#### CS 61C:

## Great Ideas in Computer Architecture MIPS Instruction Formats

Instructors:

Bernhard Boser and Randy H. Katz

http://inst.eecs.Berkeley.edu/~cs61c/fa16

#### Review: Basic Structure of a Function

#### Prologue

```
entry_label:
addi $sp,$sp, -framesize
sw $ra, framesize-4($sp) # save $ra
save other regs if need be

Body ... (call other functions...)
```

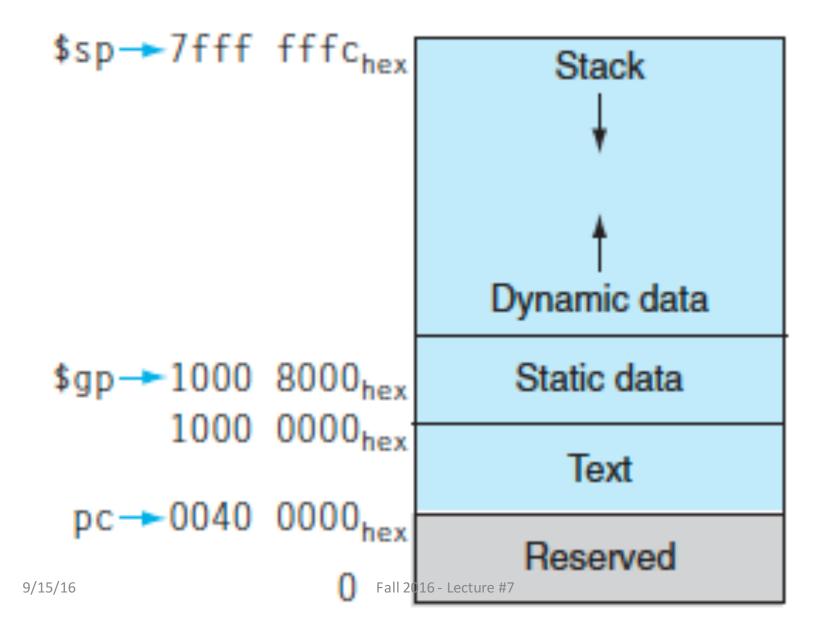
#### **Epilogue**

```
restore other regs if need be
lw $ra, framesize-4($sp) # restore $ra
addi $sp,$sp, framesize
jr $ra
```

# Review: Where is the Stack in Memory?

- MIPS convention
- Stack starts in high memory and grows down
  - Hexadecimal (base 16): 7fff fffc<sub>hex</sub>
- MIPS programs (text segment) in low end
  - $-00400000_{\text{hex}}$
- Static data segment (constants and other static variables) above text for static variables
  - MIPS convention global pointer (\$gp) points to static
- Heap above static for data structures that grow and shrink; grows up to high addresses

#### Review: MIPS Memory Allocation



# Review: Register Allocation and Numbering

Name	Register number	Usage	Preserved on call?
\$zero	0	The constant value 0	h.a.
\$v0-\$v1	2-3	Values for results and expression evaluation	no
\$a0-\$a3	4-7	Arguments	no
\$t0-\$t7	8-15	Temporaries	no
\$s0 <b>-</b> \$s7	16-23	Saved	yes
\$t8_\$t9	24-25	More temporaries	no
\$gp	28	Global pointer	yes
\$sp	29	Stack pointer	yes
\$fp	30	Frame pointer	yes
\$ra	31	Return address	yes

\$1 used by assembler (later in this lecture)

\$26, \$27 used by the operating system

#### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

#### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

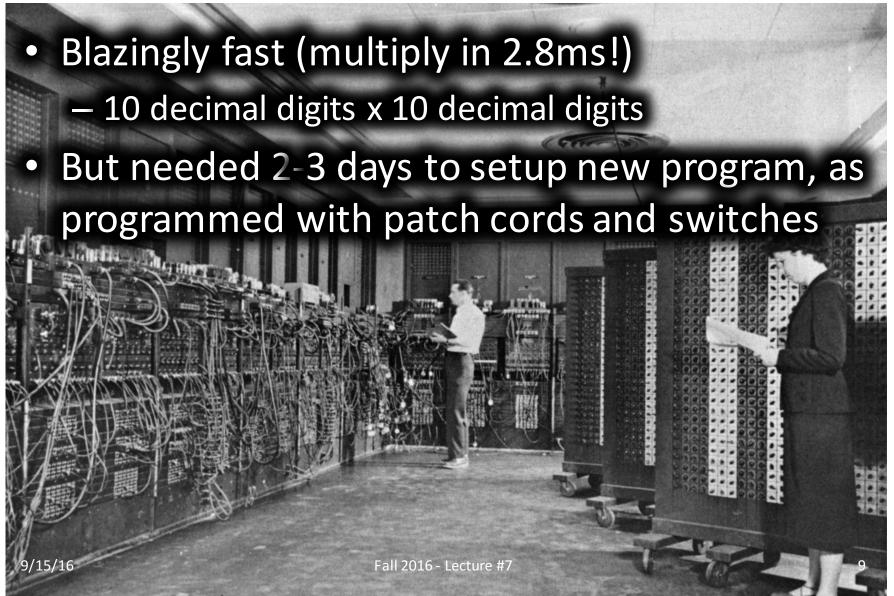
## Levels of Representation/Interpretation

```
temp = v[k];
v[k] = v[k+1];
        High Level Language
          Program (e.g., C)
                                               v[k+1] = temp;
                    Compiler
                                                    $t0, 0($2)
                                                                   Anything can be represented
         Assembly Language
                                                    $t1, 4($2)
                                                                                 as a number,
                                                    $t1, 0($2)
         Program (e.g., MIPS)
                                                SW
                                                                        i.e., data or instructions
                                                    $t0, 4($2)
                                                SW
                    Assembler
         Machine Language
           Program (MIPS)
                                                                       1100
                                                                             0110
Machine
Interpretation
                                                   Register File
  Hardware Architecture Description
         (e.g., block diagrams)
                                                      ALU
Architecture
Implementation
         Logic Circuit Description
       (Circuit Schematic Diagrams)
```

9/15/16

Fall 2016 - Lecture #

#### ENIAC (U.Penn., 1946) First Electronic General-Purpose Computer



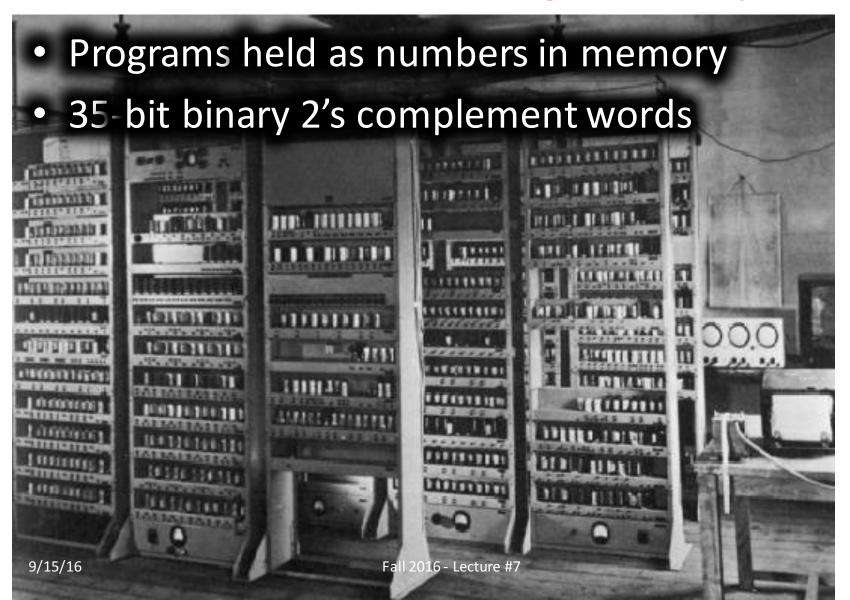
## Big Idea: Stored-Program Computer

First Draft of a Report on the EDVAC
by
John von Neumann
Contract No. W-670-ORD-4926
Between the
United States Army Ordnance Department and the
University of Pennsylvania
Moore School of Electrical Engineering
University of Pennsylvania

June 30, 1945

- Instructions are represented as bit patterns can think of these as numbers
- Therefore, entire programs can be stored in memory to be read or written just like data
- Can reprogram quickly (seconds), don't have to rewire computer (days)
- Known as the "von Neumann" computers after widely distributed tech report on EDVAC project
  - Wrote-up discussions of Eckert and Mauchly
  - Anticipated earlier by Turing and Zuse

#### EDSAC (Cambridge, 1949) First General Stored-Program Computer



# Consequence #1: Everything Addressed

- Since all instructions and data are stored in memory, everything has a memory address: instructions, data words
  - Both branches and jumps use these
- C pointers are just memory addresses: they can point to anything in memory
  - Unconstrained use of addresses can lead to nasty bugs; up to you in C; limited in Java by language design
- One register keeps address of instruction being executed:
   "Program Counter" (PC)
  - Basically a pointer to memory: Intel calls it Instruction Pointer (a better name)

# Consequence #2: Binary Compatibility

- Programs are distributed in binary form
  - Programs bound to specific instruction set
  - Different version for Macintoshes and PCs
- New machines want to run old programs ("binaries")
  as well as programs compiled to new instructions
- Leads to "backward-compatible" instruction set evolving over time
- Selection of Intel 8086 in 1981 for 1<sup>st</sup> IBM PC is major reason latest PCs still use 80x86 instruction set; could still run program from 1981 PC today

#### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

### Instructions as Numbers (1/2)

- Currently all data we work with is in words (32-bit chunks):
  - Each register is a word
  - lw and sw both access memory one word at a time
- So how do we represent instructions?
  - Remember: Computer only understands 1s and 0s, so
     "add \$t0,\$0,\$0" is meaningless
  - MIPS/RISC seeks simplicity: since data is in words,
     make instructions be fixed-size 32-bit words also

### Instructions as Numbers (2/2)

- One word is 32 bits, so divide instruction word into "fields"
- Each field tells processor something about instruction
- We could define different fields for each instruction, but MIPS seeks simplicity, so define three basic types of instruction formats:
  - R-format
  - I-format
  - J-format

#### Instruction Formats

- I-format: used for instructions with immediates, lw and sw (since offset counts as an immediate), and branches (beq and bne)
  - (but not the shift instructions; later)
- J-format: used for j and jal
- R-format: used for all other instructions
- It will soon become clear why the instructions have been partitioned in this way

#### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

## R-Format Instructions (1/5)

• Define "fields" of the following number of bits each: 6 + 5 + 5 + 5 + 5 + 6 = 32



• For simplicity, each field has a name:

opcode rs	rt	rd	shamt	funct
-----------	----	----	-------	-------

- Important: On these slides and in book, each field is viewed as a 5- or 6-bit unsigned integer, not as part of a 32-bit integer
  - Consequence: 5-bit fields can represent any number 0-31, while
     6-bit fields can represent any number 0-63

### R-Format Instructions (2/5)

- What do these field integer values tell us?
  - opcode: partially specifies what instruction it is
    - Note: This number is equal to 0 for all R-Format instructions
  - funct: combined with opcode, this number exactly specifies the instruction
- Question: Why aren't opcode and funct a single 12-bit field?
  - We'll answer this later

### R-Format Instructions (3/5)

#### More fields:

- <u>rs</u> (Source Register): *usually* used to specify register containing first operand
- <u>rt</u> (Target Register): *usually* used to specify register containing second operand (note that name is misleading)
- <u>rd</u> (Destination Register): usually used to specify register which will receive result of computation

### R-Format Instructions (4/5)

- Notes about register fields:
  - Each register field is exactly 5 bits, which means that it can specify any unsigned integer in the range 0-31. Each of these fields specifies one of the 32 registers by number
  - The word "usually" was used because there are exceptions that we'll see later

### R-Format Instructions (5/5)

#### Final field:

- shamt: This field contains the amount a shift instruction will shift by. Shifting a 32-bit word by more than 31 is useless, so this field is only 5 bits (so it can represent the numbers 0-31)
- This field is set to 0 in all but the shift instructions
- For a detailed description of field usage for each instruction, see MIPS Green Card



## P S Reference Data



#### CORE INSTRUCTION SET

OPCODE

FOR-

/ FUNCT

NAME, MNEMONIC

MAT

OPERATION (in Verilog)

(Hex)

Add

add R R[rd] = R[rs] + R[rt]

(1)  $0/20_{hex}$ 

#### BASIC INSTRUCTION FORMATS

R	opco	ode	1	rs		rt	1	rd	shamt	funct	
	31	26	25	21	20	16	15	11	10 6	5	0
I	opco	ode	1	rs		rt			immediate	2	
	31	26	25	21	20	16	15				0
$\mathbf{J}$	opco	ode		address							
	31	26	25								0

(1) May cause overflow exception

## R-Format Example (1/2)

MIPS Assembly Instruction:

```
add $8,$9,$10
```

```
opcode = 0 (look up in table in book)
funct = 32 (look up in table in book)
rd = 8 (destination)
rs = 9 (first operand)
rt = 10 (second operand)
shamt = 0 (not a shift)
```

## R-Format Example (2/2)

MIPS Assembly Instruction:

add \$8,\$9,\$10

Decimal number per field representation:

0 9 10 8 0 32

Binary number per field representation:

000000 01001 01010 01000 00000 100000

hex representation:  $012A 4020_{hex}$ 

Called a <u>Machine Language Instruction</u>

opcode rs rt rd shamt funct

#### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

#### I-Format Instructions (1/4)

- What about instructions with immediates?
  - 5-bit field only represents numbers up to the value 31: immediates may be much larger than this
  - Ideally, MIPS would have only one instruction format (for simplicity): unfortunately, we need to compromise
- Define new instruction format that is partially consistent with R-format:
  - First notice that, if instruction has immediate, then it uses at most 2 registers

## I-Format Instructions (2/4)

• Define "fields" of the following number of bits each:

$$6 + 5 + 5 + 16 = 32$$
 bits

– Again, each field has a name:

opcode	rs r	t imme	ediate
--------	------	--------	--------

Key Concept: Only one field is inconsistent with R-format.
 Most importantly, opcode is still in same location!

#### I-Format Instructions (3/4)

- What do these fields mean?
  - opcode: same as before except that, since there's no
     funct field, opcode uniquely specifies an instruction in
     I-format
  - Also answers why R-format has two 6-bit fields to identify instruction instead of a single 12-bit field: to be as consistent as possible with other formats while leaving as much space as possible for immediate field
  - <u>rs</u>: specifies a register operand (if there is one)
  - <u>rt</u>: specifies register which will receive result of computation (this is why it's called the *target* register "rt") or other operand for some instructions

#### I-Format Instructions (4/4)

- The Immediate Field:
  - addi, slti, sltiu, the immediate is signextended to 32 bits ... it's treated as a signed integer
  - 16 bits → can be used to represent immediate up to 2<sup>16</sup> different values
  - Large enough to handle the offset in a typical
     lw or sw, plus a vast majority of values that will be used in the slti instruction
  - Later, we'll see what to do when a value is too big for 16 bits



## P S Reference Data



#### CORE INSTRUCTION SET

OPCODE

FOR-

/ FUNCT

NAME, MNEMONIC MAT

OPERATION (in Verilog)

(Hex)

Add Immediate

addi I R[rt] = R[rs] + SignExtImm

(1,2)

 $8_{\text{hex}}$ 

#### BASIC INSTRUCTION FORMATS

R	opcod	e		rs		rt		rd	shamt	funct	
	31	26	25	21	20	16	15	11	10 6	5	0
I	opcod	e		rs		rt			immediate	e	
	31	26	25	21	20	16	15			(	0
$\mathbf{J}$	opcod	e						address			
	31	26	25							(	0

- (1) May cause overflow exception
- (2) SignExtImm = { 16{immediate[15]}, immediate }

## I-Format Example (1/2)

MIPS Assembly Instruction:

```
addi $21,$22,-50
```

```
opcode = 8 (look up in table in book)
rs = 22 (register containing operand)
rt = 21 (target register)
immediate = -50 (by default, this is decimal in assembly code)
```

## I-Format Example (2/2)

MIPS Assembly Instruction:

```
addi $21,$22,-50
```

#### Decimal/field representation:

8 22 21	-50
---------	-----

#### **Binary/field representation:**

```
001000 10110 10101 1111111111001110
```

hexadecimal representation: 22D5  $FFCE_{hex}$ 

### Clicker/Peer Instruction

Which instruction has same representation as integer  $35_{ten}$ ?

- a) add \$0, \$0, \$0
- b) subu \$s0,\$s0,\$s0
- c) lw \$0, 0(\$0)
- d) addi \$0, \$0, 35
- e) subu \$0, \$0, \$0

•				ten			
opcode	rs	rt	rd	shamt	funct		
opcode	rs	rt	rd	shamt	funct		
opcode	rs	rt		offset			
opcode	rs	rt	immediate				
opcode	rs	rt	rd	shamt	funct		

Registers numbers and names: 0: \$0, .. 8: \$t0, 9:\$t1, ..15: \$t7, 16: \$s0, 17: \$s1, .. 23: \$s7

Opcodes and function fields:

add: opcode = 0, funct = 32

**subu**: opcode = 0, funct = 35

addi: opcode = 8

**1w**: opcode = 35



## P S Reference Data



#### CORE INSTRUCTION SET

FOR-

NAME, MNEMONIC MAT

OPERATION (in Verilog)

 $0/23_{hex}$ 

OPCODE

/ FUNCT

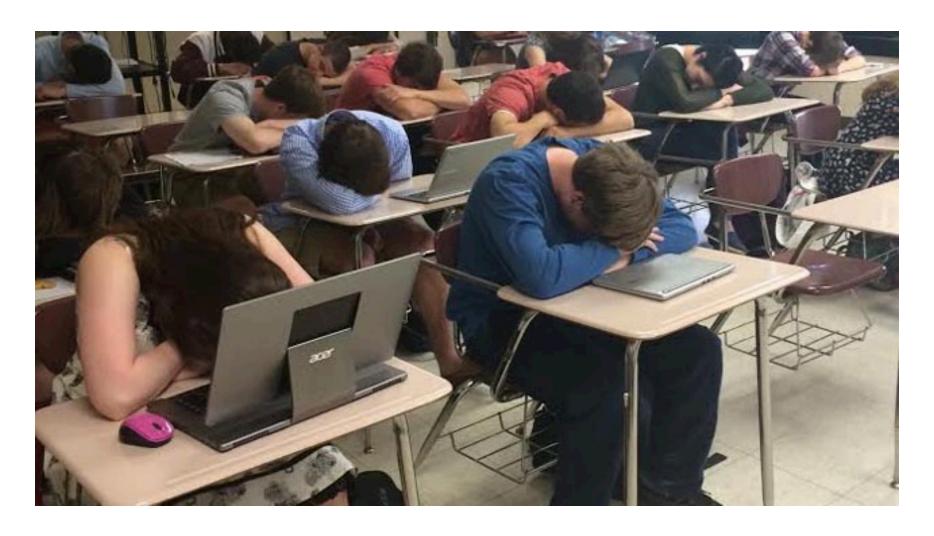
(Hex)

Subtract Unsigned subu R R[rd] = R[rs] - R[rt]

#### BASIC INSTRUCTION FORMATS

R	opcode			rs		rt		rd	shamt		funct	
	31	26	25	21	20	1	6	15 11	10	5 5		0
I	opcode			rs		rt			immediat	te		
	31	26	25	21	20	1	6	15				0
$\mathbf{J}$	opcode							address				
	31	26	25									0

# Break!



### Administrivia

- Project #1 due September 22 @ 11:59:59 PM
  - C Guerrilla sections were a great success! keep your eye on piazza for more C help sessions
- Midterm #1 in 1.5 weeks: September 27!
  - Covers Number Representations, C (including Project #1),
     MIPS Assembly and Machine Language,
     Compiler/Assembly/Linking/Loading (thru next Tuesday's lecture)
  - Two sided 8.5" x 11" cheat sheet + MIPS Green Card that we give you
  - Review session preceding Sunday (September 25) 1-3 PM Dwinelle 155
  - DSP students: please make sure we know about your special accommodations (contact Derek and Stephan the co-Head TAs)

38

9/15/16 Fall 2016 - Lecture #7

### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

# **Branching Instructions**

- beq and bne
  - Need to specify a target address if branch taken
  - Also specify two registers to compare
- Use I-Format:

opcode rs rt immediate

- opcode specifies beq (4) vs. bne (5)
- rs and rt specify registers
- How to best use immediate to specify addresses?

# **Branching Instruction Usage**

- Branches typically used for loops (if-else, while, for)
  - Loops are generally small (< 50 instructions)</li>
  - Function calls and unconditional jumps handled with jump instructions (J-Format)
- Recall: Instructions stored in a localized area of memory (Code/Text)
  - Largest branch distance limited by size of code
  - Address of current instruction stored in the program counter (PC)

## **PC-Relative Addressing**

- PC-Relative Addressing: Use the immediate field as a two's complement offset to PC
  - Branches generally change the PC by a small amount
  - Can specify ± 2<sup>15</sup> addresses from the PC

### **Branch Calculation**

If we don't take the branch:

```
- PC = PC + 4 = next instruction
```

• If we do take the branch:

```
-PC = (PC+4) + (immediate*4)
```

#### Observations:

- immediate is number of instructions to jump (remember, specifies words) either forward (+) or backwards (-)
- Branch from PC+4 for hardware reasons; will be clear why later in the course



# MIPS Reference Data



#### **CORE INSTRUCTION SET**

FOR-

NAME, MNEMONIC

Branch On Equal

beg

MAT .

OPERATION (in Verilog) if(R[rs]==R[rt])

PC=PC+4+BranchAddr

/ FUNCT

OPCODE

(Hex)

(4) 4<sub>hex</sub>

#### BASIC INSTRUCTION FORMATS

R		opcode		rs		rt		rd	shamt	funct	
	31	26	25	21	20	16	15	11	10 6	5	0
I		opcode		rs		rt			immediate	•	
	31	26	25	21	20	16	15				0
$\mathbf{J}$		opcode		address							
	31	26	25								0

(4) BranchAddr = { 14{immediate[15]}, immediate, 2'b0 }

# Branch Example (1/2)

#### • MIPS Code:

Start counting from instruction AFTER the branch

I-Format fields:

```
opcode = 4
rs = 9
rt = 0
immediate = 3
```

(look up on Green Sheet) (first operand) (second operand)

# Branch Example (2/2)

Loop: beq \$9,\$0,End

addu \$8,\$8,\$10

#### MIPS Code:

```
addiu $9,$9,-1
j Loop
End:

31 Field representation (decimal):

4 9 0 3

Field representation (binary):

000100 01001 00000 000000000000011
```

## Questions on PC-addressing

- Does the value in branch immediate field change if we move the code?
  - If moving individual lines of code, then yes
  - If moving all of code, then no
- What do we do if destination is  $> 2^{15}$  instructions away from branch?
  - Other instructions save us

### Questions on PC-addressing

- Does the value in branch immediate field change if we move the code?
  - If moving individual lines of code, then yes
  - If moving all of code, then no
- What do we do if destination is  $> 2^{15}$  instructions away from branch?
  - Other instructions save us

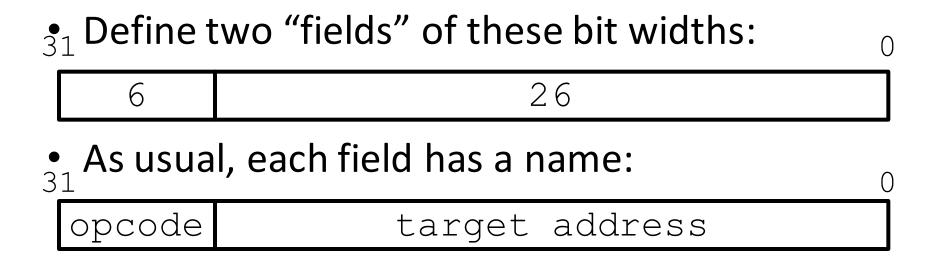
# Break!



## J-Format Instructions (1/4)

- For branches, we assumed that we won't want to branch too far, so we can specify a change in the PC
- For general jumps (j and jal), we may jump to anywhere in memory
  - Ideally, we would specify a 32-bit memory address to jump to
  - Unfortunately, we can't fit both a 6-bit opcode
     and a 32-bit address into a single 32-bit word

# J-Format Instructions (2/4)



### Key Concepts:

- Keep opcode field identical to R-Format and I-Format for consistency
- Collapse all other fields to make room for large target address Fall 2016 - Lecture #7

## J-Format Instructions (3/4)

- We can specify 2<sup>26</sup> addresses
  - Still going to word-aligned instructions, so add 0b00
     as last two bits (multiply by 4)
  - This brings us to 28 bits of a 32-bit address
- Take the 4 highest order bits from the PC
  - Cannot reach everywhere, but adequate almost all of the time, since programs aren't that long
  - Only problematic if code straddles a 256MB boundary
- If necessary, use 2 jumps or jr (R-Format) instead

### J-Format Instructions (4/4)

- Jump instruction:
  - New PC = { (PC+4)[31..28], target address, 00 }
- Notes:
  - - Book uses || instead
  - Array indexing: [31..28] means highest 4 bits
  - For hardware reasons, use PC+4 instead of PC

53



# P S Reference Data



#### CORE INSTRUCTION SET

OPCODE / FUNCT

FOR-NAME, MNEMONIC MAT

OPERATION (in Verilog)

(Hex)

Jump

J PC=JumpAddr

(5)

 $2_{\text{hex}}$ 

#### BASIC INSTRUCTION FORMATS

R	opc	code		rs	r	t		rd	shamt	funct		
	31	26	25	21	20	16	15	11	10 6	5	0	
I	opc	code		rs	r	t			immediate	е		
	31	26	25	21	20	16	15				0	
$\mathbf{J}$	opc	ode		address								
	31	26	25								0	

(5)  $JumpAddr = \{ PC+4[31:28], address, 2'b0 \}$ 

### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

### Assembler Pseudo-Instructions

- Certain C statements are implemented unintuitively in MIPS
  - e.g., assignment (a=b) via add \$zero
- MIPS has a set of "pseudo-instructions" to make programming easier
  - More intuitive to read, but get translated into actual instructions later
- Example:

move dst, src

56

### Assembler Pseudo-Instructions

- Certain C statements are implemented unintuitively in MIPS
  - e.g. assignment (a=b) via add \$zero
- MIPS has a set of "pseudo-instructions" to make programming easier
  - More intuitive to read, but get translated into actual instructions later
- Example:

```
move dst, src
```

### translated into

```
addi dst, src, 0
```

### Assembler Pseudo-Instructions

List of pseudo-instructions:

http://en.wikipedia.org/wiki/MIPS\_architecture#Pseudo\_instructions

- List also includes instruction translation
- Load Address (la)
  - -la dst, label
  - Loads address of specified label into dst
- Load Immediate (li)
  - -li dst, imm
  - Loads 32-bit immediate into dst
- MARS has additional pseudo-instructions
  - See Help (F1) for full list



Home

Features

Download

License

Papers

Help & Info

Contact Us

#### MARS (MIPS Assembler and Runtime Simulator)

#### An IDE for MIPS Assembly Language Programming

MARS is a lightweight interactive development environment (IDE) for programming in MIPS assembly language, intended for educational-level use with Patterson and Hennessy's Computer Organization and Design.



Feb. 2013: "MARS has been tested in the Softpedia labs using several industry-leading security solutions and found to be completely clean of adware/spyware components. ... Softpedia guarantees that MARS 4.3 is 100% FREE, which means it does not contain any form of malware, including spyware, viruses, trojans and backdoors."

<u>Download MARS from Softpedia</u> (version on Softpedia may lag behind the version on this page).

#### Download MARS 4.5 software! (Aug. 2014)

New for 4.0: new editor, featuring multiple files, context-sensitive input, and color-coding.

- See a <u>screenshot</u> (1478 x 889 pixels, 198 KB JPEG)
- Tutorial materials
- Sample MIPS assembly program to run under MARS <u>Fibonacci.asm</u>

MARS features overview: (List of features by version)

# Assembler Register

#### Problem:

- When breaking up a pseudo-instruction, the assembler may need to use an extra register
- If it uses a regular register, it'll overwrite whatever the program has put into it

#### Solution:

- Reserve a register (\$1 or \$at for "assembler temporary") that assembler will use to break up pseudo-instructions
- Since the assembler may use this at any time, it's not safe to code with it

# Dealing With Large Immediates

- How do we deal with 32-bit immediates?
  - Sometimes want to use immediates > ± 2<sup>15</sup> with addi, lw, sw and slti
  - Bitwise logic operations with 32-bit immediates
- Solution: Don't mess with instruction formats, just add a new instruction
- Load Upper Immediate (lui)
  - -lui reg, imm
  - Moves 16-bit imm into upper half (bits 16-31) of reg and zeros the lower half (bits 0-15)

### lui Example

- Want: addiu \$t0,\$t0,0xABABCDCD
  - This is a pseudo-instruction!
- Translates into:

### lui Example

- Want: addiu \$t0,\$t0,0xABABCDCD
  - This is a pseudo-instruction!
- Translates into:

```
lui $at,0xABAB # upper 16
ori $at,$at,0xCDCD# lower 16
addu $t0,$t0,$at # move
Only the assembler gets to use $at ($1)
```

 Now we can handle everything with a 16-bit immediate!

# Pseudo Instructions (Green Card)

#### **PSEUDOINSTRUCTION SET**

NAME	MNEMONIC	OPERATION
Branch Less Than	blt	if(R[rs] < R[rt]) PC = Label
Branch Greater Than	bgt	if(R[rs]>R[rt]) PC = Label
Branch Less Than or Equal	ble	$if(R[rs] \le R[rt]) PC = Label$
Branch Greater Than or Equal	bge	$if(R[rs] \ge R[rt]) PC = Label$
Load Immediate	li	R[rd] = immediate
Move	move	R[rd] = R[rs]

### MAL vs. TAL

- True Assembly Language (TAL)
  - The instructions a computer understands and executes

- MIPS Assembly Language (MAL)
  - Instructions the assembly programmer can use (includes pseudo-instructions)
  - Each MAL instruction becomes 1 or more TAL instruction

### Outline

- Stored Program Computers
- MIPS Instruction Formats
- Register Format
- Immediate Format
- Branch/Jump Format
- Other Instruction Considerations
- And in Conclusion ...

## And in Conclusion, ...

- I-Format: instructions with immediates,
   lw/sw (offset is immediate), and beq/bne
  - But not the shift instructions

Branches use PC-relative addressing

I: opcode rs rt immediate

- J-Format: j and jal (but not jr)
  - Jumps use absolute addressing

J: opcode target address

R-Format: all other instructions

R: opcode rs rt rd shamt funct