National Yang Ming Chiao Tung University
Computer Architecture & System Lab

Lab1: RISC-V Assembly Language Programming

CS10014 Computer Organization

Tsung Tai Yeh
Department of Computer Science
National Yang Ming Chiao University

Acknowledgements and Disclaimer

- Slides were developed in the reference with
  - RISC-V Programming
    - https://riscv-programming.org/book/riscv-book.html
  - CENG3420, CUHK
    - https://www.cse.cuhk.edu.hk/~byu/CENG3420/2022Spring/slides/lab1-1.pdf

Outline

- RISC-V Assembly Programming
  - Labels
  - Symbols
  - Directives
- RARS RISC-V Assembly Simulator

Generating native programs

- A native program is a program
  - Encoded using instructions that can be directly executed by the computer hardware

C Code

```
int main ()
{
   int r = func (10);
   return r+1;
}
```

RISC-V assembly code

```
.text
.align 2
main:
   addi sp, sp, -16
   li   a0, 10
   sw   ra, 12(sp)
   jal  func
   lw   ra, 12(sp)
   addi a0, a0, 1
   addi sp, sp, 16
   ret
```

Generating native programs

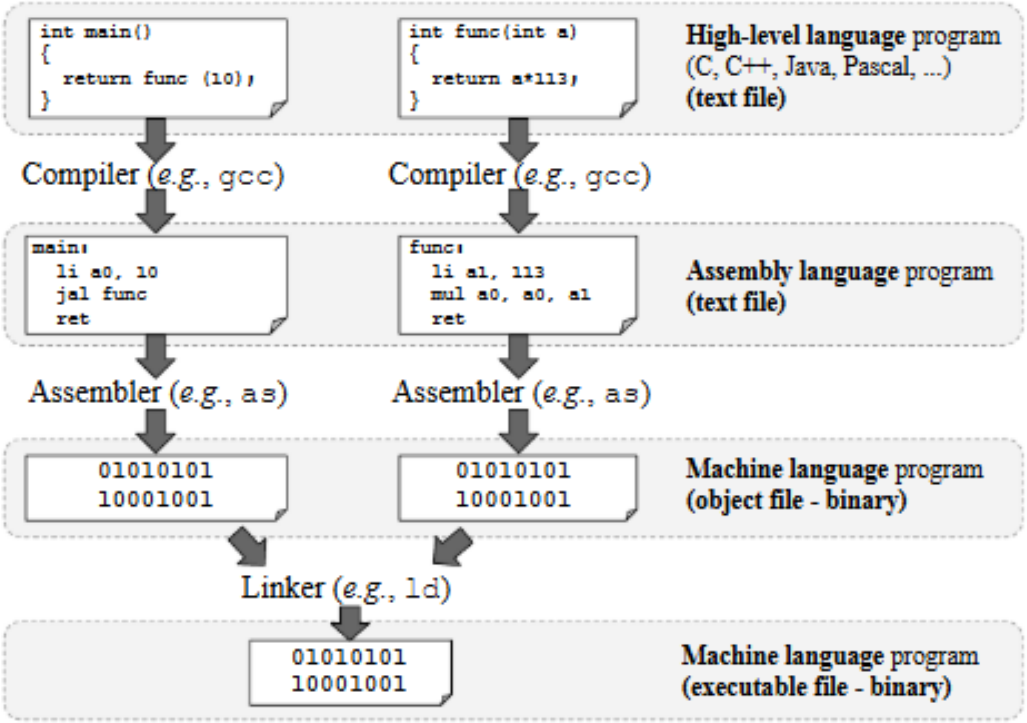
- **A compiler**
  - translate a program from one language to another
  - riscv64-unknown-elf-gcc -mabi=ilp32 -march=rv32i -S main.c -o main.s
  - The RV32I assembly program will be stored on the main.s file
- **An Assembler**
  - A assembler is a tool that translates a program in assembly language into a program in machine language
  - The GNU assembler tool (as) is an assembler
  - The assembler produces object files (.o) that are encoded in binary and contains code in machine language
  - riscv64-unknown-elf-as -mabi=ilp32 -march=rv32i main.s -o main.o

Generating native programs

- **A linker**
  - A tool that ‘links’ together one or more object files
  - Produces an executable file

riscv64-unknown-elf-ld -m elf32lriscv main.o mylib.o -o main.x

main.x is an executable file



Labels

- A label in assembly language
  - As a marker that represent program location
  - ‘x:’ label identifies a program location that contains a variable, which is allocated and initialized by the directive .word 10
  - ‘sum10:’ label identifies the program location that contains the first instruction of the sum10 routine

RISC-V assembly code

```
x:
   .word 10
sum10:
   lw   a0, x
   addi a0, a0, 10
   ret
```

Program Symbol

- Program symbols
  - “names” that are associated with numerical values
  - The “symbol table” is a data structure that maps each program system to its value
  - “nm” tool helps us to inspect the symbol table of a program

```
$ riscv64-unknown-elf-nm sum10.o
00000004 t .L0
00000004 t sum10
00000000 t x
```

Program Symbol

- Using the “.set” directive
  - Explicitly define symbols
  - The follow example that uses **.set** directive to define a symbol named answer and assign value 42 to it
  - Two symbols: answer, get_answer

```
.set   answer, 42
get_answer
   li   a0, answer
   ret
```

Global vs local Symbols

- Local symbols
  - Only visible on the same file
  - By default, the assembler registers labels as local symbols
- The .global directive
  - Instructs the assembler to register a label as a global symbol

```
.global  exit
exit:
   li   a0, 0
   li   a7, 93
   ecall
```

Program Entry Point

- The entry point is defined by an address
  - The address of the first instruction that must be executed
  - The linker sets the entry point field on the executable file and looks for a symbol named start
  - The linker sets the entry point to a default value (the address of the first instruction of the program) if the linker cannot find “start” symbol

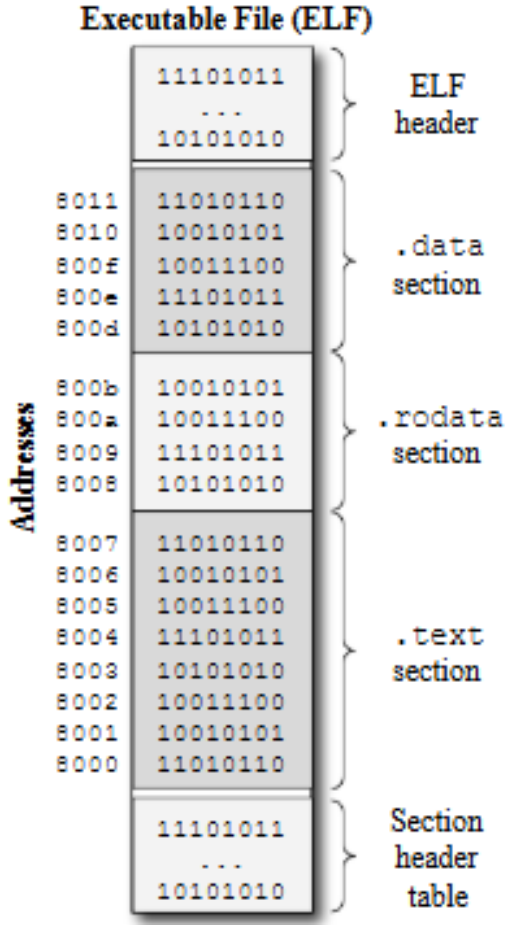
```
.global  start
start:
   li   a0, 10
   li   a1, 20
   jal  exit
```

Program Section

- The assembly program is usually organized in ‘sections’
  - A section may contain data or instructions
  - The contents of each section are mapped to a set of consecutive main memory addresses
  - .text
    - Store the program instructions
  - .data
    - Store initialized global variables
  - .bss
    - Store uninitialized global variables
  - .rodata
    - Store constants

Executable File (ELF)

- When linking multiple object files
  - The linker groups information from sections with the same name and places them together into a single section on the exec. File
  - .text are mapped to addresses 8000 to 8007
  - .data are mapped to addresses 800d to 8011



Assembly Language

- Assembly program contains
  - Comment
  - Labels
    - Usually defined by a name ended with the suffix “:”
  - Assembly instructions
    - Converted by the assembler into machine instructions
    - E.g. addi a0, a1, 1
  - Assembly directives
    - Commands used to coordinate the assembling process
    - E.g. .word 10
      - Instruct the assembler to assemble a 32-bit value (10) into the program

Program Structure I

- Plain text file with data declarations
- Data declaration section is followed by program code section

Data Declarations

- Identified with assembler directive **.data**
- Declares variable names used in program
- Storage allocated in main memory (*e.g.*, RAM)
- `<name>:  .<datatype> <value>`

Program Structure II

Code

- placed in section of text identified with assembler directive **.text**
- contains program code (instructions)
- starting point for code e.g. execution given label **start:**

Comments

Anything following # on a line

Program Structure III

The structure of an assembly program looks like this:

Program outline

```
# Comment giving name of program and description
# Template.asm
# Bare-bones outline of RISC-V assembly language program

.globl _start

.data    # variable declarations follow this line
         # ...
.text    # instructions follow this line

_start: # indicates start of code
# ...

# End of program, leave a blank line afterwards is preferred
```

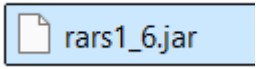
An Example RISC-V Assembly Program

```
1  .global _start
2
3  .data
4  welcome_msg: .asciz "Welcome!"
5
6  .text
7  _start:
8                  # STDOUT = 1
9                  addi a0, x0, 1
10                 # Load the address of 'welcome_msg'
11                 la a1, welcome_msg
12                 # Length of the string
13                 addi a2, x0, 8
14                 # Linux write system call
15                 addi a7, x0, 64
16                 # Call linux service to output the string
17                 ecall
```

RISC-V ISA Simulator - RARS

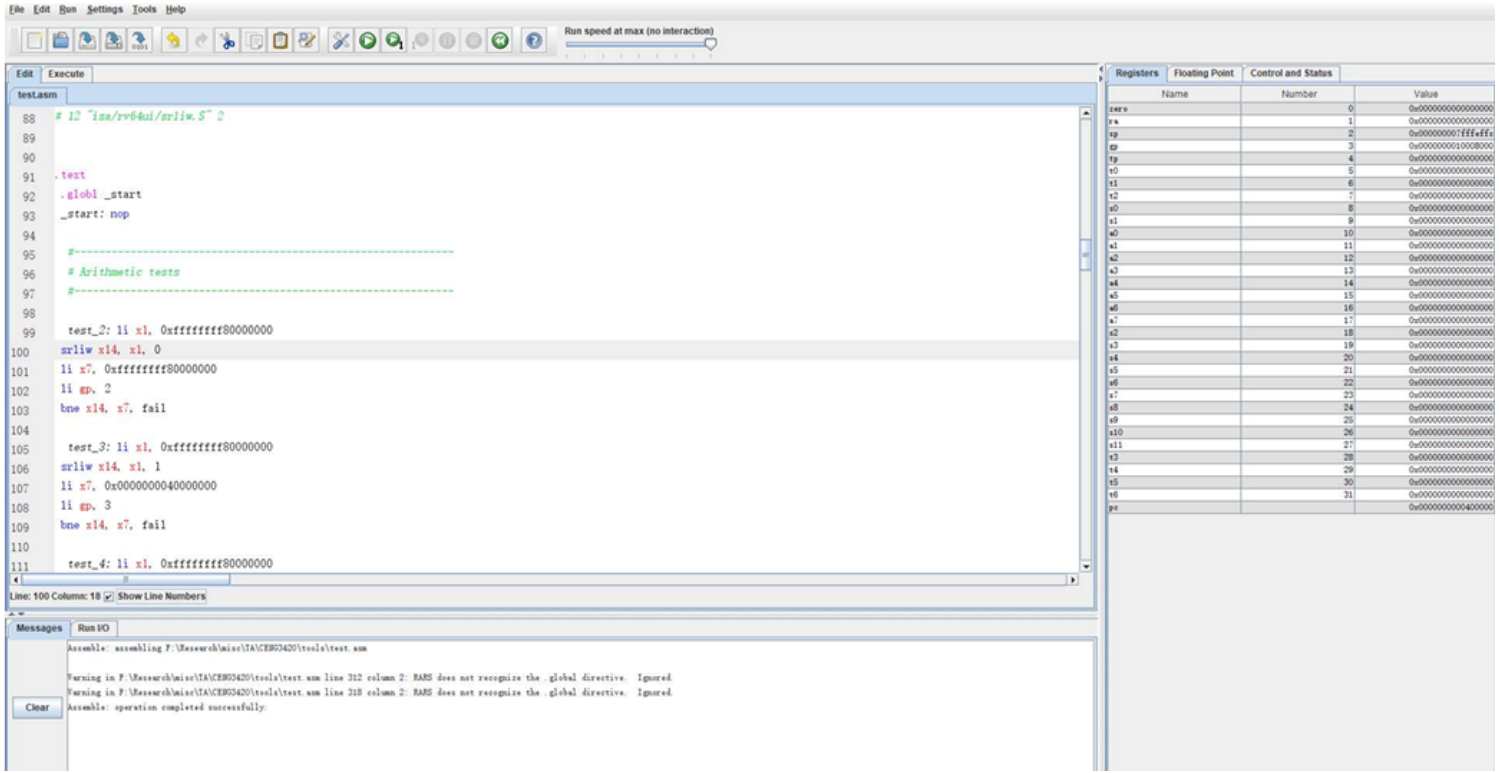
- RARS
  - The RISC-V Assembler, Runtime, and Simulator for RISC-V assembly language programs
  - **RARS** supports RISC-V IMFDN ISA base (riscv32 & riscv64)
  - **RARS** supports debugging using breakpoints like ebreak
  - **RARS** supports side by side comparison from pseudo-instruction to machine with intermediate steps
  - You need to Java environment to run RARS
    - https://www.java.com/download/ie_manual.jsp
  - Download RARA
    - https://github.com/TheThirdOne/rars/releases/tag/v1.6

RISC-V ISA Simulator - RARS

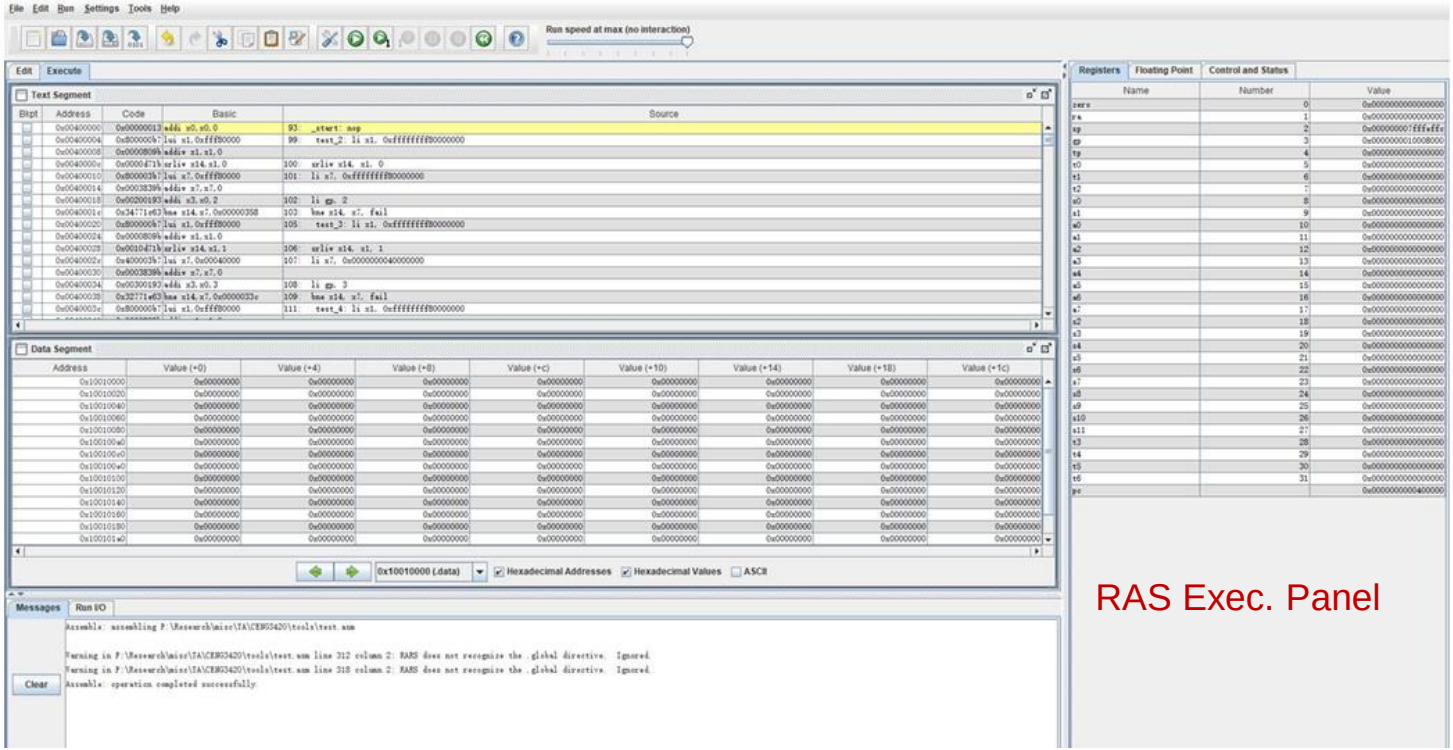
- How to run RARS?
  - Execute the command to start RARS: java -jar <rars jar path>
  - In Windows OS
    - Ensure you have installed Java Environment
    - Right click 

    - Open with -> Java (TM) SE Platform

RISC-V ISA Simulator - RARS

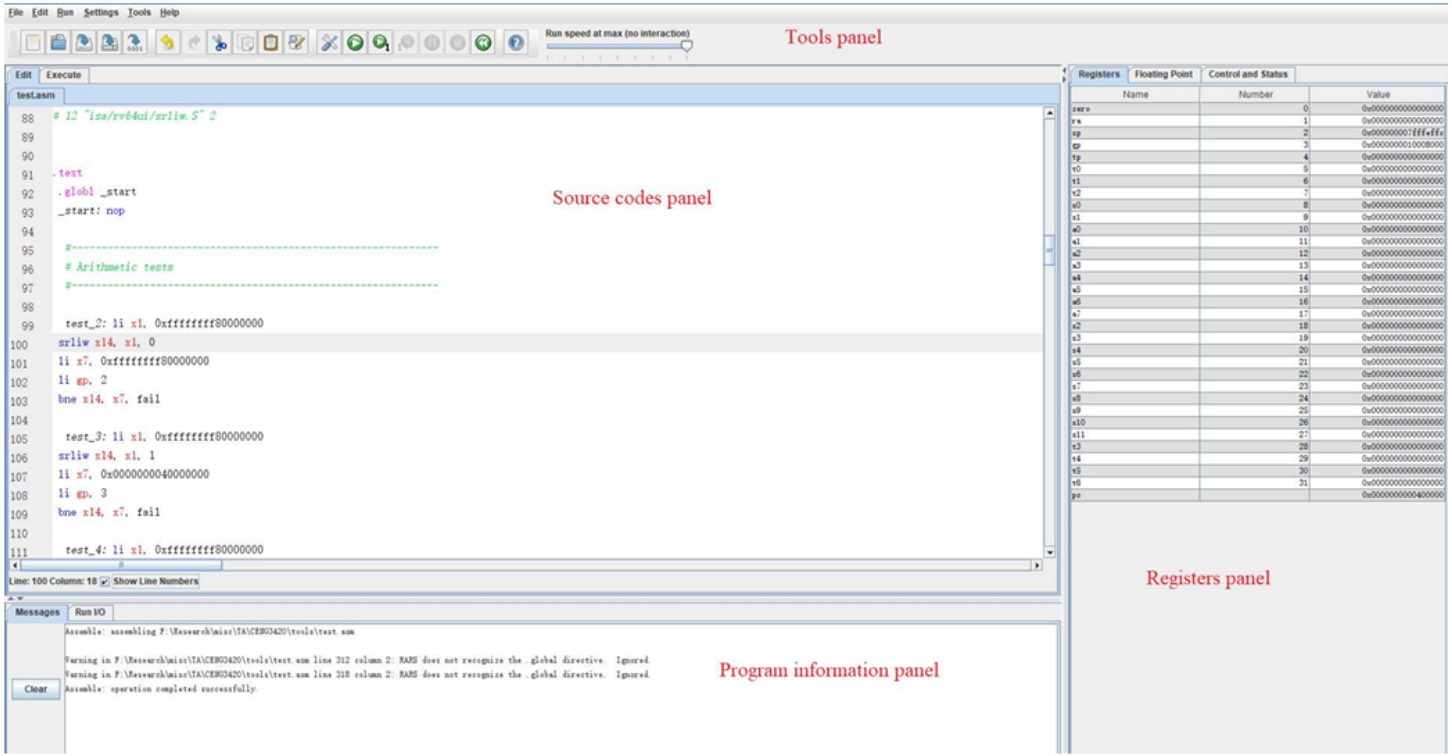


RISC-V ISA Simulator - RARS

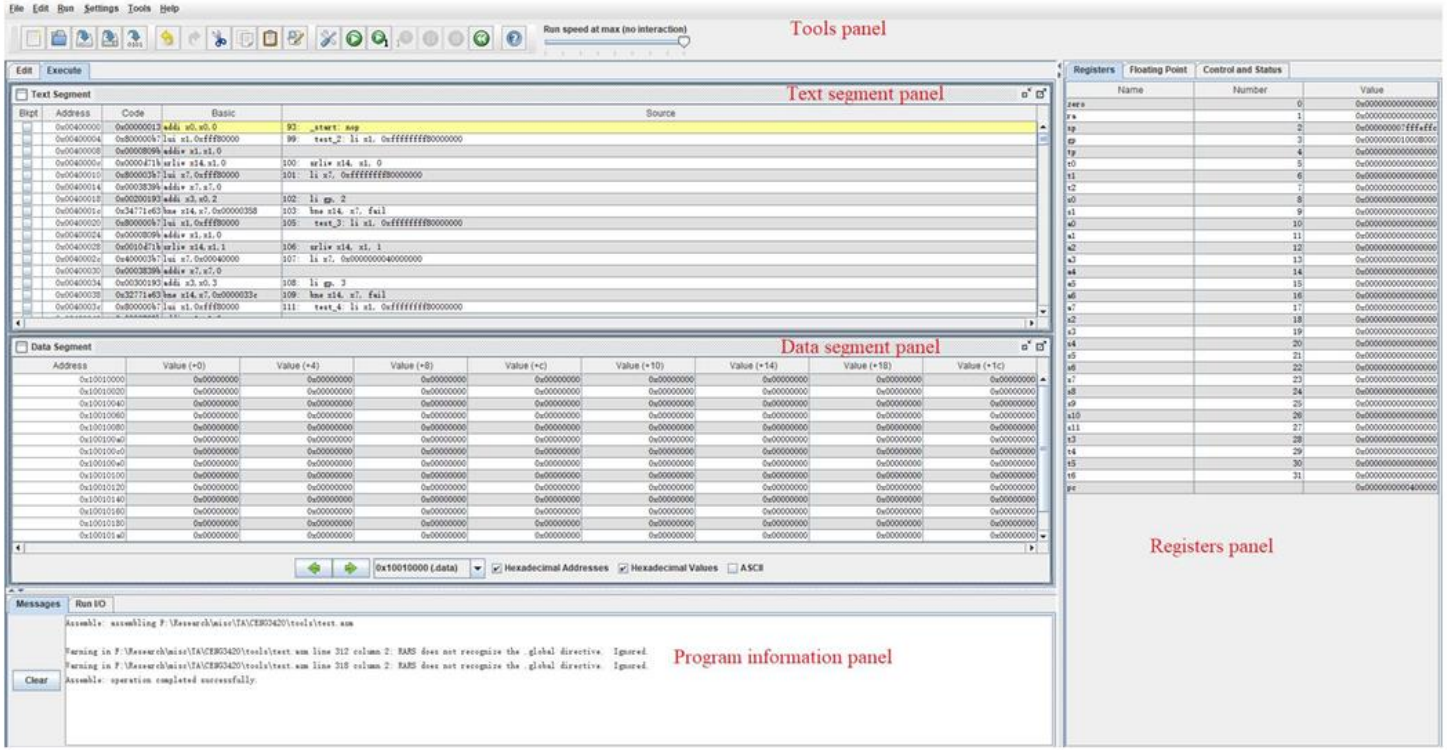


RAS Exec. Panel

RISC-V ISA Simulator - RARS



RISC-V ISA Simulator - RARS



RISC-V ISA Simulator - RARS

- RARS shortcut in windows OS
  - Create a new source file: Ctrl + N
  - Close the current source file: Ctrl + W
  - Assemble the source code: F3
  - Execute the current source code: F5
  - Step running: F7
  - Instructions & System call query: F1