Computer Structure and Language

Hamid Sarbazi-Azad

assembly instruction.

Department of Computer Engineering Sharif University of Technology (SUT) Tehran, Iran



Assembly Language

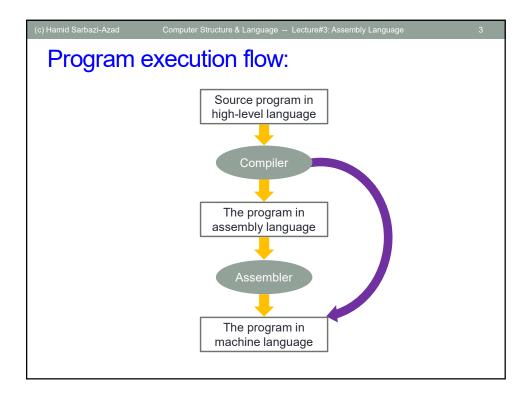
Example Machine 1: A 2-address machine with 4 instructions
Add (opcode=00),
Subtract (opcode =01),
Move (opcode=10), and
No-operation (opcode=11), and a128-location memory.

Since writing, reading and understanding programs of machine instructions in binary (machine) format is obscure and complex, we use symbolic representation of instructions, named

For example, the machine instruction 01 0001111 0001010 may be written, in a symbolic form, as:

sub num1,num2

where num1 and num2 are symbolic equivalents of operands M_{15} and M_{10} and sub is symbolic equivalent (Mnemonic) of opcode=01.



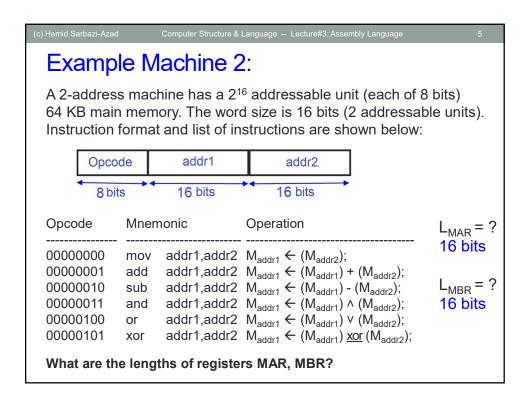
Advantages of Assembly Language:

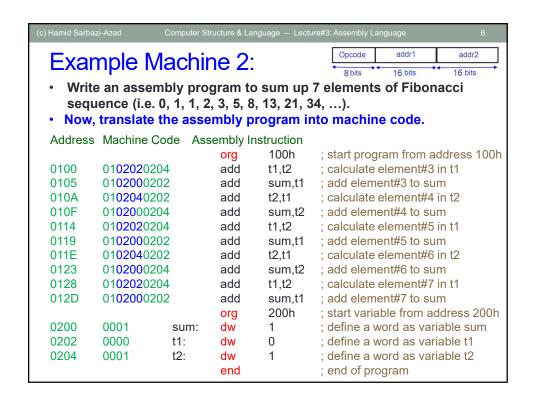
Understanding of assembly language provides knowledge of:

- Interface of programs with OS, processor and BIOS (basic input/output system);
- Representation of data in memory and other external devices;
- How processor accesses and executes instructions;
- How instructions access and process data;
- How a program accesses external devices.

Advantages of using assembly language are:

- It requires less memory and execution time;
- It allows hardware-specific complex jobs in an easier way;
- It is suitable for time-critical jobs;
- It is most suitable for writing interrupt service routines and other memory resident programs;
- Used to crack locked software;
- Used to write viruses/worms/....





Computer Structure & Language -- Lecture#3: Assembly Language

7

Registers

An n-bit Register is made of n flip-flops controlled in group.

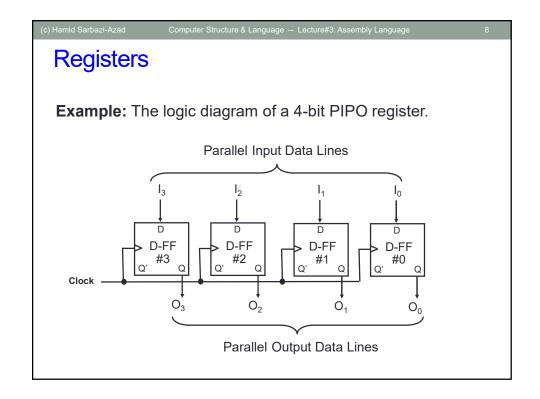
Based on the availability of flip-flops' output and their access method, registers can be:

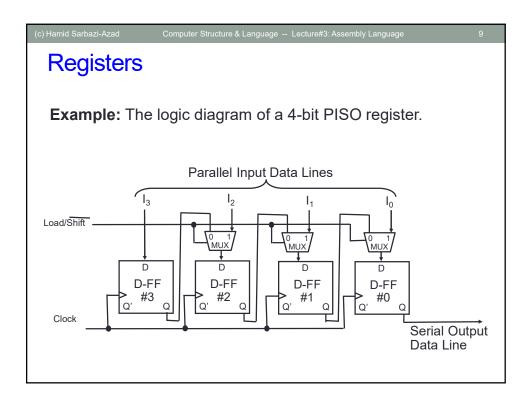
PIPO: Parallel In, Parallel Out

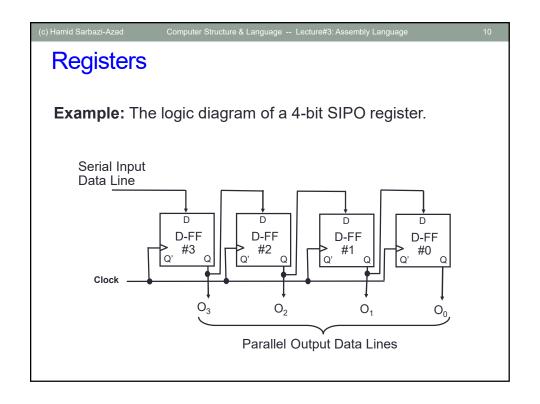
SIPO: Serial In, Parallel Out

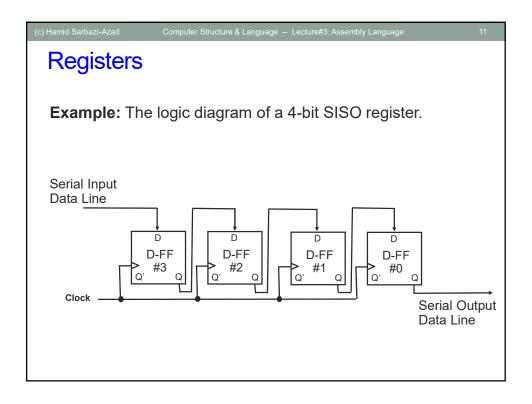
• PISO: Parallel In, Serial Out

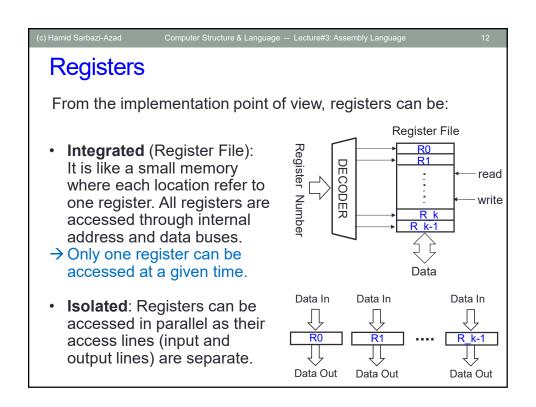
SISO: Serial In, Serial Out











Computer Structure & Language -- Lecture#3: Assembly Language

13

Registers

From the addressability point of view, registers can be:

- Addressable: Assembly programmer can address them in the instructions.
 - e.g. R0, R1, ..., R15 in IBM360 family or AX, BX, CX, DX, ... in 8086/88 processor We will see them later when learning their assembly programming.
- Non-addressable: They are not addressable by the programmer, but are used in the machine.
 e.g. MAR, MBR, PC, IR, ...
 We will see them later.

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

14

Registers

From the usage point of view, registers can be:

- General purpose registers: Addressable registers that can be used for all purposes defined by the machine instruction set.
 - e.g. Registers 0, 1, ...15 in IBM360/370 family.
- Special purpose registers: Addressable registers that may be used for special purposes defined by some machine instruction set.
 - e.g. Registers CX (for counting) or DX (for division, multiplication, and I/O instructions in 8086/88 family.

Computer Structure & Language -- Lecture#3: Assembly Language

15

Internal Machine Registers

Machine non-addressable registers include:

- MAR (Memory Address Register) keeps memory location address during a memory cycle.
- MBR (Memory Buffer Register) holds the data to be read/written from/into memory.
- **IR** (Instruction Register) holds the instruction fetched from main memory into CU.
- **PC** (Program Counter) holds the address of the next instruction to be fetched from main memory and executed.

We will see some other registers later.

Machine addressable registers include general-purpose and special-purpose registers.

The number of such registers and their purpose are different from machine to machine.

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

16

Addressing modes

- 1. **Direct addressing**: Address of operand is given in the IF (instruction format). It can be:
 - Register direct (register number is given in IF)
 - Memory direct (memory address is given in IF)

Example 1: mov ax, array ; in 8086/88 processor

Here, ax is a symbolic address for register ax which is included in the IF. Array is also a symbolic address mentioned in the IF to directly address a word in main memory.

Example 2: L 2,NUM ; in IBM 360 family

Here, 2 indicates register#2 and NUM is the symbolic address of a word in main memory.

Almost all machines employ direct addressing modes.

Computer Structure & Language -- Lecture#3: Assembly Language

17

Addressing modes

- 2. **Indirect addressing**: The address included in the IF indicates the register or memory location that contains the address of the operand in memory). It can be:
 - Register indirect
 - Memory indirect

Example 1: L 2,0(3) ; in IBM360 family

Here, register 2 (indicated in the IF) is loaded with a word in main memory whose address is in register 3.

Example 2: call far ptr [p1]; in 8086/88 processor

A procedure is called whose address is stored in a word stored in location p1. p1 is the symbolic address of the pointer included in the IF, i.e. (M_{p1}) = address of procedure.

Almost all machines employ indirect addressing modes.

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

18

Addressing modes

2. Indirect addressing:

Indirect addressing can be of 1, 2, or more levels.

The 1-level indirect addressing uses one indirection, like we saw in the previous slide.

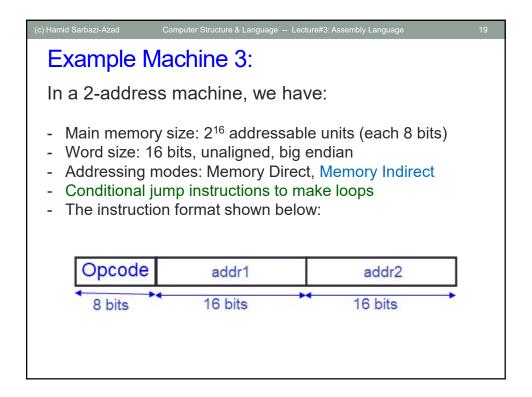
ightarrow M_(M addr) is the actual operand address.

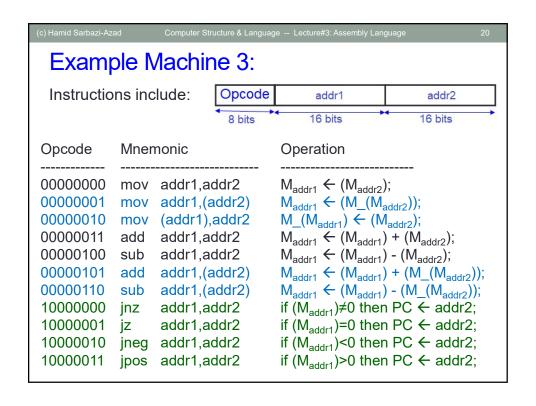
The 2-level indirect addressing use two consecutive indirection to get the actual operand.

ightarrow M_{(M_(M_addr))} is the actual operand address.

... and so on.

Most machines with indirect addressing use one level indirection; very few use two levels of indirection.





Example Machine 3:

· What is the size of machine registers?

 L_{MAR} = 16 bits; L_{MBR} = 16 bits; L_{PC} = 16 bits;

 Write an assembly program to calculate the sum of first 100 elements of alternating Fibonacci sequence as:

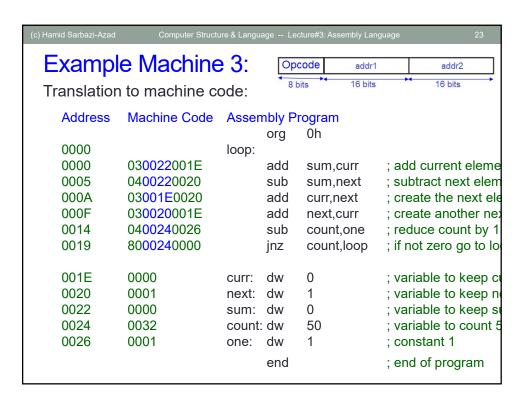
0, -1, 1, -2, 3, -5, 8, -13, 21, -34, ...

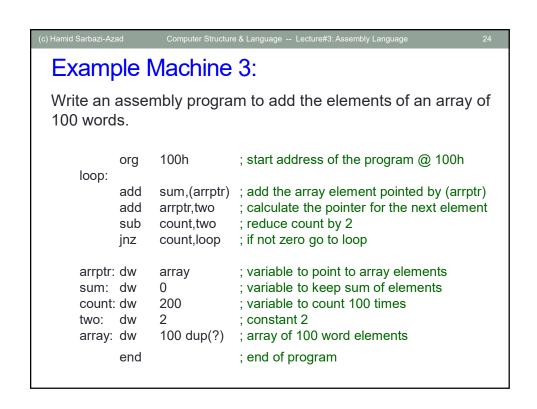
- Translate your assembly program to machine language.
- Write a program to add the elements of an array of 100 words.
- Translate your assembly program to machine language.

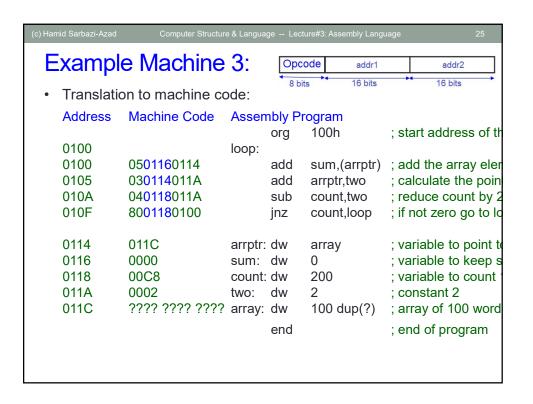
Example Machine 3:

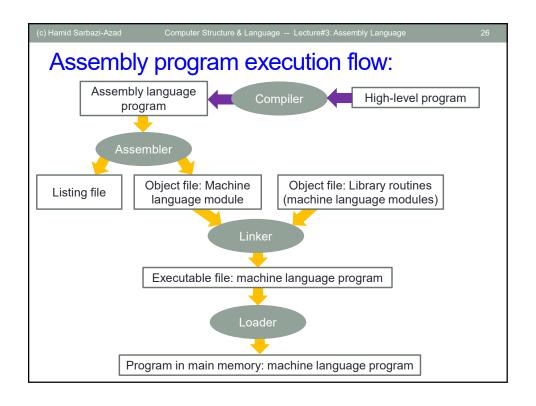
Write an assembly program to sum up the first 100 elements of series 0, -1, 1, -2, 3, -5, 8, -13, 21, -34,

```
0
      org
loop:
                         ; add current element to sum
      add
            sum,curr
            sum,next
                         ; subtract next element from sum
      sub
      add
            curr,next
                         ; create the next element in curr
                         : create another next element in next
      add
            next,curr
      sub
            count,one
                         ; reduce count by 1
      jnz
            count,loop ; if not zero go to loop
curr: dw
            0
                          ; variable to keep current element
                          ; variable to keep next element
next: dw
            1
            0
                          ; variable to keep sum of elements
sum: dw
                          : variable to count 50 times
count: dw
            50
one: dw
             1
                          ; constant 1
      end
                          ; end of program
```









Computer Structure & Language -- Lecture#3: Assembly Language

27

Assembler:

Assembler converts assembly language into object files. It also creates listing files that contains readable text.

Object files contains a combination of machine instructions, data, and information needed to place program instructions and data properly in memory.

Most Assemblers work in two passes:

Pass 1: Reads each line of the source program and records all labels in a Symbol Table.

Pass 2: Uses information in the Symbol Table to produce the actual machine code for each line of source program.

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

28

Object file format:

Object file Text Data header segment	Relocation information	Symbol table	Debugging information
--------------------------------------	------------------------	-----------------	-----------------------

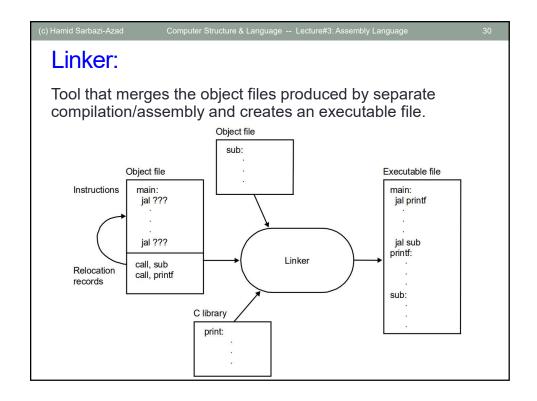
- Object file header describes the size and position of the other pieces of the file
- Text segment contains the machine instructions
- Data segment contains binary representation of data in assembly file
- Relocation info identifies instructions and data that depend on absolute addresses
- Symbol table associates addresses with external labels and lists unresolved references
- · Debugging info



A tool that merges the object files produced by separate compilation/assembly and creates an executable file.

It has three tasks:

- 1. Searches the program to find library routines used by program, e.g. printf(), scanf(), math routines,
- 2. Determines the memory locations that code from each module will occupy and relocates its instructions by adjusting absolute references
- 3. Resolve references among files



Computer Structure & Language -- Lecture#3: Assembly Language

31

Loader:

A part of Operating System that brings an executable file residing on disk into memory and starts its running.

It is done through the following steps:

- 1. Read executable file's header to determine the size of text and data segments;
- 2. Creates (via OS) a new address space for the program;
- 3. Copies instruction and data into allocated address space;
- 4. Copies arguments passed to the program on the stack;
- 5. Initialize the machine registers including stack pointer;
- 6. Jumps to a startup routine that copies the program's arguments from the stack to registers and calls the program's main routine by loading the PC.

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

32

Addressing modes

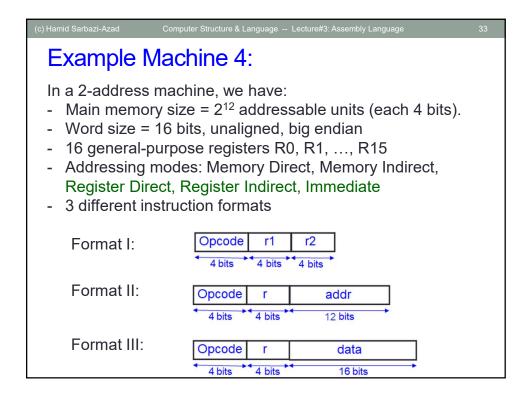
- 1. Direct
- 2. Indirect
- **3. Immediate**: The operand is included in the IF (no register/memory access is required).

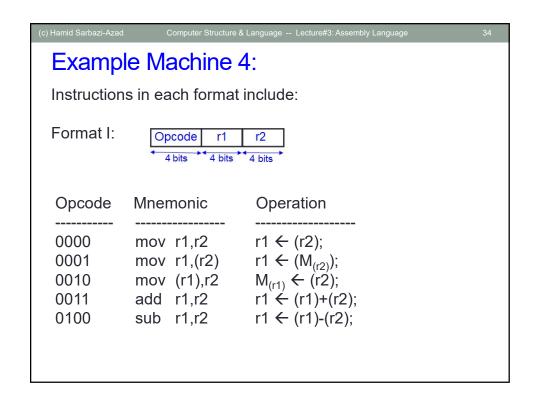
Example 1: sub cx,33 ; in 8086/88 processor

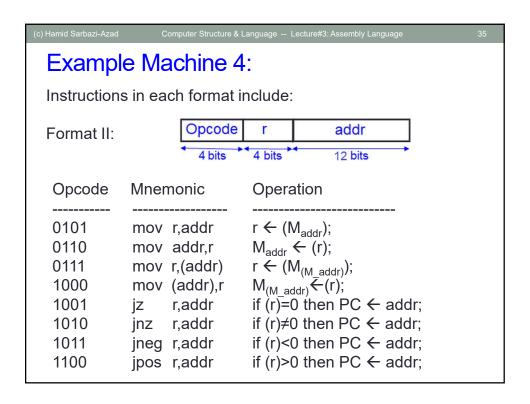
Here, 33 is an immediate data which is reduced from the content of cx register.

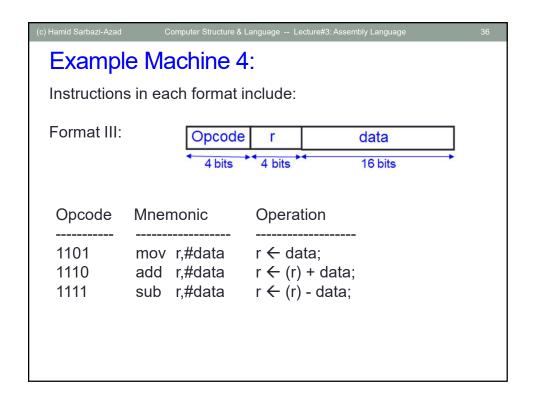
Example 2: addi \$s0,255 ; in MIPS processor Here, 255 is added to register \$s0 as an immediate data.

All available machines employ immediate addressing modes.









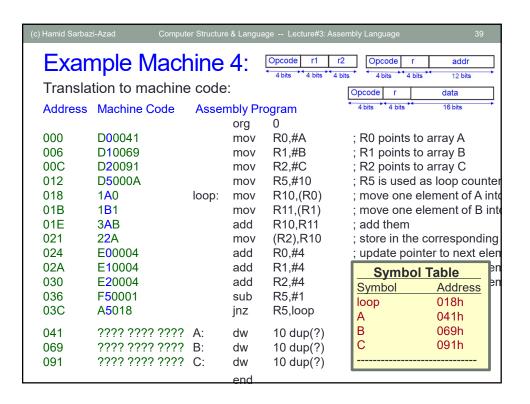
Example Machine 4:

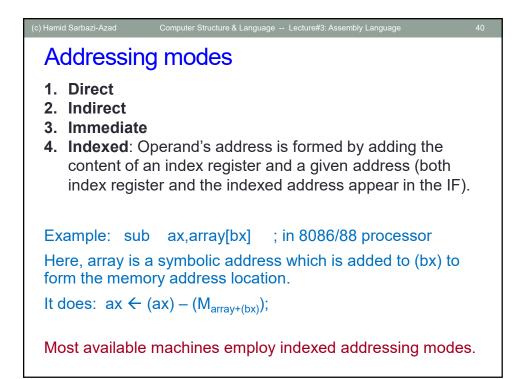
· What are the lengths of machine registers?

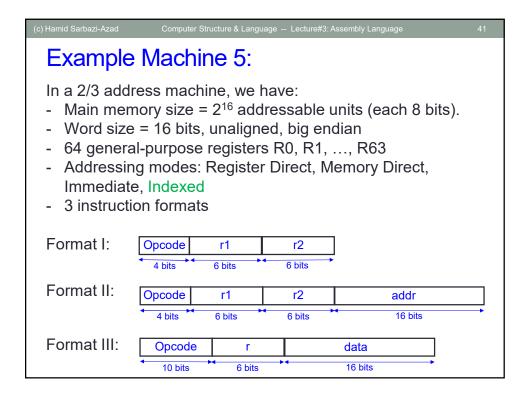
MAR = ? MBR = ? IR = ? PC = ? R0...R15 = ? 12 bits 16 bits 24 bits 12 bits 16 bits

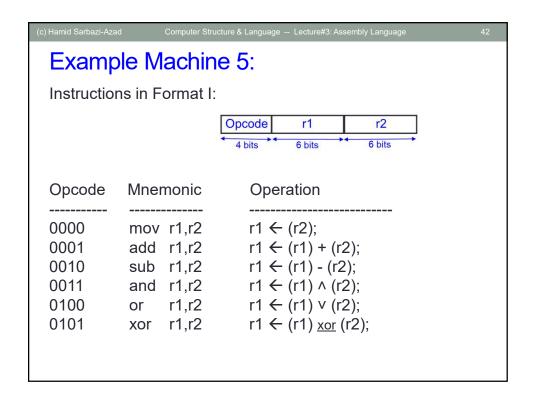
- Write a program to add two 10-element arrays A and B and save the resulted vector in 10-element array C.
- Translate your program to machine code.

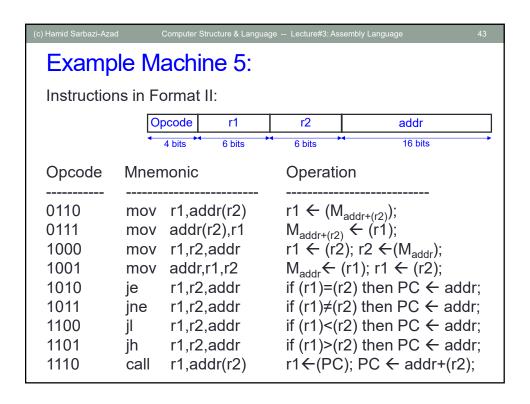
```
Example Machine 4:
Write a program to implement vector add C \leftarrow A + B;
                org
                       R0,#A
                                     ; R0 points to array A
                mov
                mov
                       R1,#B
                                     ; R1 points to array B
                       R2,#C
                mov
                                     ; R2 points to array C
                mov
                       R5,#10
                                     ; R5 is used as loop counter
          loop:
                       R10,(R0)
                                     ; move one element of A into R10
                mov
                       R11,(R1)
                                     ; move one element of B into R11
                mov
                add
                       R10,R11
                                     ; add them
                       (R2),R10
                mov
                                     ; store in the corresponding element of C
                add
                       R0,#4
                                     ; update pointer to next element of A
                add
                       R1,#4
                                     ; update pointer to next element of B
                add
                       R2,#4
                                     ; update pointer to next element of C
                                     ; decrement counter
                sub
                       R5,#1
                       R5,loop
                                     ; if not zero go to loop
                jnz
         A:
                dw
                       10 dup(?)
          B:
                dw
                       10 dup(?)
                       10 dup(?)
          C:
                dw
                end
```

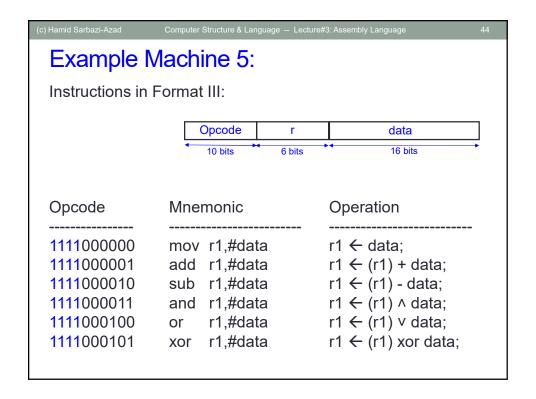












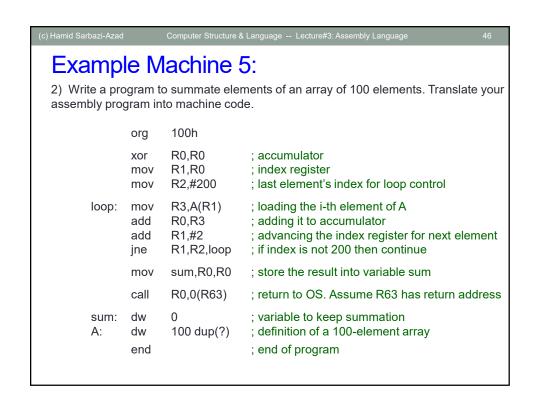
Example Machine 5:

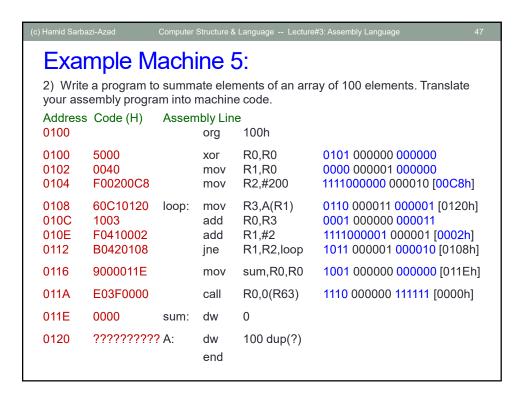
1) Determine the length of machine registers.

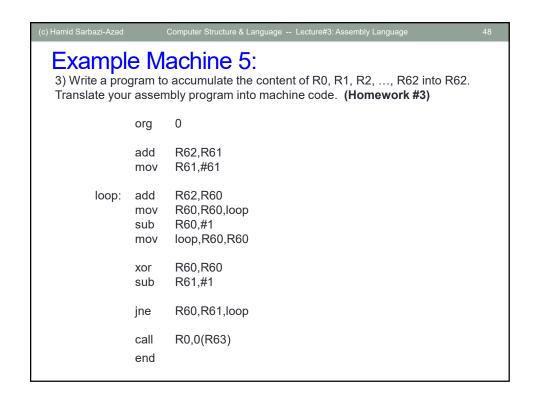
Answer:

```
L_{MAR} = 16 bits L_{MBR} = 16 bits L_{PC} = 16 bits L_{R0..R63} = 16 bits
```

- 2) Write a program to summate elements of an array of 100 elements. Translate your assembly program into machine code.
- 3) Write a program to accumulate the content of R0, R1, ..., R62 into R62. Translate your assembly program into machine code. (**Homework #3**)







Addressing modes

1. Direct
2. Indirect
3. Immediate
4. Indexed
5. Implied/Inherent: Nothing in IF but CPU knows where is the operand!

Example 1: loop instruction in 80806/88 processor.

loop address; $cx \leftarrow (cx)-1$;
; if $(cx)\neq 0$ then PC \leftarrow address;

Example 2: Multiply register instruction in IBM360 family.

MR 2,4; $R_2:R_3 \leftarrow (R_3) \times (R_4)$

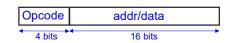
One-address Machine

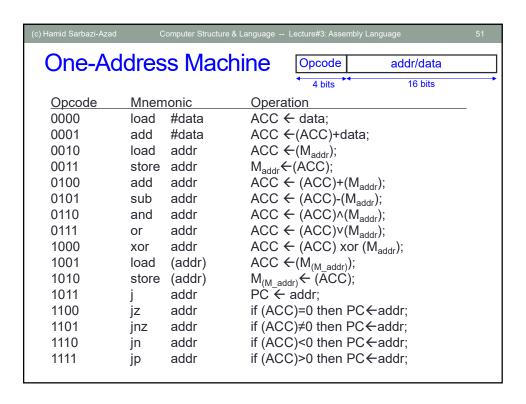
One operand is kept in a special purpose register named ACC (Accumulator) and addressed inherently.

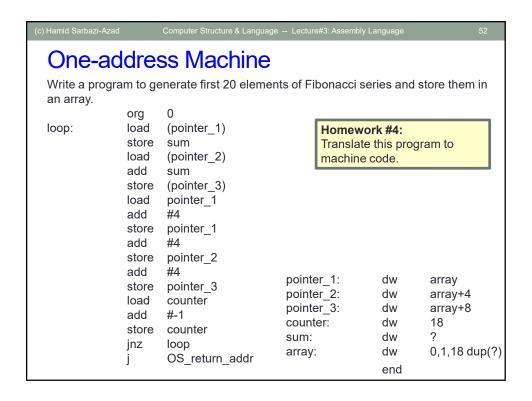
Example: In a one-address machine:

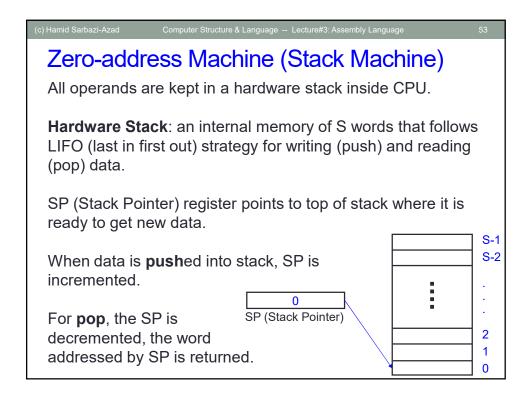
- main memory size = 2¹⁶ addressable units (each 4 bits)
- word size is 16 bits (unaligned, big endian)
- Implied, direct, indirect, and immediate addressing modes

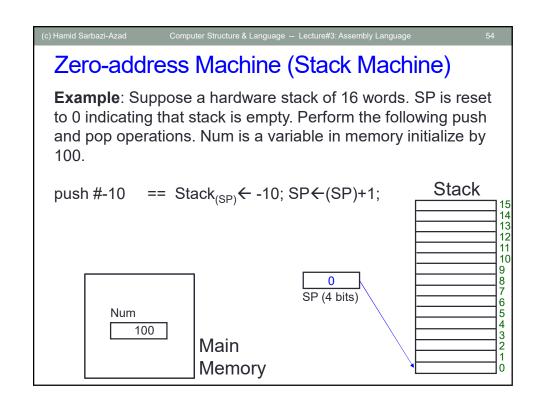
Instructions are coded in one format.

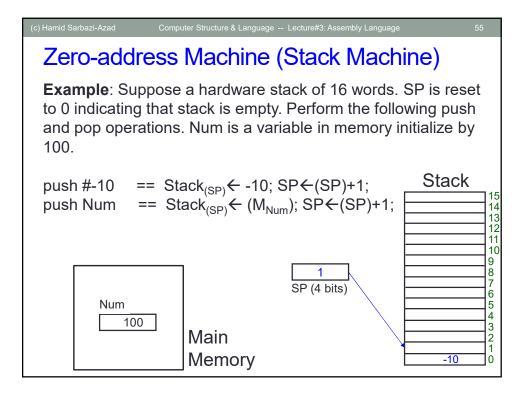


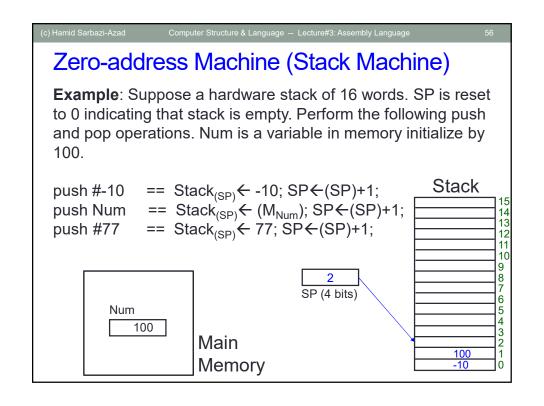


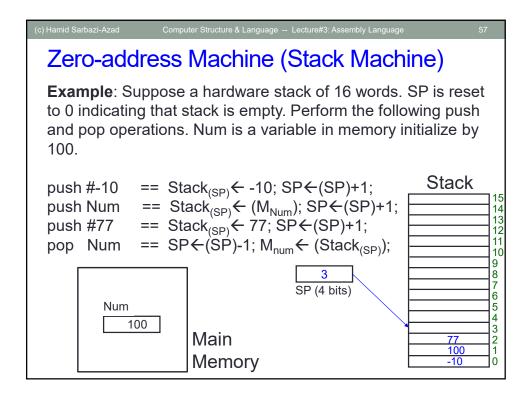


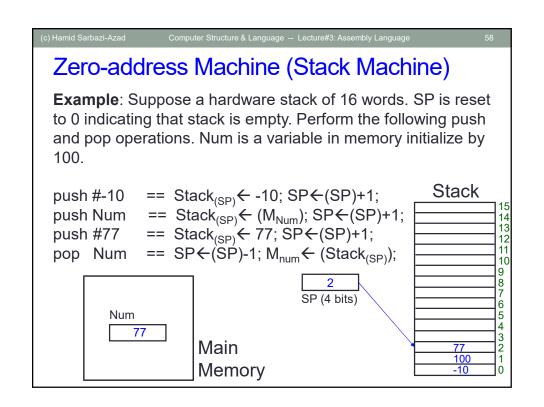


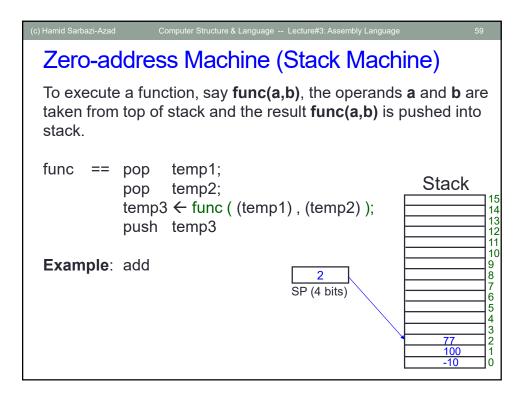


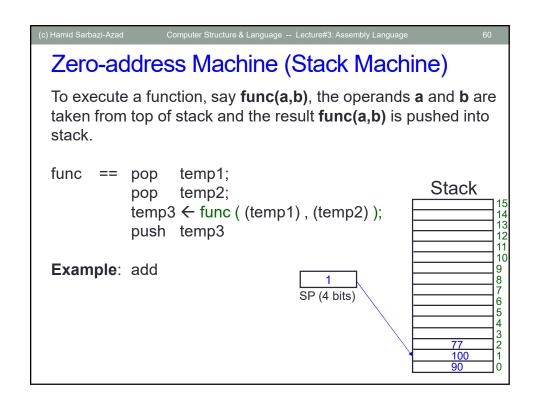












Polish Notation and Stack Machines

The normal notation we use in maths is best known as infix notation where the operator comes between operands.

Example: x + y or z * t

$$\chi + \gamma$$

But alternative notation can be used to write mathematical expressions, e.g. prefix notation as:

+xy or *zt

Postfix notation (so called Polish notation) as:

Polish Notation and Stack Machines

Example: Write the following expression in Polish notation:

$$x + y - 10 - t / d + 10*s - z$$

$$20*t - z / t + 100*w$$

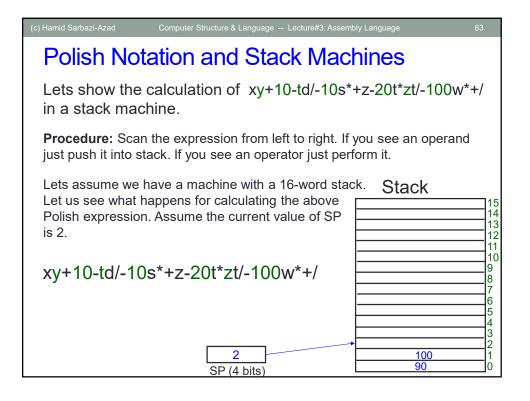
Solution:

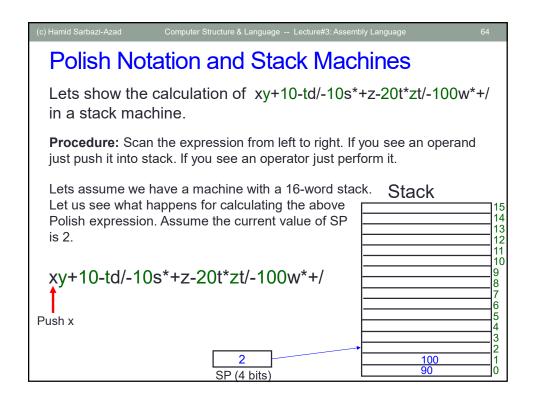
Positive point:

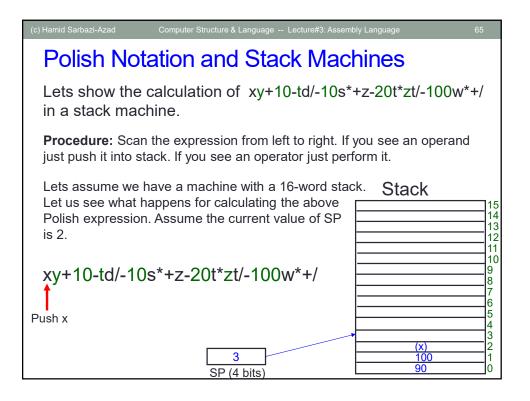
- No need to worry about priority of operations. No parentheses.
- Suitable for calculation on a stack machine.

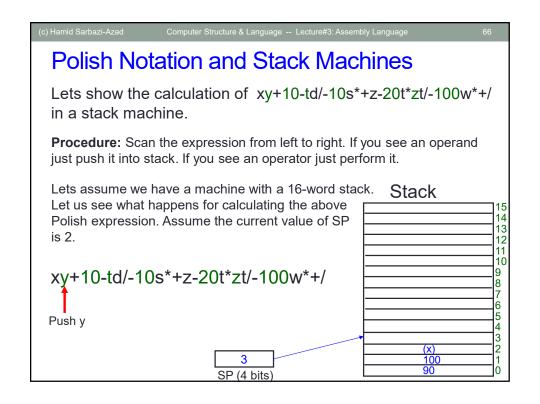
Negative point:

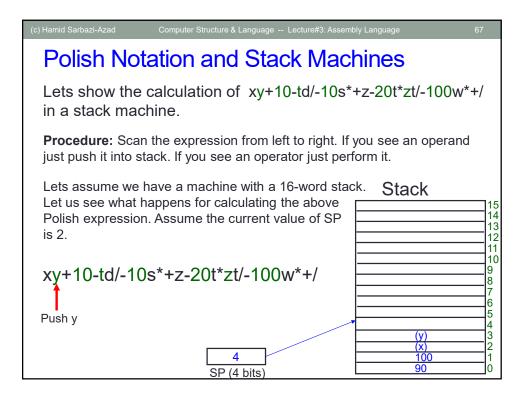
- Not good for human perception.

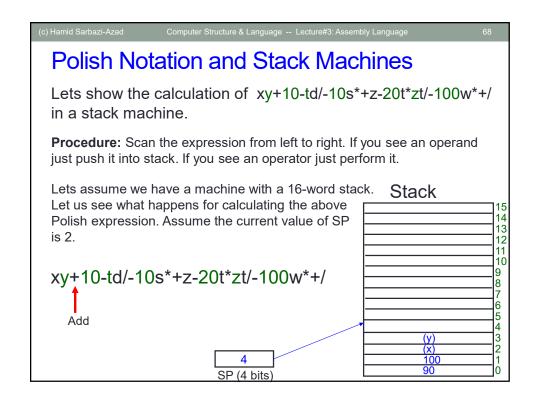


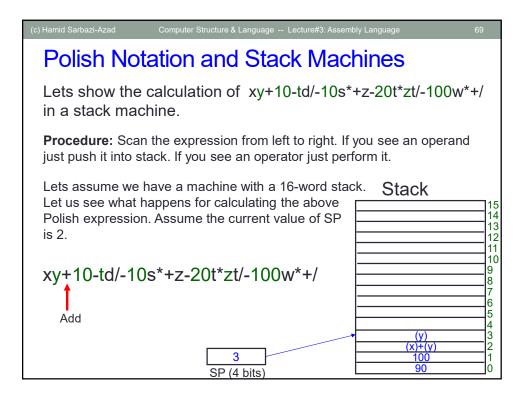


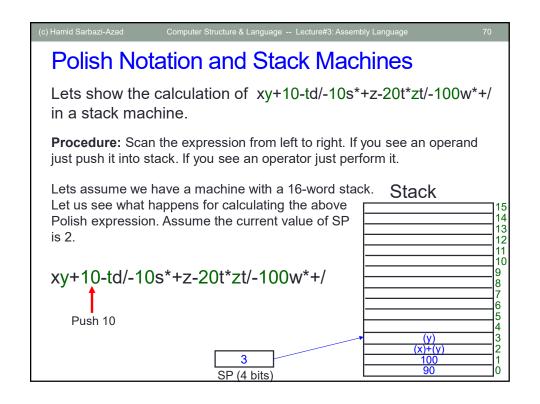


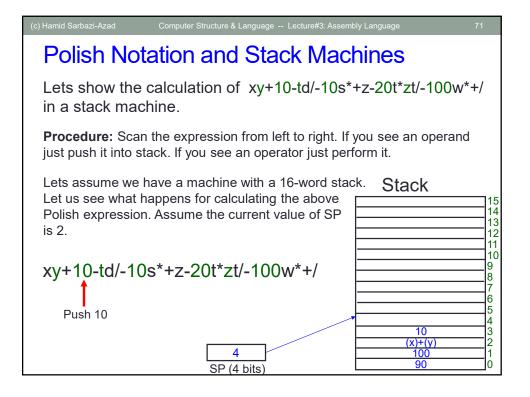


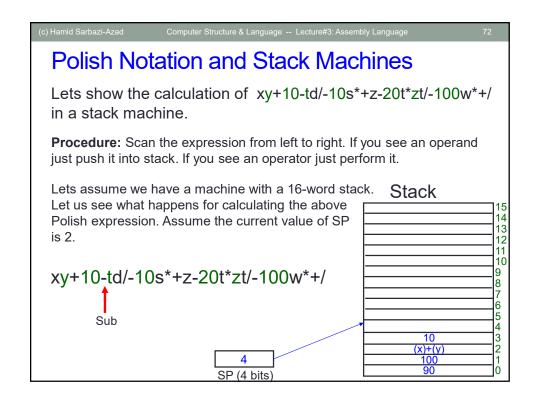


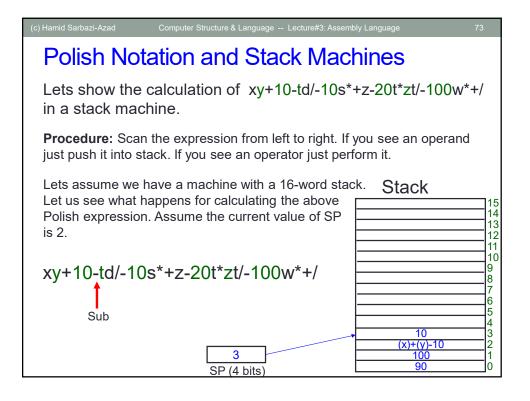


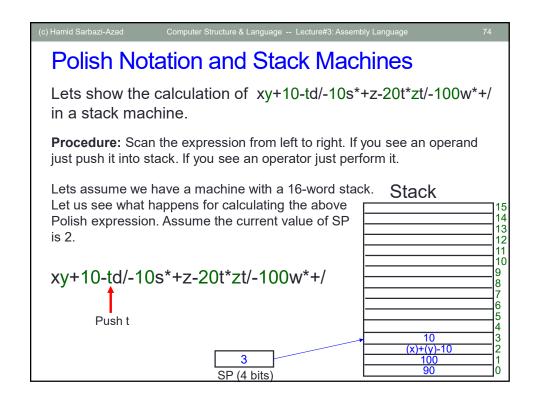


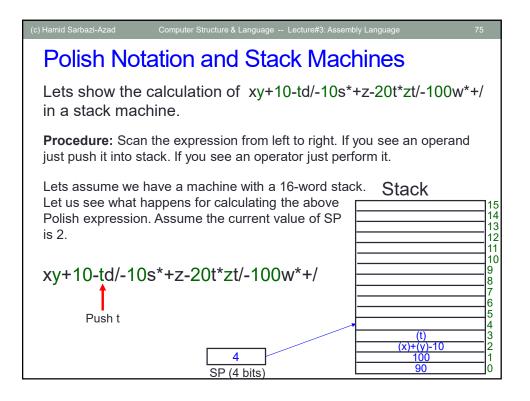


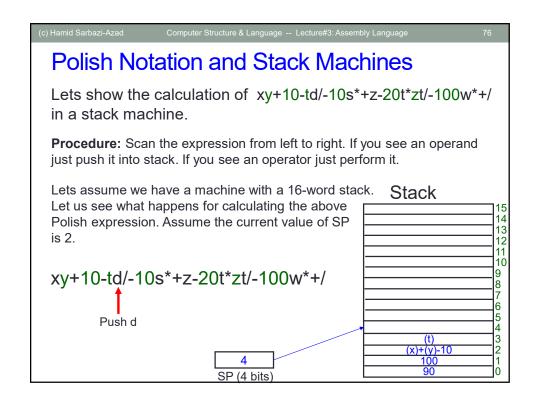


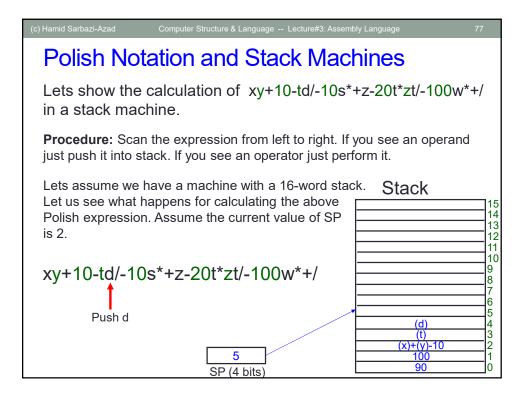


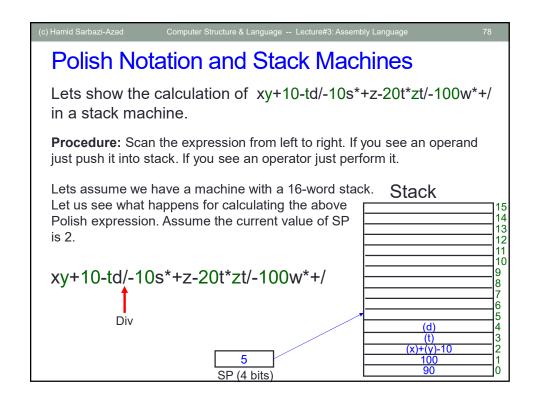


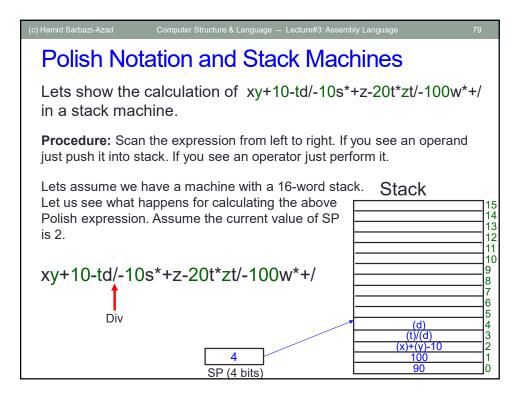


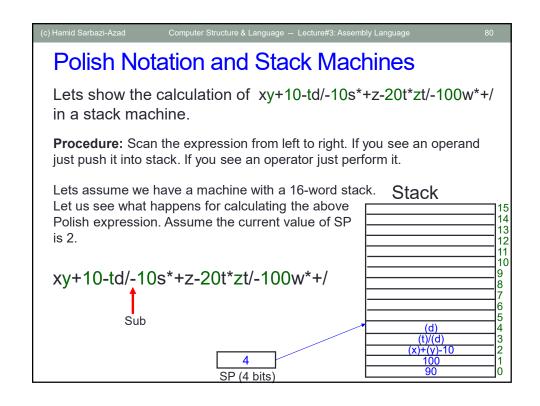


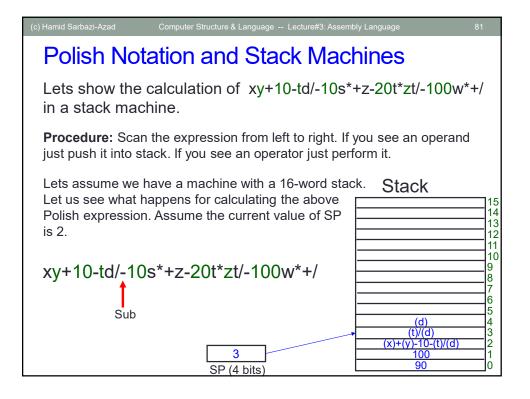


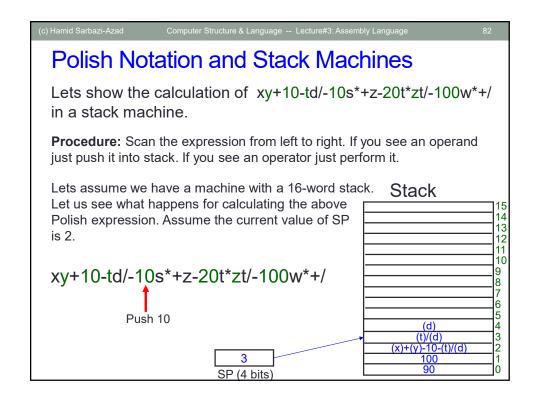


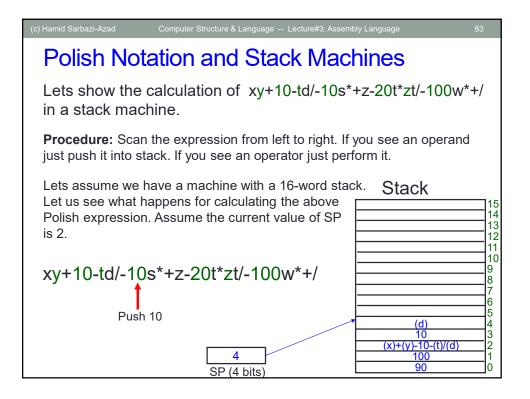


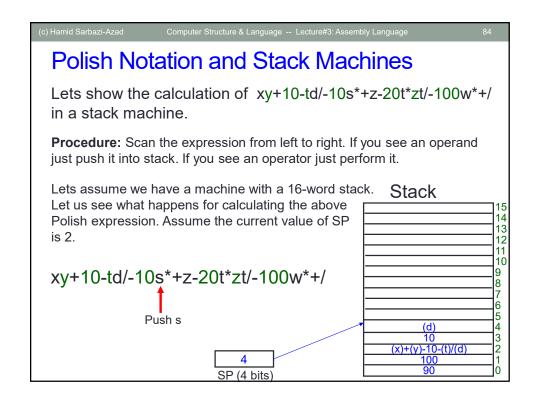


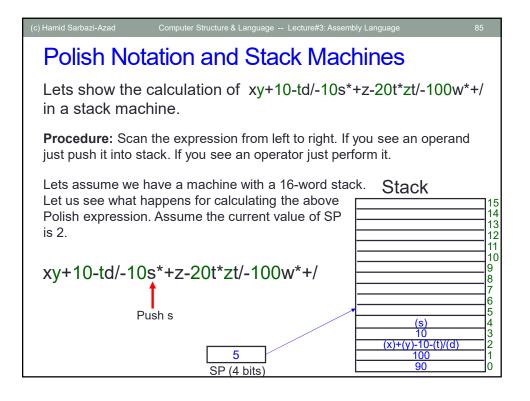


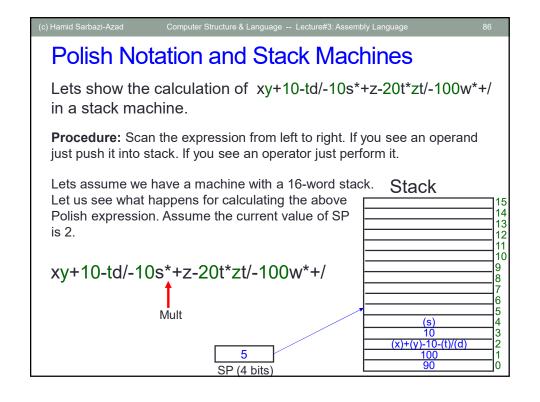


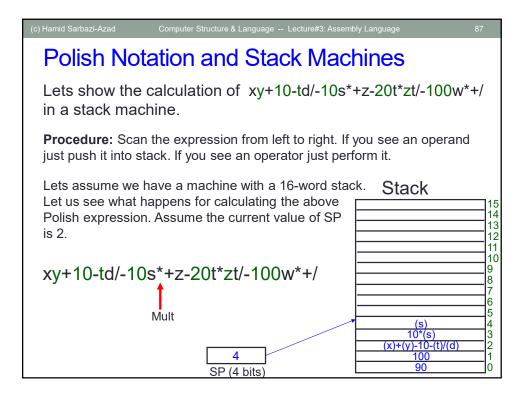


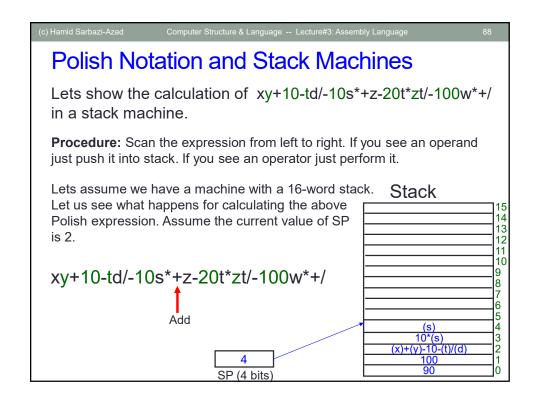


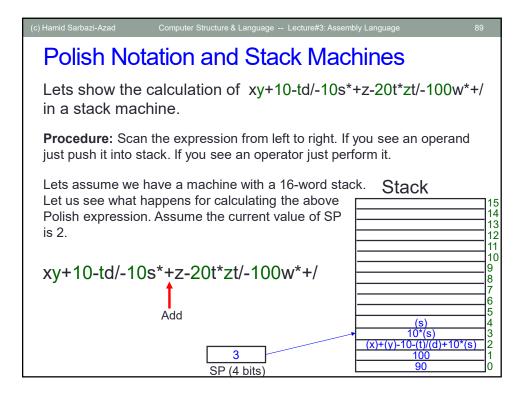


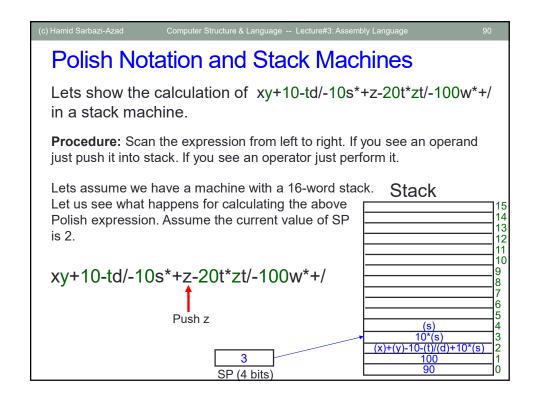


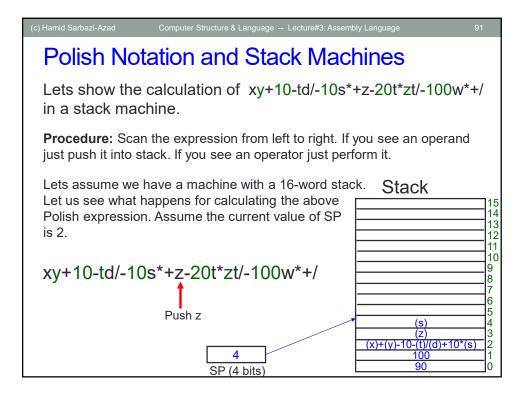


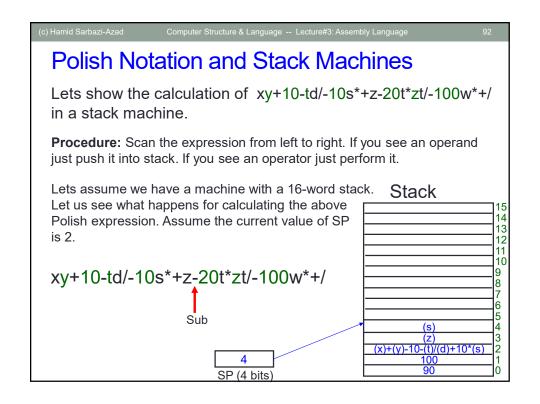


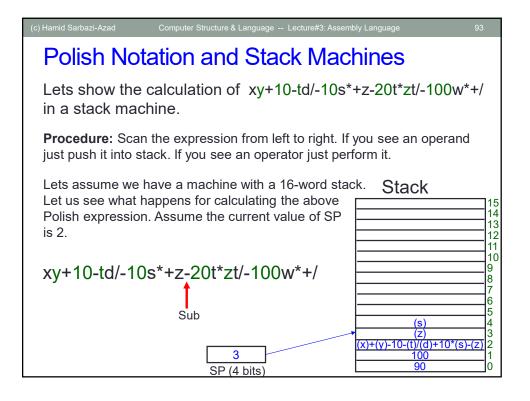


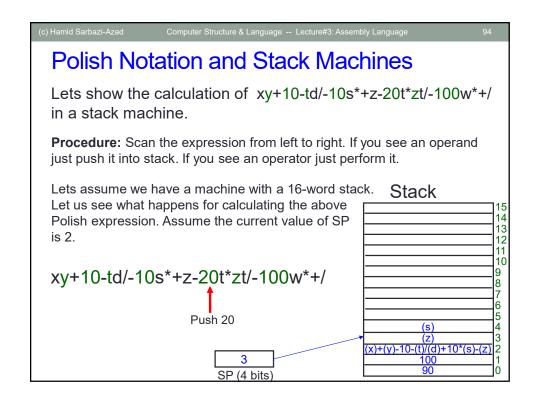


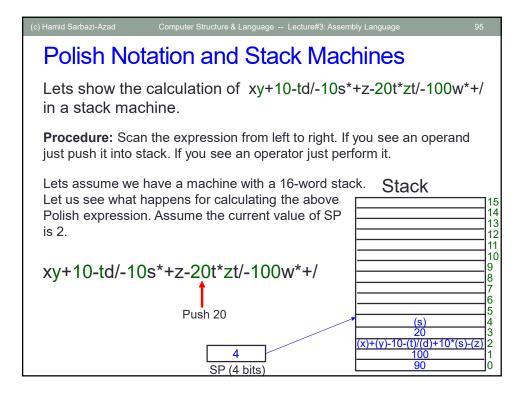


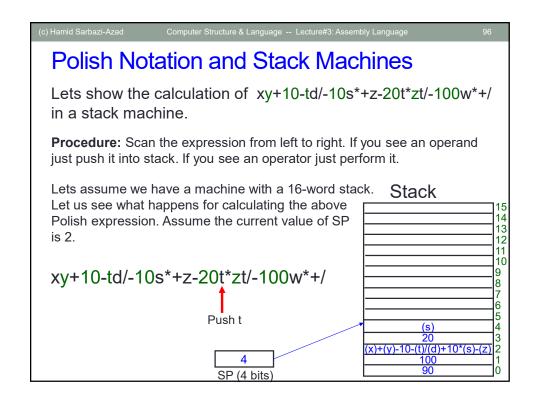


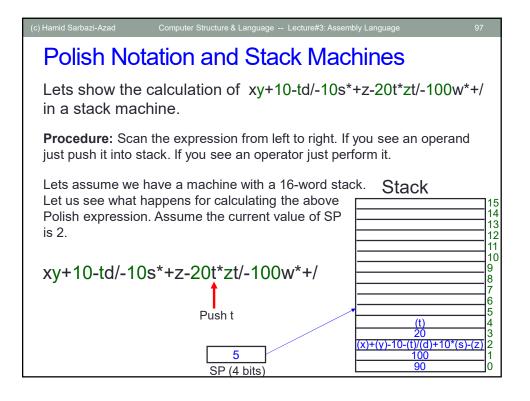


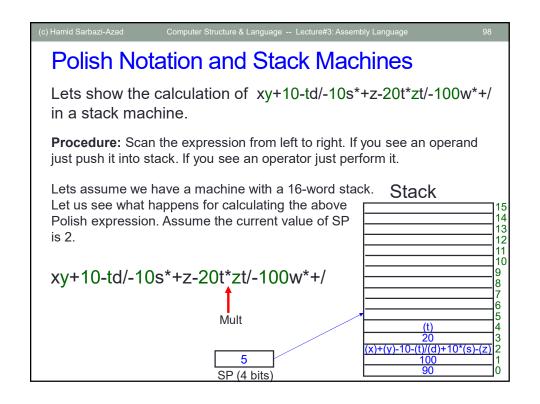


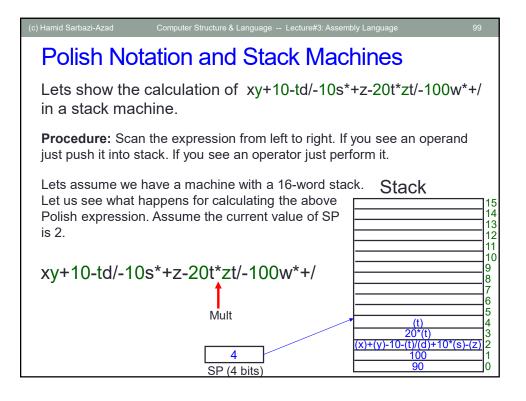


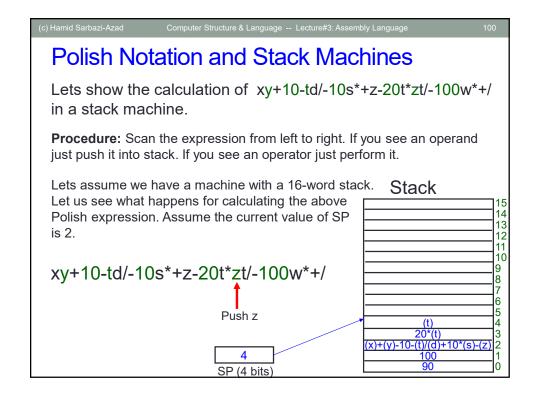


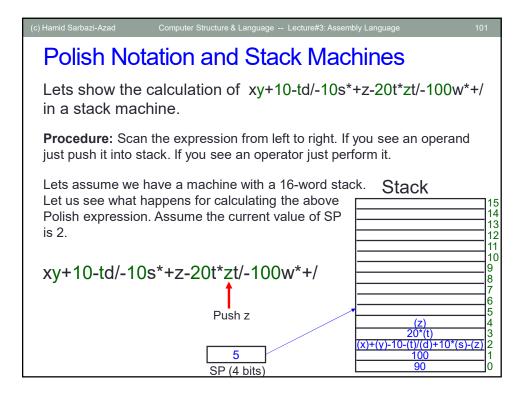


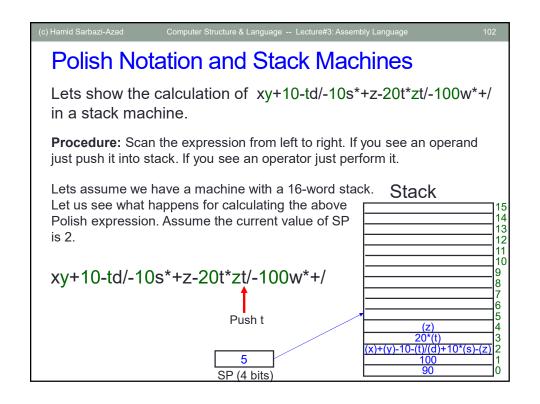


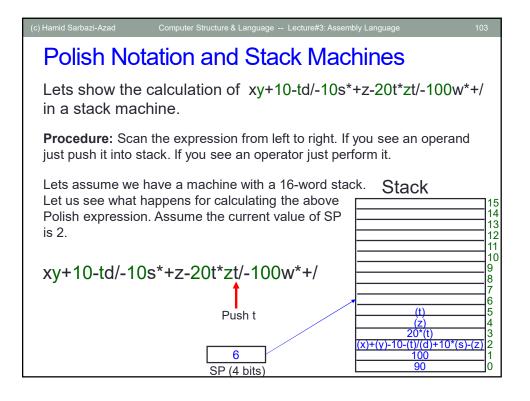


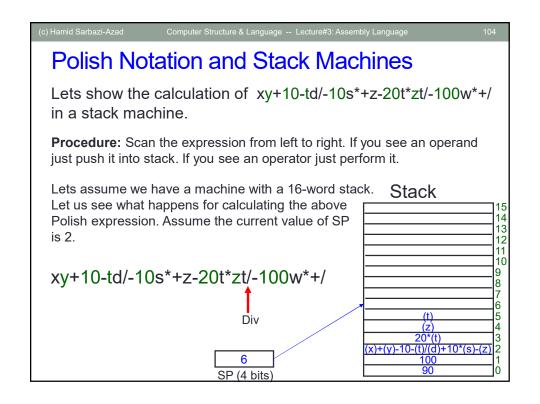


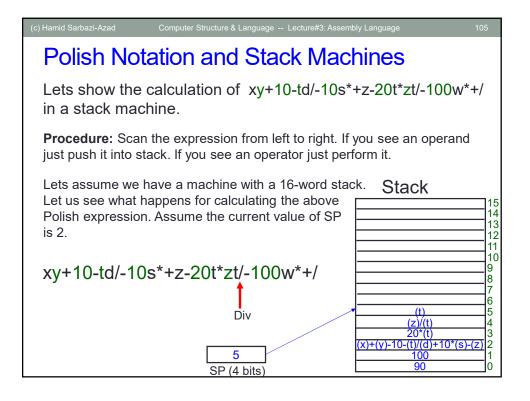


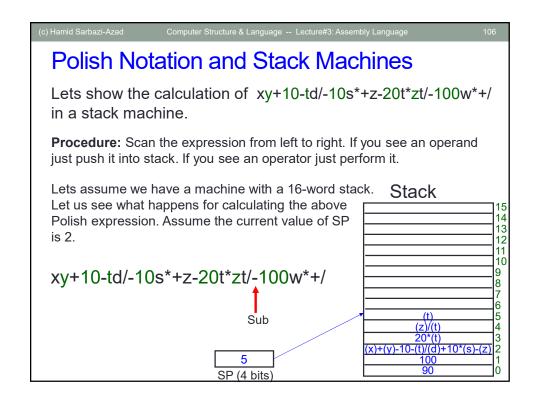


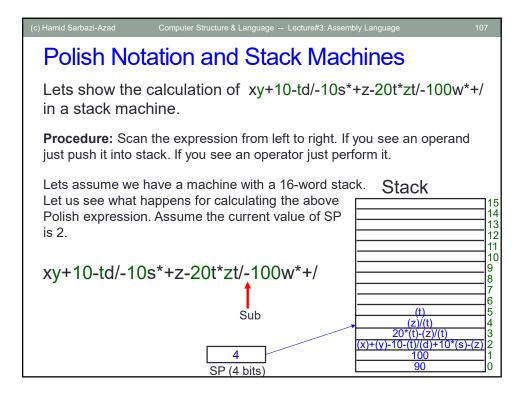


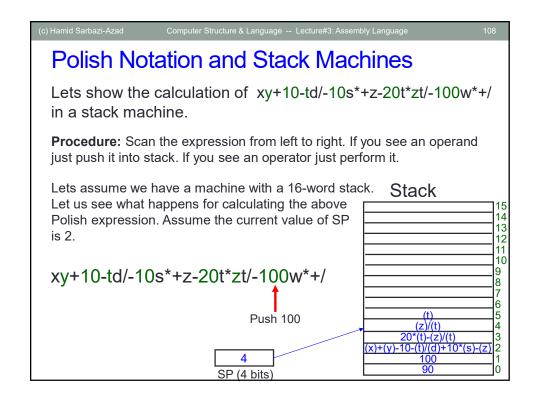


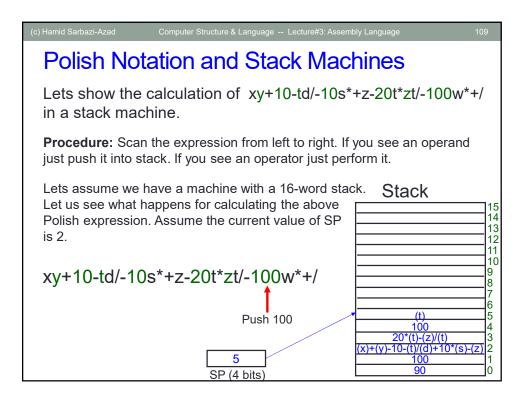


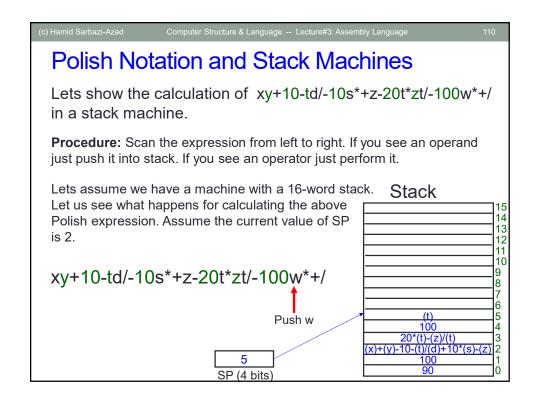


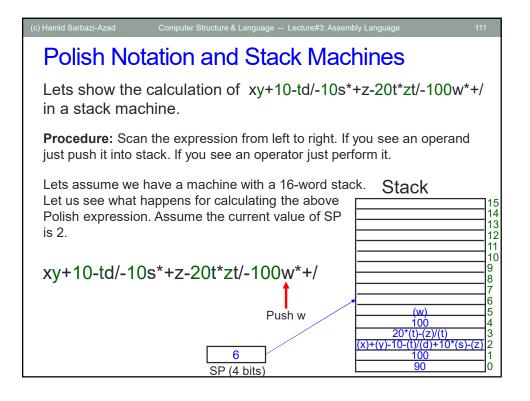


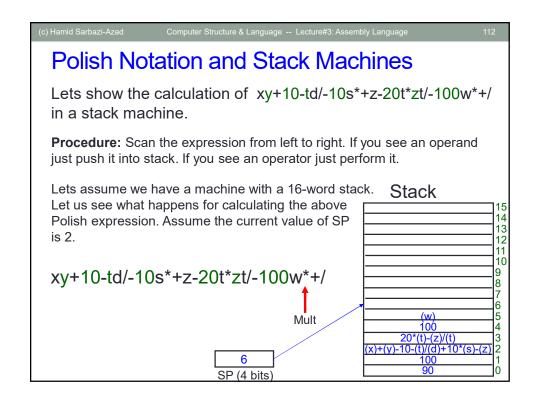


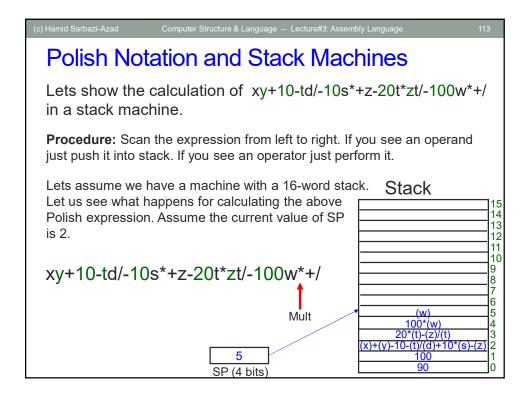


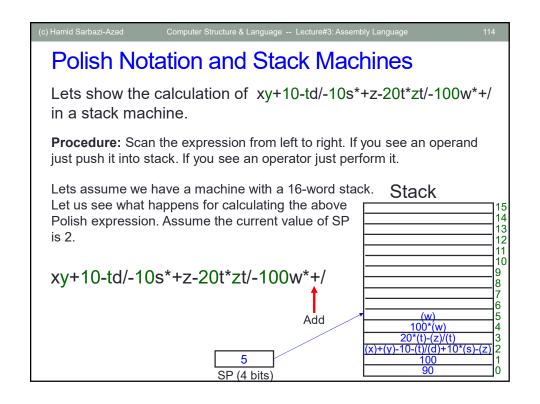


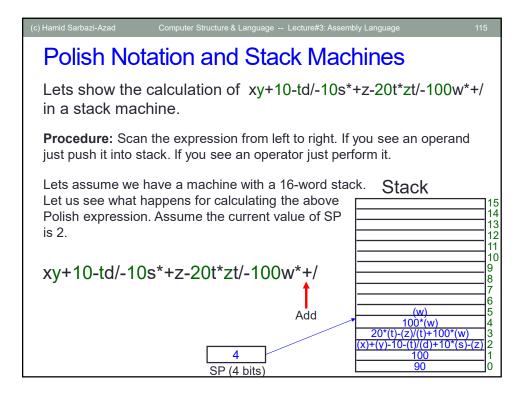


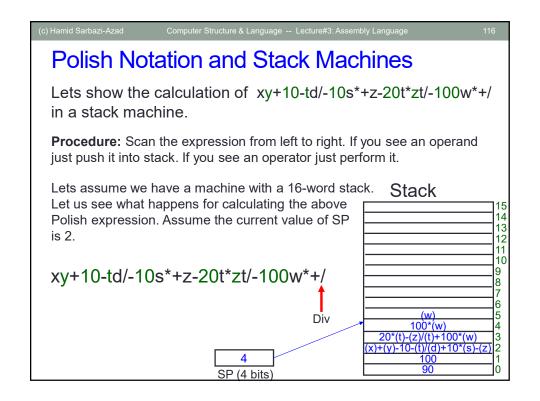


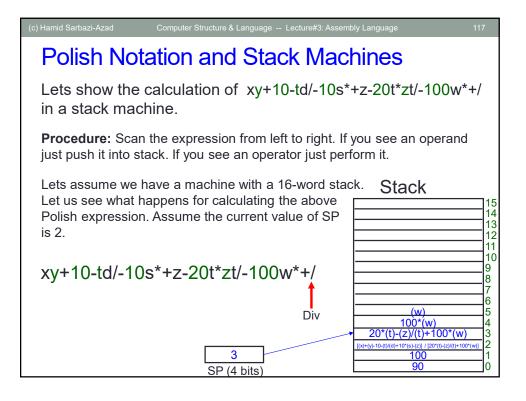


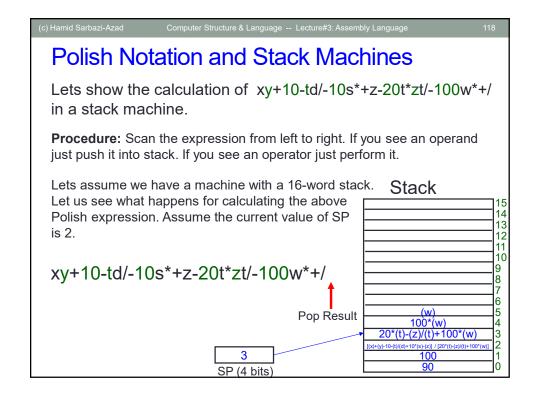


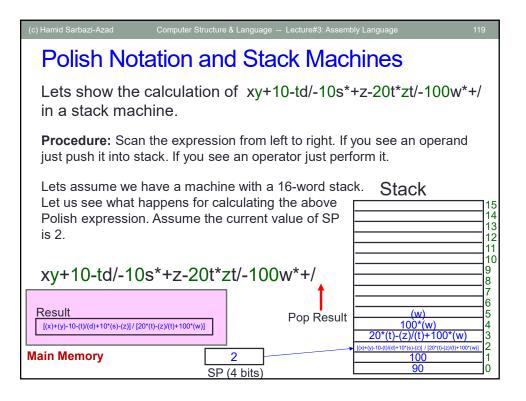


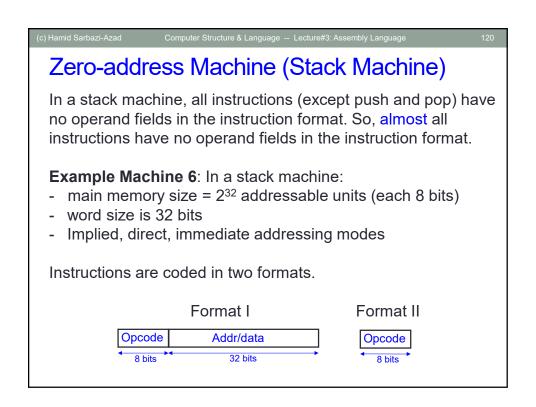


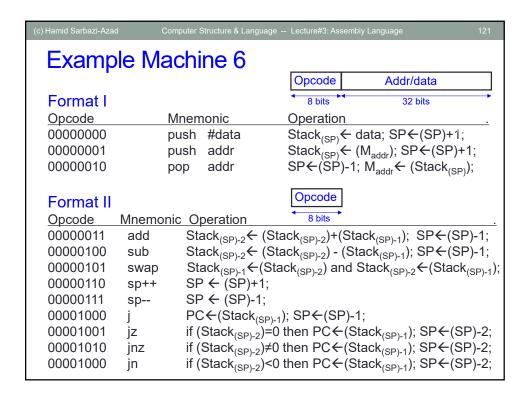


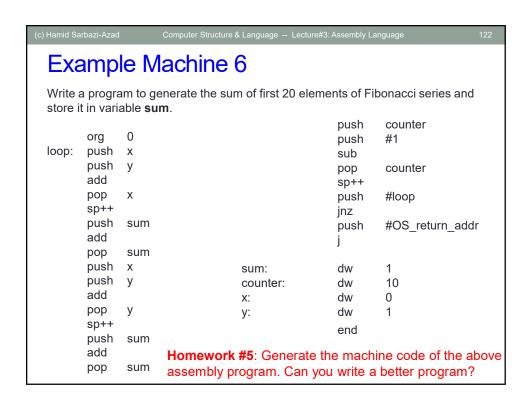












c) Hamid Sarbazi-Aza

Computer Structure & Language -- Lecture#3: Assembly Language

123

Infix to Postfix Conversion

The algorithm includes two steps:

- 1. Fully parenthesize the input expression.
- 2. Starting from inner pairs of parentheses, replace the right parenthesis with operator and delete the left parenthesis.

Example: Convert below expression into Polish notation:

$$x + y - 10 - t / d + 10*s - z$$

-----20*t - z / t + 100*w

(c) Hamid Sarbazi-Azad

Computer Structure & Language -- Lecture#3: Assembly Language

124

Infix to Postfix Conversion

Example: Convert below expression into Polish notation:

$$\frac{x + y - 10 - t / d + 10*s - z}{20*t - z / t + 100*w}$$

Step 1:

```
Infix to Postfix Conversion

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

(((((xy+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))
```

```
Infix to Postfix Conversion

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

(((((xy+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+y)-10)-(t/d))+((10*s-z)) / (((20*t-zt/)+100w*)))
```

```
Infix to Postfix Conversion

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / ((20t*-zt/)+100w*))

(((xy+10--td/)+10s*z-) / (20t*zt/-+100w*))
```

```
Infix to Postfix Conversion

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / ((20t*-zt/)+100w*))

(((xy+10--td/)+10s*z-) / (20t*zt/-+100w*))
```

```
Infix to Postfix Conversion

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / (((20t*-zt/)+100w*)))

(((xy+10-td/)+10s*z-) / (20t*zt/-+100w*))

(((xy+10-td/-+10s*z-)) / 20t*zt/-100w*+)
```

```
Infix to Postfix Conversion

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / (((20t*-zt/)+100w*))

(((xy+10-td/)+10s*z-) / (20t*zt/-+100w*))

(((xy+10-td/-+10s*z-) / 20t*zt/-100w*+)
```

```
Infix to Postfix Conversion

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / (((20t*-zt/)+100w*))

(((xy+10-td/)+10s*z-) / (20t*zt/-100w*+)

((xy+10-td/-10s*z-+/20t*zt/-100w*+)
```

```
Infix to Postfix Conversion

((((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / ((20t*-zt/)+100w*))

(((xy+10-td/)+10s*z-) / (20t*zt/-+100w*))

((xy+10-td/-+10s*z-) / 20t*zt/-100w*+)

(xy+10-td/- 10s*z-+/ 20t*zt/-100w*+)
```

```
Infix to Postfix Conversion

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

Step 2:

(((((x+y)-10)-(t/d))+((10*s)-z)) / (((20*t)-(z/t))+(100*w)))

((((xy+-10)-td/)+(10s*-z)) / (((20*t-zt/)+100w*))

(((xy+10-td/)+10s*z-) / (20t*zt/-100w*))

((xy+10-td/-10s*z-+/20t*zt/-100w*+)

xy+10-td/-10s*z-+20t*zt/-100w*+/
```

END OF SLIDES