**RISC Virtual Machine**

The RISC machine is deliberately kept simple. There are two classes of instructions, which are separated by bit 31 , the most significant bit. It has a flat memory space of 32 bit words, the size being a power of 2. The first 64 bytes / 16 words are 16 general 32 bit registers R0-R15 (actual addresses 0,4,8,…,60)

There are two flags, Sign and Zero. R15 is the program counter, R14 the link register. This is designed for simplicity not efficiency. It borrows some ideas from my 20 bit RISC for the GI Chip, and some ideas from the ARM chip.

CALL / BRANCH / RETURN

This occurs when bit 31 is ‘1’.

Bits 30,29,28 are the condition bits (see later)

Bit 27 is the link bit. When this is set, R15 is copied to R14 before the branch is performed, but after the fetch, effecting an ARM style subroutine call. (return involves copying R14 to R15). So R14 contains the instruction after this one.

Bits 26…0 are the offset address to branch to, divided by 4, sign extended giving a branch range of +/- 2^26 (nearly). This is added to the current value of R15 (after the fetch, so B 0 will actually branch to the next instruction)

ALU Instruction

This occurs when bit 30 is ‘0’

Bits 30,29,28 are the condition bits (see later)

Bits 25 and 24 are the function code (see later)

Bits 23..16 modify the source value (in the following order)

* When bit 23 is logic ‘1’ the source byte is sign extended to 32 bits and used as a value.
* Bits 20-16 form a 32 bit value, the source value is rotated left this many times (circularly)

Bits 8..15 specify the target address

Bits 0..7 specify the source address

Addresses

The effective address value is the value of 4 x the lower 4 bits if the most significant bit is ‘0’, and the contents of that memory location if the most significant bit is ‘1’ (direct vs indirect). If the indirection is done via R15 it is auto-incremented (allowing MOV R0,@R15 ; WORD 3222)

The exception to this rule occurs when bit 23 of the modifier is ‘1’ then the target then the fetched value is the sign extended value of bits 0..7, giving a short constant range of 0..127

Note all addresses are 32 bit word addresses. 8 bit data will have to be compacted in and out. Addresses should all be on a 4 byte boundary. It is implementation dependent how this is handled (e.g. memory can be implemented as 2^27 longs or 2^29 bytes. The former is probably more efficient but requires addresses to be scaled by four.

Functions

* 00 MOV:Move target to source. No flags change
* 01 ADD:Add source into target. Sets Z and S flags
* 10 AND:And source into target. Sets Z and S flags
* 11 XOR:Xor source into target. Sets Z and S flags.

Condition Codes

* 000 Always execute
* 001 Z:Execute if zero set
* 010 LT:Execute if sign set
* 011 LE:Execute if zero **OR** sign set.
* 100 Never execute
* 101 NZ:Execute if zero clear
* 110 GE:Execute if sign clear
* 111 GT:Execute if sign clear **AND** zero clear

Order

All instructions are conditional, so you cannot do a long data move conditionally, because this will mean if the condition fails it will execute the following data byte. Instructions should be processed in this order.

1. Read the instruction from address R15
2. Increment R15
3. Calculate the source address and read it or source value dependent on bit 23
4. Rotate the read value left according to bits 16-20
5. Calculate the target address
6. Calculate the result
7. Set the flags
8. Store in the target

Reset State

In the reset state, R15 is guaranteed to be 64 for the first instruction. Code should be loaded into 64. (note, this is a byte offset hence 16 x 4 byte registers)

Mnemonics

Branch instructions

* B[L][<condition>] address
* <Instruction>[<condition>] [@]target,[@]source or constant, [ rotate ]
* RET is a synonym for move R15,R14

The constant syntax applies whether it is a short constant in a single word (-128 … 127) or a long constant in the next word. Rotate can be any number, but will be anded with 31 (so you can use -1)

So if you do mov r4,$2A

It will compile the following which uses the short constant form (broken into bytes for clarity)

$00 $80 $04 $2A

If you do move r4,$12A it will compile the following words which are move r4,@r15 ; $12A

$00 $00 $04 $8F $00 $00 $01 $2A

Memory Layout

Initial versions of the header data are compiled inline, taking advantage of the fact that any instruction beginning x100 will never be executed. In the x86 this code will have to be branched over if the preceding code is not a return.

The first byte is the distance back to the head of the previous record in bytes, with C in the top nibble. If there is no previous record, it contains C00000000. So you can go to the previous record by just subtracting the contents of the memory location pointed to anded with 0FFFFFFF.

+0 0100 <shifted right twice address of previous entry. For first entry this is zero.

+4 0100 <first 4 characters of instruction>

+8 0100 <second 4 characters of instruction>

+12 1100 <last 4 characters of instruction>

The words are shifted in so, AT (which is in binary 100 0001 and 101 0011) would be stored as follows:-

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| End+Cond | | | | Unused so all zero. | | | | | | | | | | | | | | A | | | | | | | T | | | | | | |
| C0 | | | | | | | | 00 | | | | | | | | 20 | | | | | | | | D3 | | | | | | | |

Assembler

One element per line. Single pass, reverse patches at end. Spaces are irrelevant except at least one is required between the instruction and its operands.

// anything after a double slash is a comment

:<Definition> Compile inline definition for given phrase. Clear all labels, backpatch first.

.n Label n where n = 0 – 9. .n in a *branch only* mnemonic

*(opcode and operand)*

<Mnemonic> Mnemonic code.

+nn Advance code pointer by nnn words, padding with 0

Ret<Cond> A special case for clarity, equivalent to mov r15,r14

*(copy link register to program counter)*