Chapter 4: Data Transfers, Addressing, and Arithmetic

CS 238 Assembly Language Programming
Amarnath Jasti

Chapter Overview

- Data Transfer Instructions
- Addition and Subtraction
- Data-Related Operators and Directives
- Indirect Addressing
- JMP and LOOP Instructions

Data Transfer Instructions

- Operand Types
- Instruction Operand Notation
- Direct Memory Operands
- MOV Instruction
- Zero & Sign Extension
- XCHG Instruction
- Direct-Offset Instructions

Operand Types

- Three basic types of operands:
  - Immediate – a constant integer (8, 16, or 32 bits)
    - value is encoded within the instruction
  - Register – the name of a register
    - register name is converted to a number and encoded within the instruction
  - Memory – reference to a location in memory
    - memory address is encoded within the instruction, or a register holds the address of a memory location

Instruction Operand Notation

Operand	Description
r8	8-bit general-purpose register: AH, AL, BH, BL, CH, CL, DH, DL
r16	16-bit general-purpose register: AX, BX, CX, DX, SI, DI, SP, BP
r32	32-bit general-purpose register: EAX, EBX, ECX, EDX, ESI, EDI, ESP, EBP
reg	any general-purpose register
sreg	16-bit segment register: CS, DS, SS, ES, FS, GS
imm	8-, 16-, or 32-bit immediate value
imm8	8-bit immediate byte value
imm16	16-bit immediate word value
imm32	32-bit immediate doubleword value
r/m8	8-bit operand which can be an 8-bit general register or memory byte
r/m16	16-bit operand which can be a 16-bit general register or memory word
r/m32	32-bit operand which can be a 32-bit general register or memory doubleword
mem	an 8-, 16-, or 32-bit memory operand

Direct Memory Operands

- A direct memory operand is a named reference to storage in memory
- The named reference (label) is automatically dereferenced by the assembler

```
.data
var1 BYTE 10h
.code
mov al,var1                ; AL = 10h ; MASM only
mov al,[var1]              ; AL = 10h
```

alternate format

MOV Instruction

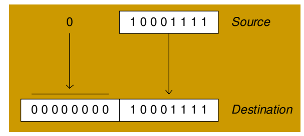
- Move from source to destination. Syntax:
  MOV *destination, source*
- No more than one memory operand permitted
- CS, EIP, and IP cannot be the destination
- No immediate to segment moves

```
.data
count BYTE 100
wVal  WORD 2
.code
    mov bl,count
    mov ax,wVal
    mov count,al

    mov al,wVal                    ; error
    mov ax,count                   ; error
    mov eax,count                  ; error
```


Zero Extension

When you copy a smaller value into a larger destination, the MOVZX instruction fills (extends) the upper half of the destination with zeros.

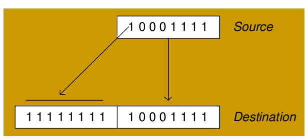



```
mov bl,10001111b
movzx ax,bl                ; zero-extension
```

The destination must be a register.

Sign Extension

The MOVSX instruction fills the upper half of the destination with a copy of the source operand's sign bit.




```
mov bl,10001111b
movsx ax, bl               ; sign extension
```

The destination must be a register.

XCHG Instruction

XCHG exchanges the values of two operands. At least one operand must be a register. No immediate operands are permitted.

```
.data
var1 WORD 1000h
var2 WORD 2000h
.code
xchg ax,bx                 ; exchange 16-bit regs
xchg ah,al                 ; exchange 8-bit regs
xchg var1,bx               ; exchange mem, reg
xchg eax,ebx               ; exchange 32-bit regs

xchg var1,var2             ; error: two memory operands
```


Direct-Offset Operands

A constant offset is added to a data label to produce an effective address (EA). The address is dereferenced to get the value inside its memory location.

```
.data
arrayB BYTE 10h,20h,30h,40h
.code
mov al,arrayB+1              ; AL = 20h
mov al,[arrayB+1]            ; alternative notation
```

Q: Why doesn't arrayB+1 produce 11h?

Direct-Offset Operands (cont)

A constant offset is added to a data label to produce an effective address (EA). The address is dereferenced to get the value inside its memory location.

```
.data
arrayW  WORD 1000h,2000h,3000h
arrayD  DWORD 1,2,3,4
.code
mov ax,[arrayW+2]               ; AX =   2000h
mov ax,[arrayW+4]               ; AX =   3000h
mov eax,[arrayD+4]              ; EAX =  00000002h
```

```
; Will the following statements assemble and run?
mov ax,[arrayW-2]          ; ??
mov eax,[arrayD+16]        ; ??
```

Addition and Subtraction

- INC and DEC Instructions
- ADD and SUB Instructions
- NEG Instruction
- Implementing Arithmetic Expressions
- Flags Affected by Arithmetic
  - Zero
  - Sign
  - Carry
  - Overflow

INC and DEC Instructions

- Add 1, subtract 1 from destination operand
  - operand may be register or memory
- INC *destination*
  - Logic: *destination* ← *destination* + 1
- DEC *destination*
  - Logic: *destination* ← *destination* – 1

INC and DEC Examples

```
.data
myWord  WORD 1000h
myDword DWORD 10000000h
.code
   inc myWord              ; 1001h
   dec myWord              ; 1000h
   inc myDword             ; 10000001h

   mov ax,00FFh
   inc ax                  ; AX = 0100h
   mov ax,00FFh
   inc al                  ; AX = 0000h
```


ADD and SUB Instructions

- ADD destination, source
  - Logic: *destination* ← *destination* + source
- SUB destination, source
  - Logic: *destination* ← *destination* – source
- Same operand rules as for the MOV instruction

ADD and SUB Examples

```
.data
var1 DWORD 10000h
var2 DWORD 20000h
.code                      ; ---EAX---
   mov eax,var1            ; 00010000h
   add eax,var2            ; 00030000h
   add ax,0FFFFh           ; 0003FFFFh
   add eax,1               ; 00040000h
   sub ax,1                ; 0004FFFFh
```


Evaluate this . . .

- We want to write a program that adds the following three bytes:

```
.data
myBytes BYTE 80h,66h,0A5h
```

- What is your evaluation of the following code?

```
mov al,myBytes
add al,[myBytes+1]
add al,[myBytes+2]
```

- What is your evaluation of the following code?

```
mov ax,myBytes
add ax,[myBytes+1]
add ax,[myBytes+2]
```

- Any other possibilities?

Evaluate this . . . (cont)

```
.data
myBytes BYTE 80h,66h,0A5h
```

- How about the following code. Is anything missing?

```
movzx ax,myBytes
mov   bl,[myBytes+1]
add   ax,bx
mov   bl,[myBytes+2]
add   ax,bx                  ; AX = sum
```

Yes: Move zero to BX before the MOVZX instruction.

NEG (negate) Instruction

Reverses the sign of an operand. Operand can be a register or memory operand.

```
.data
valB BYTE -1
valW WORD +32767
.code
   mov al,valB               ; AL = -1
   neg al                    ; AL = +1
   neg valW                  ; valW = -32767
```

Note: Actual register values are in HEX.

Suppose AX contains –32,768 and we apply NEG to it. Will the result be valid?

Implementing Arithmetic Expressions

HLL compilers translate mathematical expressions into assembly language. You can do it also. For example:

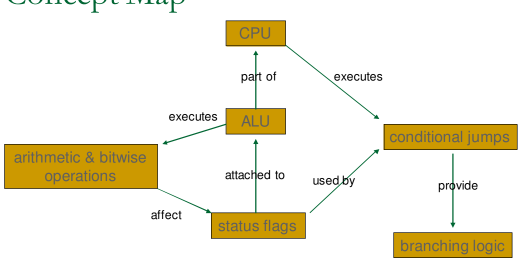
```
Rval = -Xval + (Yval – Zval)
```

```
Rval resw 1
Xval WORD 26
Yval WORD 30
Zval WORD 40
.code
   mov ax,Xval
   neg ax                    ; AX = -26
   mov bx,Yval
   sub bx,Zval               ; BX = -10
   add ax,bx
   mov Rval,ax               ; -36
```


Flags Affected by Arithmetic

- The ALU has a number of status flags that reflect the outcome of arithmetic (and bitwise) operations
  - based on the contents of the destination operand
- Essential flags:
  - Zero flag – destination equals zero
  - Sign flag – destination is negative
  - Carry flag – unsigned value out of range
  - Overflow flag – signed value out of range
- The MOV instruction never affects the flags.

Concept Map



You can use diagrams such as these to express the relationships between assembly language concepts.

Zero Flag (ZF)

Whenever the destination operand equals Zero, the Zero flag is set.

```
mov cx,1
sub cx,1                ; CX = 0, ZF = 1
mov ax,0FFFFh
inc ax                  ; AX = 0, ZF = 1
inc ax                  ; AX = 1, ZF = 0
```

A flag is set when it equals 1.

A flag is clear when it equals 0.

Sign Flag (SF)

The Sign flag is set when the destination operand is negative. The flag is clear when the destination is positive.

```
mov cx,0
sub cx,1                    ; CX = -1, SF = 1
add cx,2                    ; CX = 1, SF = 0
```

The sign flag is a copy of the destination's highest bit:

```
mov al,0
sub al,1              ; AL = 11111111b, SF = 1
add al,2              ; AL = 00000001b, SF = 0
```

29

Carry Flag (CF)

The Carry flag is set when the result of an operation generates an unsigned value that is out of range (too big or too small for the destination operand).

```
mov al,0FFh
add al,1                    ; CF = 1, AL = 00

; Try to go below zero:

mov al,0
sub al,1                    ; CF = 1, AL = FF
```

In the second example, we tried to generate a negative value. Unsigned values cannot be negative, so the Carry flag signaled an error condition.

30

Overflow Flag (OF)

The Overflow flag is set when the signed result of an operation is invalid or out of range.

```
; Example 1
mov al,+127
add al,1                 ; OF = 1,   AL = ??

; Example 2
mov al,7Fh               ; OF = 1,   AL = 80h
add al,1
```

The two examples are identical at the binary level because 7Fh equals +127. To determine the value of the destination operand, it is often easier to calculate in hexadecimal.

32

A Rule of Thumb

- When adding two integers, remember that the Overflow flag is only set when . . .
  - Two positive operands are added and their sum is negative
  - Two negative operands are added and their sum is positive

```
What will be the values of the Overflow flag?
   mov al,80h
   add al,92h               ; OF =

   mov al,-2
   add al,+127              ; OF =
```

33

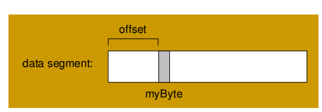
Data-Related Operators and Directives

- OFFSET Operator
- PTR Operator
- TYPE Operator
- LENGTHOF Operator
- SIZEOF Operator
- LABEL Directive

35

OFFSET Operator

- OFFSET returns the distance in bytes, of a label from the beginning of its enclosing segment
  - Protected mode: 32 bits
  - Real mode: 16 bits



The Protected-mode programs we write only have a single segment (we use the flat memory model).

36

OFFSET Examples

Let's assume that the data segment begins at 0040:4000h::

```
.data
bVal BYTE ?
wVal WORD ?
dVal DWORD ?
dVal2 DWORD ?

.code
mov esi,OFFSET bVal         ; ESI = 00404000
mov esi,OFFSET wVal         ; ESI = 00404001
mov esi,OFFSET dVal         ; ESI = 00404003
mov esi,OFFSET dVal2        ; ESI = 00404007
```


Relating to C/C++

The value returned by OFFSET is a pointer. Compare the following code written for both C++ and assembly language:

```
; C++ version:
char array[1000];
char * p = &array;
```

```
.data
myArray BYTE 1000 DUP(?)
.code
mov esi,OFFSET myArray          ; ESI is p
```


PTR Operator

Overrides the default type of a label (variable). Provides the flexibility to access part of a variable.

```
.data
myDouble DWORD 12345678h
.code
mov ax,myDouble                  ; error – why?

mov ax,WORD PTR myDouble         ; loads 5678h

mov WORD PTR myDouble,4321h      ; saves 4321h
```

To understand how this works, we need to know about little endian ordering of data in memory.

PTR Operator Examples

```
.data
myDouble DWORD 12345678h
```

doubleword	word	byte	offset	
12345678	5678	78	0000	myDouble
		56	0001	myDouble + 1
	1234	34	0002	myDouble + 2
		12	0003	myDouble + 3

```
mov al,BYTE PTR myDouble              ; AL = 78h
mov al,BYTE PTR [myDouble+1]          ; AL = 56h
mov al,BYTE PTR [myDouble+2]          ; AL = 34h
mov ax,WORD PTR [myDouble]            ; AX = 5678h
mov ax,WORD PTR [myDouble+2]          ; AX = 1234h
```


PTR Operator (cont)

PTR can also be used to combine elements of a smaller data type and move them into a larger operand. The CPU will automatically reverse the bytes.

```
.data
myBytes BYTE 12h,34h,56h,78h

.code
mov ax,WORD PTR [myBytes]          ; AX = 3412h
mov ax,WORD PTR [myBytes+2]        ; AX = 7856h
mov eax,DWORD PTR myBytes          ; EAX = 78563412h
```


TYPE Operator

The TYPE operator returns the size, in bytes, of a single element of a data declaration.

```
.data
var1 BYTE ?
var2 WORD ?
var3 DWORD ?
var4 QWORD ?

.code
mov eax,TYPE var1          ; 1
mov eax,TYPE var2          ; 2
mov eax,TYPE var3          ; 4
mov eax,TYPE var4          ; 8
```

LENGTHOF Operator

The LENGTHOF operator counts the number of elements in a single data declaration.

```
.data                                     LENGTHOF
byte1  BYTE 10,20,30                      ; 3
array1 WORD 30 DUP(?),0,0                 ; 32
array2 WORD 5 DUP(3 DUP(?))               ; 15
array3 DWORD 1,2,3,4                      ; 4
digitStr BYTE "12345678",0                ; 9

.code
mov ecx,LENGTHOF array1                   ; 32
```


SIZEOF Operator

The SIZEOF operator returns a value that is equivalent to multiplying LENGTHOF by TYPE.

```
.data                                     SIZEOF
byte1  BYTE 10,20,30                      ; 3
array1 WORD 30 DUP(?),0,0                 ; 64
array2 WORD 5 DUP(3 DUP(?))               ; 30
array3 DWORD 1,2,3,4                      ; 16
digitStr BYTE "12345678",0                ; 9

.code
mov ecx,SIZEOF array1                     ; 64
```


Spanning Multiple Lines (1 of 2)

A data declaration spans multiple lines if each line (except the last) ends with a comma. The LENGTHOF and SIZEOF operators include all lines belonging to the declaration:

```
.data
array WORD 10,20,
   30,40,
   50,60

.code
mov eax,LENGTHOF array           ; 6
mov ebx,SIZEOF array             ; 12
```


Spanning Multiple Lines (2 of 2)

In the following example, array identifies only the first WORD declaration. Compare the values returned by LENGTHOF and SIZEOF here to those in the previous slide:

```
.data
array  WORD 10,20
       WORD 30,40
       WORD 50,60

.code
mov eax,LENGTHOF array           ; 2
mov ebx,SIZEOF array             ; 4
```


LABEL Directive

- Assigns an alternate label name and type to an existing storage location
- LABEL does not allocate any storage of its own
- Removes the need for the PTR operator

```
.data
dwList   LABEL DWORD
wordList LABEL WORD
intList  BYTE 00h,10h,00h,20h
.code
mov eax,dwList             ; 20001000h
mov cx,wordList            ; 1000h
mov dl,intList             ; 00h
```


Indirect Addressing

- Indirect Operands
- Array Sum Example
- Indexed Operands
- Pointers

Indirect Operands

An indirect operand holds the address of a variable, usually an array or string. It can be dereferenced (just like a pointer).

```
.data
val1 BYTE 10h,20h,30h
.code
mov esi,OFFSET val1
mov al,[esi]              ; dereference ESI (AL = 10h)

inc esi
mov al,[esi]              ; AL = 20h

inc esi
mov al,[esi]              ; AL = 30h
```


Indirect Operands (2 of 2)

Use PTR when the size of a memory operand is ambiguous.

```
.data
myCount WORD 0

.code
mov esi,OFFSET myCount
inc [esi]                 ; error: ambiguous
inc WORD PTR [esi]        ; ok
```

Should PTR be used here?

```
add [esi],20
```


Array Sum Example

Indirect operands are ideal for traversing an array. Note that the register in brackets must be incremented by a value that matches the array type.

```
.data
arrayW WORD 1000h,2000h,3000h
.code
    mov esi,OFFSET arrayW
    mov ax,[esi]
    add esi,2             ; or: add esi,TYPE arrayW
    add ax,[esi]
    add esi,2             ; increment ESI by 2
    add ax,[esi]          ; AX = sum of the array
```

ToDo: Modify this example for an array of doublewords.

Indexed Operands

An indexed operand adds a constant to a register to generate an effective address. There are two notational forms:

[*label* + *reg*]   *label*[*reg*]

```
.data
arrayW WORD 1000h,2000h,3000h
.code
    mov esi,0
    mov ax,[arrayW + esi]     ; AX = 1000h
    mov ax,arrayW[esi]        ; alternate format
    add esi,2
    add ax,[arrayW + esi]
    etc.
```

ToDo: Modify this example for an array of doublewords.

Pointers

You can declare a pointer variable that contains the offset of another variable.

```
.data
arrayW WORD 1000h,2000h,3000h
ptrW DWORD arrayW
.code
    mov esi,ptrW
    mov ax,[esi]          ; AX = 1000h
```


JMP and LOOP Instructions

- JMP Instruction
- LOOP Instruction
- LOOP Example
- Summing an Integer Array
- Copying a String

JMP Instruction

- JMP is an unconditional jump to a label that is usually within the same procedure.
- Syntax: JMP *target*
- Logic: IP ← *target*
- Example:

```
top:
    .
    .
    jmp top
```

A jump outside the current procedure must be to a special type of label called a global label (see Section 5.5.2.3 for details).

LOOP Instruction

- The LOOP instruction creates a counting loop
- Syntax: LOOP *target*
- Logic:
  - CX ← CX – 1
  - if CX > 0, jump to *target*
- Implementation:
  - The assembler calculates the distance, in bytes, between the current location and the offset of the target label. It is called the relative offset.
  - The relative offset is added to IP.

LOOP Example

The following loop calculates the sum of the integers 5 + 4 + 3 +2 + 1:

offset	machine code	source code
00000000	66 B8 0000	mov ax,0
00000004	B9 00000005	mov ecx,5
00000009	66 03 C1	L1: add ax,cx
0000000C	E2 FB	loop L1
0000000E		

When LOOP is assembled, the current location = 0000000E. Looking at the LOOP machine code, we see that –5 (FBh) is added to the current location, causing a jump to location 00000009:

00000009 ← 0000000E + FB

Nested Loop

If you need to code a loop within a loop, you must save the outer loop counter's ECX value. In the following example, the outer loop executes 100 times, and the inner loop 20 times.

```
.data
count WORD ?
.code
    mov cx,100          ; set outer loop count
L1:
    mov count, cx       ; save outer loop count
    mov cx,20           ; set inner loop count
L2: .
    .
    loop L2             ; repeat the inner loop
    mov cx,count        ; restore outer loop count
    loop L1             ; repeat the outer loop
```


Summing an Integer Array

The following code calculates the sum of an array of 16-bit integers.

```
.data
intarray WORD 100h,200h,300h,400h
.code
    mov di, OFFSET intarray      ; address of intarray
    mov cx, LENGTHOF intarray    ; loop counter
    mov ax, 0                    ; zero the accumulator
L1:
    add ax, [di]                 ; add an integer
    add di, TYPE intarray        ; point to next integer
    loop L1                      ; repeat until ECX = 0
```


Your turn . . .

What changes would you make to the program on the previous slide if you were summing a “doubleword” array?

Copying a String

The following code copies a string from source to target.

```
.data
source  BYTE  "This is the source string",0
target  BYTE  SIZEOF source DUP(0),0

.code
    mov  si, 0                    ; index register
    mov  cx, SIZEOF source        ; loop counter
L1:
    mov  al, source[si]           ; get char from source
    mov  target[si], al           ; store it in the target
    inc  si                       ; move to next character
    loop L1                       ; repeat for entire string
```

good use of SIZEOF

64