

RISCV ISA - Documentation

Ésisar, ENSL, Université Lyon 1

2022-2023

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: **Risc-V**, 64 bits, **General 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 and CS444 and CE313 and ...

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

1 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):

```
1  addi a0, zero, 17 # initialisation of a register to 17
2  loop:
3  addi a0, a0, -1   # subtraction of an immediate
4  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/jameslzh/riscv-card/blob/master/riscv-card.pdf>

2 RISC-V Assembly Programmer's Manual - adapted for CAP and MIF08 and CE313 and CS444

2.1 Copyright and License Information - Documents

The RISC-V Assembly Programmer's Manual is

© 2017 Palmer Dabbelt palmer@dabbelt.com © 2017 Michael Clark michaeljclark@mac.com © 2017 Alex Bradbury asb@lowrisc.org

It is licensed under the Creative Commons Attribution 4.0 International License (CC-BY 4.0). The full license text is available at <https://creativecommons.org/licenses/by/4.0/>.

- 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 2019-2022

2.2 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.

2.2.1 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	<i>n/a</i>
x1	ra	return address for jumps	no
x2	sp	stack pointer	yes
x3	gp	global pointer	<i>n/a</i>
x4	tp	thread pointer	<i>n/a</i>
x5	t0	temporary register 0	no
x6	t1	temporary register 1	no
x7	t2	temporary register 2	no
x8	s0 <i>or</i> fp	saved register 0 <i>or</i> frame pointer	yes
x9	s1	saved register 1	yes
x10	a0	return value <i>or</i> function argument 0	no
x11	a1	return value <i>or</i> 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	<i>(none)</i>	program counter	<i>n/a</i>

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.

2.3 Instructions

2.3.1 Arithmetic

`add`, `addi`, `sub`, classically.

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

initialises `a0` to 42.

2.3.2 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.

2.3.3 Branching

Test and jump, within the same instruction:

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

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

2.3.4 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_L012_I msg
```

2.3.5 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
```

2.3.6 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
```

2.3.7 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"
```

2.4 Assembler directives

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)
.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

2.5 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

2.6 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

31	27	26	25	24	20	19	15	14	12	11	7	6	0	
funct7			rs2		rs1		funct3		rd		opcode			R-type
imm[11:0]					rs1		funct3		rd		opcode			I-type
imm[11:5]			rs2		rs1		funct3		imm[4:0]		opcode			S-type
imm[12:10:5]			rs2		rs1		funct3		imm[4:1 11]		opcode			B-type
imm[31:12]									rd		opcode			U-type
imm[20 10:1 11 19:12]									rd		opcode			J-type

“imm[x:y]” means “bits x to y from binary representation of imm”. “imm[y|x]” means “bits y, then x of imm”.

RV32I Base Integer Instructions - (excerpt)

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	$rd = rs1 \wedge rs2$	
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		$rd = rs1 \wedge imm$	
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 \geq	B	1100011	0x5		$if(rs1 > rs2) PC += imm$	
bltu	Branch < (U)	B	1100011	0x6		$if(rs1 < rs2) PC += imm$	zero-extends
bgeu	Branch \geq (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			$rd = imm \ll 12$	
auipc	Add Upper Imm to PC	U	0010111			$rd = PC + (imm \ll 12)$	

Pseudo Instructions (excerpt)

Pseudoinstruction	Base Instruction(s)	Meaning
la rd, symbol	auipc rd, symbol[31:12] addi rd, rd, symbol[11:0]	Load address
{lb lh lw ld} rd, symbol	auipc rd, symbol[31:12] {lb lh lw ld} rd, symbol[11:0](rd)	Load global
{sb sh sw sd} rd, symbol, rt	auipc rt, symbol[31:12] s{b h w d} rd, symbol[11:0](rt)	Store global
nop	addi x0, x0, 0	No operation
li rd, immediate	<i>Myriad sequences</i>	Load immediate
mv rd, rs	addi rd, rs, 0	Copy register
not rd, rs	xori rd, rs, -1	One's complement
neg rd, rs	sub rd, x0, rs	Two's complement
sext.w rd, rs	addiw rd, rs, 0	Sign extend word
seqz rd, rs	sltiu rd, rs, 1	Set if = zero
beqz rs, offset	beq rs, x0, offset	Branch if = zero
bnez rs, offset	bne rs, x0, offset	Branch if \neq zero
blez rs, offset	bge x0, rs, offset	Branch if \leq zero
other similar instructions are cut here		
bgt rs, rt, offset	blt rt, rs, offset	Branch if >
ble rs, rt, offset	bge rt, rs, offset	Branch if \leq
bgtu rs, rt, offset	bltu rt, rs, offset	Branch if >, unsigned
bleu rs, rt, offset	bgeu rt, rs, offset	Branch if \leq , unsigned
j offset	jal x0, offset	Jump
jal offset	jal x1, offset	Jump and link
ret	jalr x0, x1, 0	Return from subroutine
call offset	auipc x1, offset[31:12] jalr x1, x1, offset[11:0]	Call far-away subroutine
fence	fence iorw, iorw	Fence on all memory and I/O