

Arquitectura de Computadores

Clase 8 - Arquitectura RISC-V

Profesor: Germán Leandro Contreras Sagredo

Objetivos de la clase

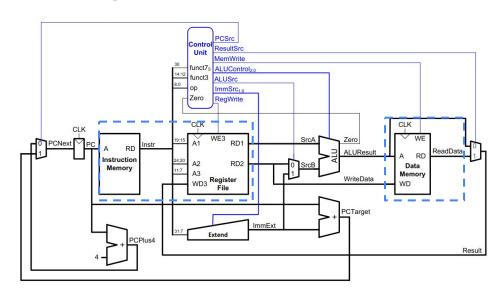
- Conocer la arquitectura RISC-V a nivel general.
- Conocer el set de instrucciones de la ISA RISC-V.
- Conocer las convenciones de llamada utilizadas en RISC-V.
- Ver ejemplos de código en clases a través del simulador RARS.

Arquitectura RISC-V (RISC "five")

- ISA de tipo **RISC** open source. **No es una microarquitectura**.
- Propuesta por la Universidad de Berkeley el 2010 (pero con muchos colaboradores externos a la fecha).
- Arquitectura de tipo load-store. Posee dos categorías de instrucción: acceso de memoria y ALU (solo entre registros).
- Posee un diseño modular, *i.e.* que posee una ISA base y un conjunto de extensiones **opcionales**.

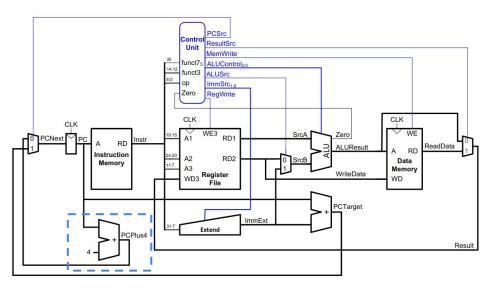
- Register File posee el banco de 16 o 32 registros de la arquitectura.
- Presenta memoria de instrucciones y datos (Harvard).





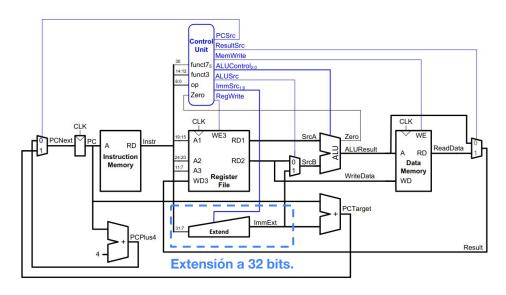
Este diagrama corresponde a una propuesta que permite implementar las instrucciones de la ISA RISC-V con ejecución en un ciclo.

- Palabras de memoria de 8 bits (1 byte).
- Direcciones de memoria de 32 bits (4 bytes).
- Instrucciones de 32 bits (4 bytes). Ocupan 4 direcciones de memoria.



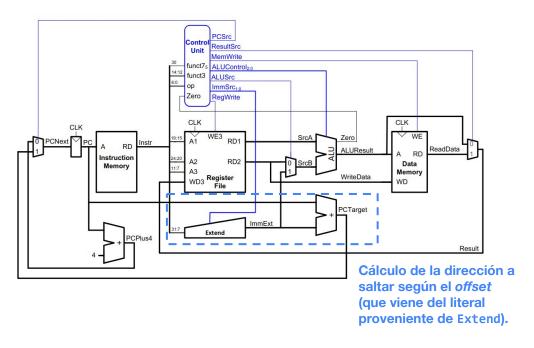
Como las instrucciones ocupan 4 direcciones de la memoria, PC + 4 apunta a la siguiente instrucción.

- Manejo de literales (llamados immediates).
- Como el literal ocupa menos de 32 bits dentro de la instrucción, Extend extiende su tamaño a 32 bits para operar con los valores de los registros. Esta extensión se realiza sobre el bit de signo.



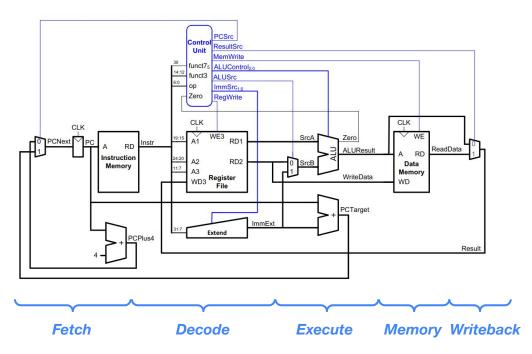
Adicionalmente, el componente Extend recibe un *input* ImmSrc de 2 bits ya que, según el tipo de instrucción, pueden existir hasta cuatro formas de distribuir los bits del literal en ellas.

Las instrucciones de salto, a diferencia de las vistas en el computador básico, cargan la dirección de la instrucción mediante un offset que se suma con el PC actual.



Se separa en 5 etapas:

- Fetch: Obtención de la instrucción.
- Decode: Obtención de señales de control.
- Execute: Ejecución (ALU).
- Memory: Lectura o escritura en memoria.
- Writeback: escritura sobre el banco de registros.

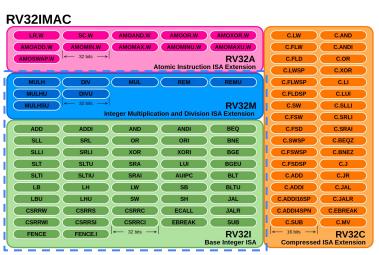


A pesar de la separación en etapas, las instrucciones se ejecutan en un ciclo. Más adelante veremos cómo paralelizar la ejecución de más de una instrucción en nuestra propia arquitectura.

Arquitectura RISC-V

- Se considera una arquitectura Harvard + RISC.
- Usaremos el conjunto base de instrucciones para números enteros (RV32I) y la extensión de multiplicación y división (RV32M).





Conjunto de instrucciones base y extensiones. En nuestro caso, usaremos la ISA RV32IM: RISC-V de 32 bits con extensiones Integer e Integer Multiplication and Division.

Arquitectura RISC-V - Microarquitectura en detalle

- 32 registros de propósito general de 32 bits dentro de un *Register File*.
- La arquitectura varía según los módulos usados. RV32I posee una sola unidad de ejecución (ALU).
- Direcciones de memoria: 32 bits; palabras de memoria: 8 bits. Almacenamiento *little endian*.
- Stack en memoria. Registro SP apunta al último elemento ingresado al stack.

La ISA de la arquitectura RISC-V es un poco más compleja que la del computador básico (a pesar de ser RISC):

- Posee instrucciones de transferencia, aritméticas, lógicas, saltos y subrutinas.
- Posee tipos de instrucción. Cada tipo posee un formato distinto (a diferencia del computador básico donde se tiene solo opcode y literal concatenados para toda instrucción).
- Acepta tipos de datos nativos de 8 y 32 bits (con y sin signo).
- Posee un solo tipo de direccionamiento: indirecto + offset.

Listado de registros

Registro(s)	Mnemotecnia ABI	Descripción
x0	zero	Registro cero. Almacena este valor y no cambia. Ignora las escrituras.
x1	ra	Return Address. Almacena la dirección de retorno de las subrutinas.
x2	sp	Stack Pointer, apunta al último elemento almacenado.
х3	gp	Global Pointer, apunta al segmento de memoria donde se almacenan las variables globales.
x4	tp	Thread Pointer, apunta al segmento de memoria donde se almacenan las variables de un thread para aplicaciones de múltiples threads.
x5-x7, x28-x31	t0-t6	Registros temporales. Pierden su valor entre llamados de subrutinas.
x8-x9, x18-x27	s0-s11	Registros guardados (saved). Preservan su valor entre llamados de subrutinas.
x10-x17	a0-a7	Registros para argumentos de subrutinas.
x10-x11	a0-a1	Si bien son de argumentos de subrutinas, también se utilizan para almacenar valores de retorno.

A nivel de Assembly, usaremos los nombres mnemotécnicos para referirnos a estos registros, respetando la ABI (Application Binary Interface).

Si bien todos son de propósito general, por convención los usaremos según las descripciones de este listado (más información en las siguientes diapositivas).

Listado de directivas de Assembler

Directiva de Assembler	Descripción		
.text	Segmento de texto (código).		
.data	Sección de datos global.		
.bss	Sección de datos globales inicializados en 0.		
.section .foo	Sección llamada "foo".		
.align N	Alinear el siguiente dato/instrucción a 2 ^N bytes de memoria.		
.balign N	Alinear el siguiente dato/instrucción a N bytes de memoria.		
.globl sym	Label sym se vuelve global.		
.string "str"	Almacena el string "str" en memoria.		
.word w1, w2,, wN	Almacena N valores de 32 bits en palabras de memoria sucesivas.		
.byte w1, w2,, wN	Almacena N valores de 8 bits en bytes de memoria sucesivos.		
.space N	Reserva N bytes para almacenar una variable.		
.equ name, constant	Define el símbolo name con valor constant.		
.end	Término del código Assembly.		

Las directivas de un Assembler consisten en órdenes para que este tome ciertas acciones o realice cambios de configuración.

¡No son instrucciones ni se traducen a código de máquina!

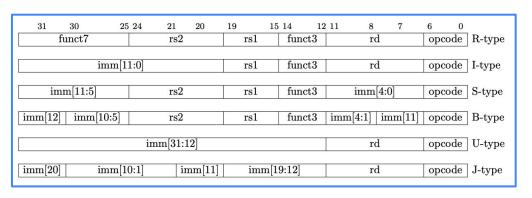
* .align, .balign agregan padding de bytes "basura" en memoria de forma que la inserción de la siguiente variable quede almacenada en una dirección "alineada" (generalmente una dirección par para optimizar direccionamiento).

Tipos de instrucción

- I-Type: *Immediate Type*. Utilizan al menos un literal de operando.
- R-Type: Register Type. Utilizan solo registros como operandos.
- S-Type: Store Type. Almacenamiento en la memoria de datos.
- B-Type: *Branching Type*, variante de S-Type. Saltos condicionales a través de *offsets*.
- U-Type: *Upper Immediate Type*. Permite cargar en registros literales de más de 12 bits (límite de instrucciones I-Type).
- J-Type: Jump Type, variante de U-Type. Saltos incondicionales para las subrutinas.

Formato de tipos de instrucción

- Instrucciones de 32 bits.
- imm = literal (*immediate*).
- rs1, rs2 = registros de operandos.
- rd = registro de destino.
- opcode = identificador de la instrucción.
- funct7, funct3 = Bits adicionales que, en conjunto con el opcode, definen la operación completa a ejecutar.



Resumen de instrucciones - Operaciones aritméticas

Mnemotecnia	Instrucción	Tipo	Descripción
ADD rd, rs1, rs2	Adición	R	rd ← rs1 + rs2
SUB rd, rs1, rs2	Sustracción	R	rd ← rs1 - rs2
ADDI rd, rs1, imm12	Adición de literal	ı	rd ← rs1 + imm12
SLT rd, rs1, rs2	Configurar "menor a"	R	rd ← rs1 < rs2 ? 1 : 0
SLTI rd, rs1, rs2	Configurar "menor a" literal	I	rd ← rs1 < imm12 ? 1 : 0
SLTU rd, rs1, rs2	Configurar "menor a" sin signo	R	rd ← rs1 < rs2 ? 1 : 0
SLTIU rd, rs1, imm12	Configurar "menor a" literal sin signo	I	rd ← rs1 < imm12 ? 1 : 0
LUI rd, imm20	Cargar literal "superior" (20 bits)	U	rd ← imm20 << 12 (SHL 12)
AUIPC rd, imm20	Sumar literal "superior" a PC (20 bits)	U	rd ← PC + imm20 << 12 (SHL 12)

^{*} Para restar un literal, usamos ADDI con un literal negativo.

Resumen de instrucciones - Operaciones lógicas 1/2

Mnemotecnia	Instrucción	Tipo	Descripción
AND rd, rs1, rs2	Operación AND	R	rd ← rs1 & rs2
OR rd, rs1, rs2	Operación OR	R	rd ← rs1 rs2
XOR rd, rs1, rs2	Operación XOR	R	rd ← rs1 ^ rs2
ANDI rd, rs1, imm12	Operación AND con literal	I	rd ← rs1 & imm12
ORI rd, rs1, imm12	Operación OR con literal	I	rd ← rs1 imm12
XORI rd, rs1, imm12	Operación XOR con literal	I	rd ← rs1 ^ imm12

Resumen de instrucciones - Operaciones lógicas 2/2

Mnemotecnia	Instrucción	Tipo	Descripción
SLL rd, rs1, rs2	Operación shift left lógico	R	rd ← rs1 << rs2
SRL rd, rs1, rs2	Operación shift right lógico	R	rd ← rs1 >> rs2
SRA rd, rs1, rs2	Operación shift right aritmético	R	rd ← rs1 >> rs2
SLLI rd, rs1, shamt	Operación shift left lógico con literal	I	rd ← rs1 << shamt
SRLI rd, rs1, shamt	Operación shift right lógico con literal	ı	rd ← rs1 >> shamt
SRAI rd, rs1, shamt	Operación shift right aritmético con literal	ı	rd ← rs1 >> shamt

^{*} shamt o shift amount es la cantidad de shifts a realizar y se codifica como un entero a partir de los 5 bits menos significativos del literal (imm12[4:0]).

Resumen de instrucciones - Operaciones de carga y almacenamiento

Mnemotecnia	Instrucción	Tipo	Descripción
LW rd, imm12(rs1)	Cargar word (32 bits)	ı	rd ← mem[rs1 + imm12]
LH rd, imm12(rs1)	Cargar half word (16 bits)	I	rd ← mem[rs1 + imm12]
LB rd, imm12(rs1)	Cargar byte (8 bits)	I	rd ← mem[rs1 + imm12]
LWU rd, imm12(rs1)	Cargar word sin signo (32 bits)	I	rd ← mem[rs1 + imm12]
LHU rd, imm12(rs1)	Cargar half word sin signo (16 bits)	I	rd ← mem[rs1 + imm12]
LBU rd, imm12(rs1)	Cargar byte sin signo (8 bits)	I	rd ← mem[rs1 + imm12]
SW rs2, imm12(rs1)	Almacenar word (32 bits)	S	$rs2 \rightarrow mem[rs1 + imm12]$
SH rs2, imm12(rs1)	Almacenar half word (16 bits)	S	$rs2(15:0) \rightarrow mem[rs1 + imm12]$
SB rs2, imm12(rs1)	Almacenar byte (8 bits)	S	$rs2(7:0) \rightarrow mem[rs1 + imm12]$

^{*} Para direccionar en LW y SW, se usa el formato offset(x), donde la dirección se almacena en el registro x. Ejemplo: 4(sp)

Resumen de instrucciones - Saltos y subrutinas

Mnemotecnia	Instrucción	Tipo	Descripción
BEQ rs1, rs2, imm12	Salto con condición "igual"	В	if rs1 == rs2: PC ← PC + imm12
BNE rs1, rs2, imm12	Salto con condición "distinto"	В	if rs1 != rs2: PC ← PC + imm12
BGE rs1, rs2, imm12	Salto con condición "mayor o igual"	В	if rs1 >= rs2: PC ← PC + imm12
BGEU rs1, rs2, imm12	Salto con condición "mayor o igual" sin signo	В	if rs1 >= rs2: PC \leftarrow PC + imm12
BLT rs1, rs2, imm12	Salto con condición "menor"	В	if rs1 < rs2: PC ← PC + imm12
BLTU rs1, rs2, imm12	Salto con condición "menor" sin signo	В	if rs1 < rs2: PC \leftarrow PC + imm12
JAL rd, imm20	Salto incondicional con "enlace"	J	rd ← PC+4; PC ← PC + imm20
JALR rd, imm12(rs1)	Salto incondicional con "enlace" a registro	I	rd ← PC+4; PC ← rs1 + imm12

^{*} En estos casos, el literal representa el offset para llegar a la instrucción deseada desde el *Program Counter*. A nivel de código, se observa como el *label* de la dirección a saltar.

Resumen de instrucciones - Pseudo-instrucciones 1/2

Mnemotecnia	Instrucción	Instrucción(es) base
LI rd, imm12	Cargar literal en registro que utiliza ≤ 12 bits	ADDI rd, zero, imm12
LI rd, imm	Cargar literal en registro que utiliza > 12 bits	LUI rd, imm[31:12]; ADDI rd, rd, imm[11:0]
LA rd, sym	Cargar dirección en registro	AUIPC rd, sym[31:12]; ADDI rd, rd, sym[11:0]
MV rd, rs	Copiar registro	ADDI rd, rs, 0
NOT rd, rs	Complemento de 1	XORI rd, rs, -1
NEG rd, rs	Complemento de 2	SUB rd, zero, rs
BGT rs1, rs2, offset	Salto si rs1 > rs2	BLT rs2, rs1, offset
BLE rs1, rs2, offset	Salto si rs1 ≤ rs2	BGE rs2, rs1, offset
BGTU rs1, rs2, offset	Salto si rs1 > rs2 sin signo	BLTU rs2, rs1, offset
BLEU rs1, rs2, offset	Salto si rs1 ≤ rs2 sin signo	BGEU rs2, rs1, offset

^{*} Las pseudo-instrucciones mnemotécnicas se traducen a las instrucciones reales de RISC-V. Esto ayuda a tener instrucciones de operaciones útiles en un lenguaje más sencillo de entender.

Resumen de instrucciones - Pseudo-instrucciones 2/2

Mnemotecnia	Instrucción	Instrucción(es) base
BEQZ rs1, offset	Salto si rs1 = 0	BEQ rs1, zero, offset
BNEZ rs1, offset	Salto si rs1 ≠ 0	BNE rs1, zero, offset
BGEZ rs1, offset	Salto si rs1 ≥ 0	BGE rs1, zero, offset
BLEZ rs1, offset	Salto si rs1 ≤ 0	BGE zero, rs1, offset
BGTZ rs1, offset	Salto si rs1 > 0	BLT zero, rs1, offset
J offset	Salto incondicional	JAL zero, offset
CALL offset12	Llamado a subrutina (dirección ≤ 12 bits)	JALR ra, ra, offset12
CALL offset*	Llamado a subrutina (dirección > 12 bits)	AUIPC ra, offset[31:12]; JALR ra, ra, offset[11:0]
RET	Retorno de la subrutina	JALR zero, 0(ra)
NOP	No se realiza ninguna operación	ADDI zero, zero, 0

^{*} Esta es la pseudo-instrucción que ocupa el emulador a utilizar, RARS.

Ejemplo de código - Multiplicación

```
a = 10
b = 200
res = 0
while (a > 0):
    res += b
    a -= 1
print(res)
```

Programa en pseudocódigo (Python).

Programa en RISC-V. Notar el uso de pseudo-instrucciones para facilitar la lectura del código.

Resumen de instrucciones - Extensión M

Mnemotecnia	Instrucción	Tipo	Descripción
MUL rd, rs1, rs2	32 bits menos significativos del producto.	R	rd ← (rs1 * rs2)[31:0]
MULH rd, rs1, rs2	32 bits más significativos del producto (rs1, rs2 con signo).	R	rd ← (rs1 * rs2)[63:32]
MULHSU rd, rs1, rs2	32 bits más significativos del producto (rs1 con signo, rs2 sin signo).	R	rd ← (rs1 * rs2)[63:32]
MULHU rd, rs1, rs2	32 bits más significativos del producto (rs1, rs2 sin signo).	R	rd ← (rs1 * rs2)[63:32]
DIV rd, rs1, rs2	División con signo	R	rd ← rs1 / rs2
DIVU rd, rs1, rs2	División sin signo	R	rd ← rs1 / rs2
REM rd, rs1, rs2	Resto con signo	R	rd ← rs1 % rs2
REMU rd, rs1, rs2	Resto sin signo	R	rd ← rs1 % rs2

Ejemplo de código - Multiplicación con extensión M

```
.globl main

.text

main:

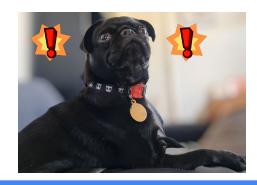
    li t0, 10  # t0 = 10

    li t1, 200  # t1 = 200

    mul t2, t0, t1  # t2 = t0 * t1 = 2000

.end
```

Si agregamos la extensión RV31M al conjunto de instrucciones base RV31I, entonces el código se simplifica de manera significativa (similar a lo que ocurre con las operaciones MUL y DIV de la arquitectura x86).



Instrucción ECALL (Environment Call)

- Instrucción especial de tipo I que permite realizar una solicitud al entorno de ejecución (sistema operativo).
- Su uso varía según la implementación, pero en este curso nos basaremos en la implementación del simulador de RISC-V RARS (RISC-V Assembler And Runtime System).

Instrucción ECALL (Environment Call)

- Se hace uso de un código que indica la llamada a efectuar. Este se almacena en el registro a7 y los argumentos en a0-a6.
- Existen <u>múltiples funciones</u>, pero usaremos dos:
 - PrintInt(code=1): Impresión de valor entero en consola.
 Solo recibe un argumento en a0, el número a imprimir.
 - Exit(code=10): Término de programa. No recibe argumentos y asume que el código de término es 0 (sin errores).

Ejemplo de código - Impresión de número y salida del programa

```
Messages Run I/O

11
— program is finished running (0) —

Clear
```

Ejemplo de programa que imprime en consola el número 11. Se incluye además el resultado del simulador RARS. Cabe destacar que RARS no reconoce la directiva .end, por lo que es necesario usar ecall para el término del programa.

Ejemplo de código - Multiplicación con extensión M en RARS

```
Messages Kun I/O

2000
— program is finished running (0) —

Clear
```

Ejemplo de código de multiplicación ejecutable en RARS.

Se hace uso de convenciones de llamada para estandarizar la forma en la que se ejecutan las subrutinas en un programa.

En el caso de RISC-V, se hace uso de una convención propia donde el almacenamiento de parámetros, valores de retorno y dirección de retorno recae en los registros y no en el *stack*. Este último se utiliza como respaldo para preservar el valor de los registros entre llamadas.

- El registro ra (x1) almacena la dirección de retorno.
- Los argumentos de una subrutina se almacenan en los registros a0-a7 (x10-x17).
- Si existen valores de retorno, se almacenan en los registros a0-a1 (x10-x11).
- Se hace uso del stack con el registro sp (x2). Se puede utilizar para incluir argumentos adicionales, variables locales y respaldo de registros.

El encargado de respaldar el valor de los registros con el llamado de una subrutina varía según el tipo de registro. En este caso, el encargado puede ser quien llama a la subrutina (caller) o la subrutina llamada (callee).

El respaldo de los registros se realiza a través del *stack pointer*, modificando su valor para reservar valores en la memoria de *stack*.

Registro(s)	Mnemotecnia ABI	Encargado del respaldo
x0	zero	-
x1	ra	Caller
x2	sp	Callee
х3	gp	-
x4	tp	-
x5-x7, x28-x31	t0-t6	Caller
x8-x9, x18-x27	s0-s11	Callee
x10-x17	a0-a7	Caller
x10-x11	a0-a1	Caller

Ejemplo de código - Duplicación de elementos de un arreglo

```
len: .word 5
   arr: .word 198, 137, 42, 63, 175
   start:
       li s0, 4
                                           # s0 = 4 bytes por dirección
       lw s1, len
                                           # s1 = largo del arreglo
       li t0, 0
                                           # Contador (i)
       la t1, arr
                                           # t1 = dirección del arreglo (inicialmente arr[0])
                                           # t2 = s0 * s1 = bytes que ocupa el arreglo de entrada
       mul t2, s0, s1
       add t2, t2, t1
                                           # t2 += t1 = primera dirección que podemos usar de .data
       while:
           lw a0, 0(t1)
                                           # a0 = arr[i]
           addi sp, sp, -12
           sw t0, 0(sp)
                                           # Respaldamos t0, t1 y ra (caller-saved). t1 se respalda
Preámbulo
           sw t1, 4(sp)
                                           # aunque no se use porque call lo modifica con la
           sw ra, 8(sp)
                                           # dirección de retorno en RARS
           call double_give_next_person
           lw t0, 0(sp)
                                           # Recuperamos t0, t1 y ra y restauramos el stack
           lw t1, 4(sp)
Prólogo
           lw ra, 8(sp)
           addi sp, sp, 12
                                           # out[i] = a0
           sw a0, 0(t2)
           addi t0, t0, 1
           beq t0, s1, end
                                           # Termina cuando se recorre todo el arreglo
           add t1, t1, s0
                                           # t1 += s0 = dirección arr[i+1]
                                           # t2 += s0 = dirección out[i+1]
           add t2, t2, s0
           j while
   double_give_next_person:
       mv t0, a0
       add a0, a0, t0
                                           # a0 = a0 + t0 = 2 * a0
       ret
       li a7, 10
       ecall
```

Programa en RISC-V que duplica los elementos de un arreglo y los almacena en las direcciones posteriores a dicha variable. El respaldo de registros *caller-saved* en la convención de llamada se realiza de la siguiente forma:

- 1. Los registros se respaldan en el stack antes de la llamada. Para ello, se desplaza el registro sp según la cantidad de bytes a respaldar (4 bytes por registro) y luego se almacenan a través de direccionamiento indirecto por registro sp con offset. Esto se conoce como el preámbulo.
- Finalizada la llamada, se restauran los registros desde el stack y se restablece el valor de sp. Esto se conoce como el prólogo.



Resultado de la ejecución en el segmento de datos en RARS.

Ejemplo de código - Duplicación de elementos de un arreglo

```
len: .word 5
arr: .word 198, 137, 42, 63, 175
start:
    li s0, 4
                                        # s0 = 4 bytes por dirección
    lw s1, len
                                        # s1 = largo del arreglo
    li t0, 0
                                        # Contador (i)
    la t1, arr
                                        # t1 = dirección del arreglo (inicialmente arr[0])
                                        # t2 = s0 * s1 = bytes que ocupa el arreglo de entrada
    mul t2, s0, s1
    add t2, t2, t1
                                        # t2 += t1 = primera dirección que podemos usar de .data
    while:
        lw a0, 0(t1)
                                        # a0 = arr[i]
        addi sp, sp, -12
        sw t0, 0(sp)
                                        # Respaldamos t0, t1 y ra (caller-saved). t1 se respalda
        sw t1, 4(sp)
                                        # aunque no se use porque call lo modifica con la
        sw ra, 8(sp)
                                        # dirección de retorno en RARS
        call double_give_next_person
        lw t0, 0(sp)
                                        # Recuperamos t0, t1 y ra y restauramos el stack
        lw t1, 4(sp)
        lw ra, 8(sp)
        addi sp, sp, 12
                                        # out[i] = a0
        sw a0, 0(t2)
        addi t0, t0, 1
        beq t0, s1, end
                                        # Termina cuando se recorre todo el arreglo
        add t1, t1, s0
                                        # t1 += s0 = dirección arr[i+1]
        add t2, t2, s0
                                        # t2 += s0 = dirección out[i+1]
        j while
double_give_next_person:
    mv t0, a0
    add a0, a0, t0
                                        # a0 = a0 + t0 = 2 * a0
    ret
    li a7, 10
    ecall
```

En este mismo ejemplo, es importante destacar que el registro ra podría no respaldarse y que se asegure el funcionamiento correcto del programa. No obstante, por convención es buena práctica hacerlo de todas formas ya que el programa podría ser ejecutado mediante una llamada desde otro archivo. Si no se respaldara ra, se perdería la dirección de retorno para el programa que ejecuta este ejemplo. El no respaldo de ra generará problemas siempre que haya llamados de subrutinas anidados.

Por otra parte, t1 se respalda solo porque la instrucción call de RARS modifica su valor con la dirección de retorno, en estricto rigor el respaldo no es necesario si no se utiliza dentro de la subrutina.

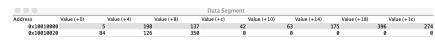
Ejemplo de código - Duplicación de elementos de un arreglo

```
len: .word 5
    arr: .word 198, 137, 42, 63, 175
   start:
        li t0, 4
                                            # s0 = 4 bytes por dirección
        lw t2, len
                                            # s1 = largo del arreglo
        li s0, 0
        la s1, arr
                                            # t1 = dirección del arreglo (inicialmente arr[0])
                                            # t2 = s0 * s1 = bytes que ocupa el arreglo de entrada
       mul s2, t0, t2
        add s2, s2, s1
                                            # t2 += t1 = primera dirección que podemos usar de .data
        while:
                                            # a0 = arr[i]
            lw a0, 0(s1)
            addi sp, sp, -4
Preámbulo
            sw ra, 0(sp)
                                            # Respaldamos ra (caller-saved)
            call double give next person
            lw ra, 0(sp)
                                            # Recuperamos ra y restauramos el stack
            addi sp, sp, 4
            sw a0, 0(s2)
                                            # out[i] = a0
            addi s0, s0, 1
            beg s0, t2, end
                                            # Termina cuando se recorre todo el arreglo
            add s1, s1, t0
                                            # s1 += t0 = dirección arr[i+1]
                                            # s2 += t0 = dirección out[i+1]
            add s2, s2, t0
            i while
    double give next person:
       'addi sp, sp, -4
       sw s0, 0(sp)
                                 # Respaldamos s0 (callee-saved)
        mv s0, a0
        add a0, a0, s0
                                 # a0 = a0 + s0 = 2 * a0
Prólogo ∫ lw s0, 0(sp)
                                 # Recuperamos s0 y restauramos el stack
        addi sp, sp, 4
        ret
        li a7, 10
        ecall
```

Veamos el mismo ejemplo pero invirtiendo los registros s* por registros t*. El respaldo de registros callee-saved en la convención de llamada se realiza de la siguiente forma:

- Los registros se respaldan en el stack al comienzo de la llamada al igual que con los registros caller-saved. Esto también es un preámbulo.
- Antes de finalizar la llamada con ret, se restauran los registros desde el stack y se restablece el valor de sp al igual que con los registros caller-saved. Esto también es un prólogo.

En este caso seguimos respaldando el registro ra por ser caller-saved y por lo comentado en la diapositiva anterior.



Resultado de la ejecución en el segmento de datos en RARS.

Ejemplo de código - Factorial, función recursiva

```
.data
   N: .word 4
                               # N = Argumento de factorial. Calcularemos N! = 4
.text
    main:
       addi sp, sp, -4
                               # Reservamos 4 bytes en el stack
       sw ra, 0(sp)
                               # Respaldamos ra
       la t0, N
                               # Dirección de memoria de N
       lw t0, 0(t0)
                               # Valor de N
       add a0, zero, t0
                               # Argumento 0 = valor de N
       call factorial
                               # factorial(N)
       li a7, 1
                               # Llamada de sistema: print int
       ecall
                               # Valor en consola: 24 (a0, valor de retorno)
       lw ra, 0(sp)
                               # Restauramos ra
       addi sp, sp, 4
                               # Restauramos el stack
       li a7, 10
                               # Llamada de sistema: exit
       ecall
    factorial:
       addi sp, sp, -8
                               # Reservamos 8 bytes en el stack
       sw ra, 0(sp)
                               # Respaldamos ra
       sw a0, 4(sp)
                               # Respaldamos N
       blez a0, factorial zero # if (N > 0){
       addi a0, a0, -1
       call factorial
                               # (N-1)! = factorial(N-1)
       lw t0, 4(sp)
                               # Recuperamos N
                               # N! = N * (N-1)! = N * factorial(N-1)
       mul a0, a0, t0
       factorial end
                               # }
       factorial zero:
                               # else {
           li a0, 1
                               # N! = 1
       factorial_end:
           lw ra, 0(sp)
                               # Restauramos solo ra, a0 ahora posee el retorno
            addi sp, sp, 8
                               # Restauramos el stack
       ret
```

Programa en RISC-V que calcula el valor del factorial de la variable *N*.

Ahora, el callee también actúa como caller por la recursión y respalda tanto el registro ra como el registro de argumento a0. Además, el callee se encarga de restaurar el registro sp en cada llamada, lo que asegura la consistencia de su valor al realizar el último retorno. En este caso no se restaura a0 ya que este posee el valor de retorno a ser utilizado en la impresión de la consola.

Si se hubieran utilizado registros s*, también habrían sido respaldados por el *callee* como en el ejemplo anterior.

```
Messages Run I/O

24
— program is finished running (0) —

Clear
```

Arquitectura RISC-V - Resumen de la convención de llamada

- Si usamos los registros t* tanto fuera como dentro de una subrutina, los respaldamos en el *stack* antes del llamado y los restauramos posterior a este.
- Si usamos los registros s* tanto fuera como dentro de una subrutina, los respaldamos en el stack al principio de la subrutina y los recuperamos justo antes del retorno.
- Si usamos los registros a*, los respaldamos fuera de la subrutina pero los recuperamos siempre que no interfieran con el valor de retorno (a2-a7).
- Por seguridad, siempre respaldamos el registro ra antes de un llamado.

Arquitectura RISC-V - Manejo de floats

Además de las instrucciones y registros anteriores, existe la extensión **RV32F** que otorga soporte operaciones sobre números de punto flotante bajo el estándar IEEE754. A continuación, se listan los 32 registros que se añaden con esta extensión.

Registro(s)	Mnemotecnia ABI	Descripción	
f0-7, f28-31	ft0-7, ft8-11	Registros temporales de punto flotante. Pierden su valor entre llamados de subrutinas.	
f8-9, f18-27	fs0-1, fs2-11	Registros guardados de punto flotante (saved). Preservan su valor entre llamados de subrutinas.	
f10-17	fa0-7	Registros de punto flotante para argumentos de subrutinas.	
f10-11	fa0-7	Si bien son de argumentos de subrutinas, también se utilizan para almacenar valores de retorno.	

^{*} Se consideran caller-saved o callee-saved bajo los mismos criterios que los registros de números enteros.

Arquitectura RISC-V - Manejo de floats

A continuación, se listan las instrucciones **más relevantes** que añade la extensión RV32F.

Mnemotecnia	Instrucción	Tipo	Descripción
FLW rd, imm12(rs1)	Cargar word de tipo float (32 bits)	1	rd ← mem[rs1 + imm12]
FSW rs2, imm12(rs1)	Almacenar word de tipo float (32 bits)	S	$rs2 \rightarrow mem[rs1 + imm12]$
FADD.S rd, rs1, rs2	Adición de floats	R	rd ← rs1 + rs2
FSUB.S rd, rs1, rs2	Sustracción de floats	R	rd ← rs1 - rs2
FMUL.S rd, rs1, rs2	Multiplicación de floats	R	rd ← rs1 * rs2
FDIV.S rd, rs1, rs2	División de floats	R	rd ← rs1 / rs2
FSQRT.S rd, rs1	Raíz cuadrada de float	R	rd ← √rs1
FCVT.S.W rd, rs1	Convierte int a float	R	rd ← float(rs1)
FCVT.W.S rd, rs1	Convierte float a int	R	rd ← int(rs1)

Arquitectura RISC-V - Manejo de floats

Ejemplo: Cómputo de la raíz cuadrada de un número *N*. Si el valor posee decimales, se truncan al transformar el número de vuelta a entero.

```
.data
 N:
         word 25
 N trunc: .word 37
.text
 main:
   lw t0, N
           # t0 = N
   fcvt.s.w ft0, t0 # ft0 = float(t0)
   fsqrt.s ft0, ft0 # ft0 = sqrt 2(ft0)
   fcvt.w.s t0, ft0 # t0 = int(t0)
   mv a0, t0 # a0 = t0
   li a7, 1  # print(a0)
   ecall
   lw t0, N trunc # t0 = N trunc
   fcvt.s.w ft0, t0 # ft0 = float(t0)
   fsqrt.s ft0, ft0 # ft0 = sqrt 2(ft0)
   fcvt.w.s t0, ft0 # t0 = int(t0)
   mv a0, t0 # a0 = t0
  ecall # print(a0)
  li a7, 10 # exit
   ecall
```

Arquitectura RISC-V - Cheat sheets

A continuación, se dejan dos *cheat sheets* útiles que resumen todas las instrucciones que pueden utilizar en RISC-V:

- Cheat Sheet 1
- Cheat Sheet 2



Actividad en clase - Instalación de RARS

Antes de hacer ejercicios, realizaremos la instalación de RARS para poder ejecutar código RISC-V en nuestros computadores.

Para ello, es necesario instalar Java 8 o superior y seguir las instrucciones del <u>repositorio principal</u>. Asegúrese de instalar: (1) el último lanzamiento estable (detallado en el README); y (2) de **instalar el archivo . jar adjunto, no clonar el repositorio**.



Actividad en clase - Instalación de Java 8

■ Windows: Descargar Java 8 en <u>el siguiente enlace</u> y ejecutar el archivo .jar.

■ Mac

- Chip M1: Seguir el siguiente tutorial.
- Modelos previos: Descargar Java 8 en el siguiente enlace y ejecutar el archivo .jar (con click derecho por seguridad).
- **Ubuntu:** Seguir <u>el siguiente tutorial</u>.



^{*} Pueden instalar versiones superiores de Java, la versión 8 es la mínima requerida.

Actividad en clase - Ejercicio de Assembly RISC-V

Ejercicio en clases - Parte 1

Desarrolle un programa en RISC-V que determine si un *input N* definido en el segmento .data es primo o no. El programa debe imprimir un 1 si lo es y un 0 en caso contrario. **Debe seguir la convención de llamada de RISC-V**.

Actividad en clase - Ejercicio de Assembly RISC-V

Ejercicio en clases - Parte 2

Extienda el programa anterior para que el programa imprima 1 si el input *N* es **primo gemelo**, *i.e.*:

- Es primo.
- Existe otro número primo M tal que |N M| = 2.

Debe seguir la convención de llamada de RISC-V.

Arquitectura RISC-V - Consideraciones respecto a RARS

■ En teoría, RISC-V no otorga forma de cargar los valores de memoria de variables directamente en registros, pero RARS sí lo permite.

```
.data
N: .word 23
.text
# Comportamiento esperado
main:
la a0, N # a0 = dirección N
lw a0, 0(a0) # a0 = Mem[a0] = N
```

Al ejecutar la pseudo-instrucción call, RARS almacena los 24 bits más significativos del literal offset en el registro t1. Por lo tanto, al definir subrutinas recursivas se sugiere no operar con este registro.

```
.text

call example

mv a0, t1

li a7, 1

ecall

li a7, 10

ecall

example:

ret
```

```
Messages Run I/O

4194304
— program is finished running (0) —

Clear
```

Ahora, veremos algunos ejercicios.

Estos se basan en preguntas de tareas y pruebas de semestres anteriores, por lo que nos servirán de preparación para las evaluaciones.



Indique justificadamente si el siguiente fragmento de código respeta la convención de llamada de RISC-V.

```
.data
           .word 23
 N:
 divisor: .word 12
.text
 main:
   addi sp, sp, -4
   sw ra, 0(sp)
   call remainder
   lw ra, 0(sp)
   addi sp, sp, 4
   j end
 remainder:
   lw a0, N
   lw a1, divisor
   rem a0, a0, a1
   ret
 end:
```

Examen, 2023-1

Indique si el siguiente fragmento de código en RISC-V termina su ejecución. Si lo hace, indique el valor del registro a0. En otro caso, justifique por qué no termina.

```
N: .word 4
main:
    addi sp, sp, -4
    sw ra, 0(sp)
    lw t0, N
    call factorial
   lw ra, 0(sp)
    addi sp, sp, 4
   i end
factorial:
    addi sp, sp, -4
    sw a0, 0(sp)
   blez a0, factorial_zero
    call factorial
        t0, 0(sp)
    mul a0, a0, t0
   i factorial end
    factorial zero:
       li a0, 1
    factorial end:
        addi sp, sp, 4
end:
```

Examen, 2023-1

Para esta pregunta, deberá programar una pequeña versión de juego FizzBuzz en RISC-V. Para esto, recibirá en la sección .data una etiqueta N, que corresponde a un entero positivo y deberá escribir, a partir de la etiqueta out, todos los números entre 0 y N como .word, con la salvedad de que si el número es divisible por 3, este se reemplaza por 70 (ASCII "F"); en caso de ser divisible por 5, se reemplaza por 66 (ASCII "B"); y si es divisible por ambos, se reemplaza por 7066. Además, si un número es primo gemelo, deberá reemplazarlo por 80 (ASCII "P"). A modo de ejemplo, si N = 15 la secuencia que se guardará en out será: 7066, 1, 2, 80, 4, 80, 70, 80, 8, 70, 66, 80, 70, 80, 14, 7066.

Tarea 3, 2023-1

Antes de terminar

¿Dudas?

¿Consultas?

¿Inquietudes?

¿Comentarios?





Arquitectura de Computadores

Clase 8 - Arquitectura RISC-V

Profesor: Germán Leandro Contreras Sagredo

Anexo - Resolución de ejercicios

ilmportante!

Estos ejercicios pueden tener más de un desarrollo correcto. Las respuestas a continuación no son más que soluciones que **no excluyen** otras alternativas igual de correctas.



Ejercicio en clases - Parte 1

Desarrolle un programa en RISC-V que determine si un *input N* definido en el segmento .data es primo o no. El programa debe imprimir un 1 si lo es y un 0 en caso contrario. **Debe seguir la convención de llamada de RISC-V**.

Respuesta en la siguiente diapositiva.

```
.data
           .word 17
 N:
.text
 main:
                                              \# a0 = N = parameter
   lw a0, N
   li s0, 1
                                              \# Constant s0 = 1
   addi sp, sp, -4
                                              # Save ra
   sw ra, 0(sp)
   call check is prime
   lw ra, 0(sp)
                                              # Restore ra and stack
   addi sp, sp, 4
   li a7, 1
                                              # print
   ecall
   li a7, 10
                                              # exit
   ecall
  check is prime:
   ble a0, s0, is not prime
                                              # a0 <= 1 -> is not prime
   # Check divisors from 2 to sqrt 2(N).
   fcvt.s.w ft0, a0
   fsqrt.s ft0, ft0
   fcvt.w.s t1, ft0
   li t0, 2
                                              # Starting divisor. t0 = 2
   division loop:
     rem t2, a0, t0
                                              # If a0 % t0 == 0 -> Divisible, not prime.
     beqz t2, is_not_prime
     addi t0, t0, 1
                                              # t0 += 1
     ble t0, t1, division loop
                                              # While t0 <= t1, keep looking for divisors.
   is prime:
     li a0, 1
     i end
   is_not_prime:
     li a0, 0
   end:
      ret
```

Ejercicio en clases - Parte 2

Extienda el programa anterior para que el programa imprima 1 si el input *N* es **primo gemelo**, *i.e.*:

- Es primo.
- Existe otro número primo M tal que |N M| = 2.

Debe seguir la convención de llamada de RISC-V.

Respuesta en la siguiente diapositiva.

```
.data
          .word 19
 N:
.text
 main:
                                           \# a0 = N = parameter
   lw a0, N
   li s0, 1
                                           # Constant s0 = 1
   addi sp, sp, -4
                                           # Save ra
   sw ra, 0(sp)
   call check_is_twin_prime
   lw ra, 0(sp)
                                           # Restore ra and stack
   addi sp, sp, 4
   li a7, 1
                                           # print
   ecall
   li a7, 10
                                           # exit
   ecal1
 check_is_twin_prime:
                                           # Save ra and a0
   addi sp, sp, -8
   sw ra, 0(sp)
   sw a0, 4(sp)
                                           # Check if N is prime
   call check is prime
   bne a0, s0, is not twin prime
                                           # a0 != 1 -> is not twin prime
   lw a0, 4(sp)
                                           # Restore a0
   addi a0, a0, -2
   call check_is_prime
                                           # Check if N - 2 is prime
   beg a0, s0, is twin prime
                                           # If N - 2 is prime -> N is twin prime
   lw a0, 4(sp)
                                           # Restore a0
   addi a0, a0, 2
                                           # Check if N + 2 is prime
   call check is prime
   beq a0, s0, is_twin_prime
                                           # If N + 2 is prime -> N is twin prime
   # Else: N is not twin prime
   is not twin prime:
     li a0, 0
     i end
   is twin prime:
     li a0, 1
   end:
     lw ra, 0(sp)
                                           # Restore ra and stack
     addi sp, sp, 8
```

```
check_is_prime:
 ble a0, s0, is not prime
                                         # a0 <= 1 -> is not prime
 # Check divisors from 2 to sart 2(N).
 fcvt.s.w ft0, a0
 fsqrt.s ft0, ft0
 fcvt.w.s t1, ft0
 li t0, 2
                                         # Starting divisor. t0 = 2
 division_loop:
   rem t2, a0, t0
                                         # If a0 % t0 == 0 -> Divisible, not prime.
   beqz t2, is_not_prime
   addi t0, t0, 1
                                        # t0 += 1
   ble t0, t1, division_loop
                                        # While t0 <= t1, keep looking for divisors.
 is prime:
   li a0, 1
   j end_check_is_prime
 is_not_prime:
   li a0, 0
 end_check_is_prime:
   ret
```

Indique justificadamente si el siguiente fragmento de código respeta la convención de llamada de RISC-V.

```
.data
 N:
           .word 23
 divisor: .word 12
.text
 main:
   addi sp, sp, -4
   sw ra, 0(sp)
   call remainder
   lw ra, 0(sp)
   addi sp, sp, 4
   j end
 remainder:
   lw a0, N
   lw a1, divisor
   rem a0, a0, a1
   ret
 end:
```

Respuesta en la siguiente diapositiva.

El fragmento de código **no respeta** la convención de RISC-V ya que los registros a0, a1 se usan como argumentos de la subrutinas, por lo que deberían cargar sus valores **antes** del llamado de la subrutina y no dentro de ella.

Indique si el siguiente fragmento de código en RISC-V termina su ejecución. Si lo hace, indique el valor del registro a0. En otro caso, justifique por qué no termina.

```
N: .word 4
    addi sp, sp, -4
    sw ra, 0(sp)
    lw t0, N
    call factorial
   lw ra, 0(sp)
    addi sp, sp, 4
   i end
factorial:
    addi sp, sp, -4
    sw a0, 0(sp)
   blez a0, factorial_zero
        t0, 0(sp)
    mul a0, a0, t0
   i factorial end
    factorial zero:
       li a0, 1
    factorial end:
        addi sp, sp, 4
end:
```

Respuesta en la siguiente diapositiva.

El programa anterior **no termina** su ejecución ya que posee llamadas recursivas que **no respaldan** la dirección de retorno del registro ra. Por dicho motivo, el programa podría no terminar o arrojar errores.

Para esta pregunta, deberá programar una pequeña versión de juego FizzBuzz en RISC-V. Para esto, recibirá en la sección .data una etiqueta N, que corresponde a un entero positivo y deberá escribir, a partir de la etiqueta out, todos los números entre 0 y N como .word, con la salvedad de que si el número es divisible por 3, este se reemplaza por 70 (ASCII "F"); en caso de ser divisible por 5, se reemplaza por 66 (ASCII "B"); y si es divisible por ambos, se reemplaza por 7066. Además, si un número es primo gemelo, deberá reemplazarlo por 80 (ASCII "P"). A modo de ejemplo, si N = 15 la secuencia que se guardará en out será: 7066, 1, 2, 80, 4, 80, 70, 80, 8, 70, 66, 80, 70, 80, 14, 7066.

Respuesta en la siguiente diapositiva.

```
.globl start
    N: .word 15
.text
    start:
       lw s1. N
                                       # max number
       li s0, 0
                                       # counter
       li s2, 3
                                       # first divisor
       li s3, 5
                                       # second divisor
       while:
            call is twin prime
                                       # is_twin_prime(a0 = counter). No need to save t* registers
            bnez a0, twin prime
            rem t0, s0, s2
                                       # t0 = s0 % 3 = remainder of division by 3
                                       # t1 = s0 % 5 = remainder of division by 5
            rem t1, s0, s3
                                       # t0 == t1 -> s0 divisible by 3 and 5? -> FizzBuzz
            beq t1, t0, fizzbuzz
            beq t0, zero, fizz
                                       # t0 == 0 -> s0 divisible by 3 -> Fizz
            beg t1, zero, buzz
                                       # t1 == 0 -> s0 divisible by 5 -> Buzz
            fizzbuzz:
                bnez t1, skip
                                       # t1 != 0 -> s0 not divisible by 3 nor 5
               li a0, 7066
               i continue
            fizz:
               li a0, 70
                i continue
            buzz:
               li a0, 66
               i continue
            twin prime:
                li a0, 80
               i continue
            skip:
            continue:
               mv a1, s0
               call store
                                       # store(a0 = value to store, a1 = counter)
                addi s0, s0, 1
                bgt s0, s1, end
                                       # If s0 > s1 -> counter > N, array is finished
                i while
    store:
       addi sp, sp, -8
                                       # s* registers are callee-saved
       sw s0, 0(sp)
       sw s1, 4(sp)
                                       # s0 = N variable address
       la s0. N
        addi s0, s0, 4
                                       # s0 += 4 for first available array address
                                       # s1 = 4 (bytes stored for each array element)
       li s1, 4
       mul s1, s1, a1
```

```
add s0, s1, s0
                                   # Stored value address = Address(N)+4+4*counter
    sw a0, 0(sp)
                                   # Value stored
    lw s0, 0(sp)
                                   # s* registers restored
    lw s1, 4(sp)
    addi sp, sp, 8
end:
    li a7, 10
    ecall
is twin prime:
                                   # a0 = number to check. a0=1 if twin, a0=0 elsewhen
                                   # We store ra for the final return.
    addi sp. sp. -4
    sw ra, 0(sp)
    mv a1, a0
    call is prime
                                   # is_prime(a0=counter). a0=1 if prime, a0=0 elsewhen
                                   # a0 == 0 if counter is not prime
    begz a0, end is twin prime
    addi a0, a1, -2
    call is prime
                                   # is prime(a0 - 2)
    bgtz a0, end is twin prime
                                   # a0 > 0 if counter is twin prime
    addi a0, a1, 2
    call is prime
                                   # is prime(a0 + 2)
    end_is_twin_prime:
                                   # a0 == 1 if N is twin prime else a0 == 0
                                   # We restore ra for correct return
       lw ra, 0(sp)
       addi sp, sp, 4
    is prime:
       li t0, 1
       ble a0, t0, not prime
                                   # Edge case for counter <= 1
       fcvt.s.w ft0, a0
                                   # ft0 = float(a0)
       fsart.s ft0, ft0
                                   # ft0 = sart 2(ft0)
       fcvt.w.s t0, ft0
                                   # t0 = int(ft0)
       li t1, 2
        while_is_prime_check:
            rem t2, a0, t1
                                   # if a0 % t1 == 0 -> counter not prime
            beqz t2, not_prime
            addi t1, t1, 1
            ble t1, t0, while_is_prime_check
    prime:
       li a0, 1
       ret
    not prime:
       li a0, 0
       ret
```