneral purpose (base instruction set, and extensions for floating point, atomic and multiplications), without compressed instructions. In practice, we will use only 32 bits instructions (and very few of floating point instructions).
- Document: Laure Gonnord and Matthieu Moy, for CAP and MIF08.

This is a simplified version of the machine, which is (hopefully) conform to the chosen simulator.

A.1 Installing the simulator and getting started

To get the RISCV assembler and simulator, follow instructions of the first lab (git pull on the course lab repository).

A.2 The RISCV architecture

Here is an example of RISCV assembly code snippet (a proper `main` function would be needed to execute it, cf. course and lab):

```
  addi a0, zero, 17  # initialisation of  a register to 17
loop:
  addi a0, a0, -1    # subtraction of an immediate
  j  loop            # equivalent to jump xx
```

The rest of the documentation is adapted from `https://github.com/riscv/riscv-asm-manual/blob/master/riscv-asm.md` and `https://github.com/jameslzhu/riscv-card/blob/master/riscv-card.pdf`

A.3 RISC-V Assembly Programmer's Manual - adapted for CAP and MIF08

A.3.1 Copyright and License Information - Documents

- Official Specifications webpage: https://riscv.org/specifications/
- Latest Specifications draft repository: https://github.com/riscv/riscv-isa-manual

This document has been modified by Laure Gonnord & Matthieu Moy, in 2019.

A.3.2 Registers

Registers are the most important part of any processor. RISC-V defines various types, depending on which extensions are included: The general registers (with the program counter), control registers, floating point registers (F extension), and vector registers (V extension). We won't use control nor F or V registers.

General registers

The RV32I base integer ISA includes 32 registers, named `x0` to `x31`. The program counter `PC` is separate from these registers, in contrast to other processors such as the ARM-32. The first register, `x0`, has a special function: Reading it always returns 0 and writes to it are ignored.

In practice, the programmer doesn't use this notation for the registers. Though `x1` to `x31` are all equally general-use registers as far as the processor is concerned, by convention certain registers are used for special tasks. In assembler, they are given standardized names as part of the RISC-V **application binary interface** (ABI). This is what you will usually see in code listings. If you really want to see the numeric register names, the `-M` argument to objdump will provide them.

Register	ABI	Use by convention	Preserved?
x0	zero	hardwired to 0, ignores writes	*n/a*
x1	ra	return address for jumps	no
x2	sp	stack pointer	yes
x3	gp	global pointer	*n/a*
x4	tp	thread pointer	*n/a*
x5	t0	temporary register 0	no
x6	t1	temporary register 1	no
x7	t2	temporary register 2	no
x8	s0 *or* fp	saved register 0 *or* frame pointer	yes
x9	s1	saved register 1	yes
x10	a0	return value *or* function argument 0	no
x11	a1	return value *or* function argument 1	no
x12	a2	function argument 2	no
x13	a3	function argument 3	no
x14	a4	function argument 4	no
x15	a5	function argument 5	no
x16	a6	function argument 6	no
x17	a7	function argument 7	no
x18	s2	saved register 2	yes
x19	s3	saved register 3	yes
x20	s4	saved register 4	yes
x21	s5	saved register 5	yes
x22	s6	saved register 6	yes
x23	s7	saved register 6	yes
x24	s8	saved register 8	yes
x25	s9	saved register 9	yes
x26	s10	saved register 10	yes
x27	s11	saved register 11	yes
x28	t3	temporary register 3	no
x29	t4	temporary register 4	no
x30	t5	temporary register 5	no
x31	t6	temporary register 6	no
pc	*(none)*	program counter	*n/a*

Registers of the RV32I. Based on RISC-V documentation and Patterson and Waterman "The RISC-V Reader" (2017)

As a general rule, the **saved registers** `s0` to `s11` are preserved across function calls, while the **argument**

registers a0 to a7 and the **temporary registers** t0 to t6 are not. The use of the various specialized registers such as sp by convention will be discussed later in more detail.

A.3.3 Instructions

Arithmetic

add, addi, sub, classically.

```
addi a0, zero, 42
```

initialises a0 to 42.

Labels

Text labels are used as branch, unconditional jump targets and symbol offsets. Text labels are added to the symbol table of the compiled module.

```
loop:
        j loop
```

Jumps and branches target is encoded with a relative offset. It is relative to the beginning of the current instruction. For example, the self-loop above corresponds to an offset of 0.

Branching

Test and jump, within the same instruction:

```
    beq a0, a1, end
```

tests whether a0=a1, and jumps to 'end' if its the case.

Absolute addressing

The following example shows how to load an absolute address:

```
.section .text
.globl _start
_start:
        lui a0,       %hi(msg)       # load msg(hi)
        addi a0, a0,  %lo(msg)       # load msg(lo)
        jal ra, puts
2:      j 2b

.section .rodata
msg:
        .string "Hello World\n"
```

which generates the following assembler output and relocations as seen by objdump:

```
0000000000000000 <_start>:
   0:   000005b7            lui a1,0x0
            0: R_RISCV_HI20 msg
   4:   00858593            addi    a1,a1,8 # 8 <.L21>
            4: R_RISCV_LO12_I   msg
```

Relative addressing

The following example shows how to load a PC-relative address:

```
.section .text
.globl _start
_start:
1:      auipc a0,     %pcrel_hi(msg) # load msg(hi)
        addi  a0, a0, %pcrel_lo(1b)  # load msg(lo)
        jal ra, puts
2:      j 2b

.section .rodata
msg:
        .string "Hello World\n"
```

which generates the following assembler output and relocations as seen by objdump:

```
0000000000000000 <_start>:
   0:   00000597            auipc   a1,0x0
            0: R_RISCV_PCREL_HI20   msg
   4:   00858593            addi    a1,a1,8 # 8 <.L21>
            4: R_RISCV_PCREL_LO12_I .L11
```

Load Immediate

The following example shows the `li` pseudo instruction which is used to load immediate values:

```
        li a0, 0x76543210
```

which generates the following assembler output as seen by objdump (generated code will be different depending on the constant):

```
   0:   76543537                lui     a0,0x76543
   4:   2105051b                addiw   a0,a0,528
```

Load Address

The following example shows the `la` pseudo instruction which is used to load symbol addresses:

```
.section .text
.globl _start
_start:

        la a0, msg

.section .rodata
msg:
        .string "Hello World\n"
```

A.3.4 Assembler directives for CAP and MIF08

Both the RISC-V-specific and GNU .-prefixed options.

The following table lists assembler directives:

Directive	Arguments	Description
.align	integer	align to power of 2 (alias for .p2align)

Directive	Arguments	Description
.file	“filename”	emit filename FILE LOCAL symbol table
.globl	symbol_name	emit symbol_name to symbol table (scope GLOBAL)
.local	symbol_name	emit symbol_name to symbol table (scope LOCAL)
.section	[{.text,.data,.rodata,.bss}]	emit section (if not present, default .text) and make current
.size	symbol, symbol	accepted for source compatibility
.text		emit .text section (if not present) and make current
.data		emit .data section (if not present) and make current
.rodata		emit .rodata section (if not present) and make current
.string	“string”	emit string
.equ	name, value	constant definition
.word	expression [, expression]*	32-bit comma separated words
.balign	b,[pad_val=0]	byte align
.zero	integer	zero bytes

A.3.5 Assembler Relocation Functions

The following table lists assembler relocation expansions:

Assembler Notation	Description	Instruction / Macro
%hi(symbol)	Absolute (HI20)	lui
%lo(symbol)	Absolute (LO12)	load, store, add
%pcrel_hi(symbol)	PC-relative (HI20)	auipc
%pcrel_lo(label)	PC-relative (LO12)	load, store, add

A.3.6 Instruction encoding

Credit This is a subset of the RISC-V greencard, by James Izhu, licence CC by SA, `https://github.com/jameslzhu/riscv-card`

Core Instruction Formats

<table>
<tr><th>31 27</th><th>26 25</th><th>24 20</th><th>19 15</th><th>14 12</th><th>11 7</th><th>6 0</th><th></th></tr>
<tr><td colspan="2">funct7</td><td>rs2</td><td>rs1</td><td>funct3</td><td>rd</td><td>opcode</td><td>R-type</td></tr>
<tr><td colspan="3">imm[11:0]</td><td>rs1</td><td>funct3</td><td>rd</td><td>opcode</td><td>I-type</td></tr>
<tr><td colspan="2">imm[11:5]</td><td>rs2</td><td>rs1</td><td>funct3</td><td>imm[4:0]</td><td>opcode</td><td>S-type</td></tr>
<tr><td colspan="2">imm[12|10:5]</td><td>rs2</td><td>rs1</td><td>funct3</td><td>imm[4:1|11]</td><td>opcode</td><td>B-type</td></tr>
<tr><td colspan="5">imm[31:12]</td><td>rd</td><td>opcode</td><td>U-type</td></tr>
<tr><td colspan="5">imm[20|10:1|11|19:12]</td><td>rd</td><td>opcode</td><td>J-type</td></tr>
</table>

RV32I Base Integer Instructions - CAP subset

Inst	Name	FMT	Opcode	funct3	funct7	Description (C)	Note
add	ADD	R	0110011	0x0	0x00	rd = rs1 + rs2	
sub	SUB	R	0110011	0x0	0x20	rd = rs1 - rs2	
xor	XOR	R	0110011	0x4	0x00	rd = rs1 ^ rs2	
or	OR	R	0110011	0x6	0x00	rd = rs1 | rs2	
and	AND	R	0110011	0x7	0x00	rd = rs1 & rs2	
slt	Set Less Than	R	0110011	0x2	0x00	rd = (rs1 < rs2)?1:0	
sltu	Set Less Than (U)	R	0110011	0x3	0x00	rd = (rs1 < rs2)?1:0	zero-extends
addi	ADD Immediate	I	0010011	0x0		rd = rs1 + imm	
xori	XOR Immediate	I	0010011	0x4		rd = rs1 ^ imm	
ori	OR Immediate	I	0010011	0x6		rd = rs1 | imm	
andi	AND Immediate	I	0010011	0x7		rd = rs1 & imm	
lb	Load Byte	I	0000011	0x0		rd = M[rs1+imm][0:7]	
lw	Load Word	I	0000011	0x2		rd = M[rs1+imm][0:31]	
lbu	Load Byte (U)	I	0000011	0x4		rd = M[rs1+imm][0:7]	zero-extends
sb	Store Byte	S	0100011	0x0		M[rs1+imm][0:7] = rs2[0:7]	
sw	Store Word	S	0100011	0x2		M[rs1+imm][0:31] = rs2[0:31]	
beq	Branch ==	B	1100011	0x0		if(rs1 == rs2) PC += imm	
bne	Branch !=	B	1100011	0x1		if(rs1 != rs2) PC += imm	
blt	Branch <	B	1100011	0x4		if(rs1 < rs2) PC += imm	
bge	Branch ≥	B	1100011	0x5		if(rs1 >= rs2) PC += imm	
bltu	Branch < (U)	B	1100011	0x6		if(rs1 < rs2) PC += imm	zero-extends
bgeu	Branch ≥ (U)	B	1100011	0x7		if(rs1 >= rs2) PC += imm	zero-extends
jal	Jump And Link	J	1101111			rd = PC+4; PC += imm	
jalr	Jump And Link Reg	I	1100111	0x0		rd = PC+4; PC = rs1 + imm	
lui	Load Upper Imm	U	0110111			rd = imm << 12	
auipc	Add Upper Imm to PC	U	0010111			rd = PC + (imm << 12)	

Pseudo Instructions

<table>
<tr><th>Pseudoinstruction</th><th>Base Instruction(s)</th><th>Meaning</th></tr>
<tr><td>la rd, symbol</td><td>auipc rd, symbol[31:12]
addi rd, rd, symbol[11:0]</td><td>Load address</td></tr>
<tr><td>{lb|lh|lw|ld} rd, symbol</td><td>auipc rd, symbol[31:12]
{lb|lh|lw|ld} rd, symbol[11:0](rd)</td><td>Load global</td></tr>
<tr><td>{sb|sh|sw|sd} rd, symbol, rt</td><td>auipc rt, symbol[31:12]
s{b|h|w|d} rd, symbol[11:0](rt)</td><td>Store global</td></tr>
<tr><td>{flw|fld} rd, symbol, rt</td><td>auipc rt, symbol[31:12]
fl{w|d} rd, symbol[11:0](rt)</td><td>Floating-point load global</td></tr>
<tr><td>{fsw|fsd} rd, symbol, rt</td><td>auipc rt, symbol[31:12]
fs{w|d} rd, symbol[11:0](rt)</td><td>Floating-point store global</td></tr>
<tr><td>nop</td><td>addi x0, x0, 0</td><td>No operation</td></tr>
<tr><td>li rd, immediate</td><td>Myriad sequences</td><td>Load immediate</td></tr>
<tr><td>mv rd, rs</td><td>addi rd, rs, 0</td><td>Copy register</td></tr>
<tr><td>not rd, rs</td><td>xori rd, rs, -1</td><td>One’s complement</td></tr>
<tr><td>neg rd, rs</td><td>sub rd, x0, rs</td><td>Two’s complement</td></tr>
<tr><td>negw rd, rs</td><td>subw rd, x0, rs</td><td>Two’s complement word</td></tr>
<tr><td>sext.w rd, rs</td><td>addiw rd, rs, 0</td><td>Sign extend word</td></tr>
<tr><td>seqz rd, rs</td><td>sltiu rd, rs, 1</td><td>Set if = zero</td></tr>
<tr><td>snez rd, rs</td><td>sltu rd, x0, rs</td><td>Set if ≠ zero</td></tr>
<tr><td>sltz rd, rs</td><td>slt rd, rs, x0</td><td>Set if < zero</td></tr>
<tr><td>sgtz rd, rs</td><td>slt rd, x0, rs</td><td>Set if > zero</td></tr>
<tr><td>fmv.s rd, rs</td><td>fsgnj.s rd, rs, rs</td><td>Copy single-precision register</td></tr>
<tr><td>fabs.s rd, rs</td><td>fsgnjx.s rd, rs, rs</td><td>Single-precision absolute value</td></tr>
<tr><td>fneg.s rd, rs</td><td>fsgnjn.s rd, rs, rs</td><td>Single-precision negate</td></tr>
<tr><td>fmv.d rd, rs</td><td>fsgnj.d rd, rs, rs</td><td>Copy double-precision register</td></tr>
<tr><td>fabs.d rd, rs</td><td>fsgnjx.d rd, rs, rs</td><td>Double-precision absolute value</td></tr>
<tr><td>fneg.d rd, rs</td><td>fsgnjn.d rd, rs, rs</td><td>Double-precision negate</td></tr>
<tr><td>beqz rs, offset</td><td>beq rs, x0, offset</td><td>Branch if = zero</td></tr>
<tr><td>bnez rs, offset</td><td>bne rs, x0, offset</td><td>Branch if ≠ zero</td></tr>
<tr><td>blez rs, offset</td><td>bge x0, rs, offset</td><td>Branch if ≤ zero</td></tr>
<tr><td>bgez rs, offset</td><td>bge rs, x0, offset</td><td>Branch if ≥ zero</td></tr>
<tr><td>bltz rs, offset</td><td>blt rs, x0, offset</td><td>Branch if < zero</td></tr>
<tr><td>bgtz rs, offset</td><td>blt x0, rs, offset</td><td>Branch if > zero</td></tr>
<tr><td>bgt rs, rt, offset</td><td>blt rt, rs, offset</td><td>Branch if ></td></tr>
<tr><td>ble rs, rt, offset</td><td>bge rt, rs, offset</td><td>Branch if ≤</td></tr>
<tr><td>bgtu rs, rt, offset</td><td>bltu rt, rs, offset</td><td>Branch if >, unsigned</td></tr>
<tr><td>bleu rs, rt, offset</td><td>bgeu rt, rs, offset</td><td>Branch if ≤, unsigned</td></tr>
<tr><td>j offset</td><td>jal x0, offset</td><td>Jump</td></tr>
<tr><td>jal offset</td><td>jal x1, offset</td><td>Jump and link</td></tr>
<tr><td>jr rs</td><td>jalr x0, rs, 0</td><td>Jump register</td></tr>
<tr><td>jalr rs</td><td>jalr x1, rs, 0</td><td>Jump and link register</td></tr>
<tr><td>ret</td><td>jalr x0, x1, 0</td><td>Return from subroutine</td></tr>
<tr><td>call offset</td><td>auipc x1, offset[31:12]
jalr x1, x1, offset[11:0]</td><td>Call far-away subroutine</td></tr>
<tr><td>tail offset</td><td>auipc x6, offset[31:12]
jalr x0, x6, offset[11:0]</td><td>Tail call far-away subroutine</td></tr>
<tr><td>fence</td><td>fence iorw, iorw</td><td>Fence on all memory and I/O</td></tr>
</table>