

#### 请现场的同学们:

1. 打开雨课堂,点击页面右下角喇 叭按钮调至静音状态

#### 本次课程是

# 线上+线下

## 融合式教学

#### 请远程上课的同学们:

- 1. 打开雨课堂,点击页面右下角喇 叭按钮调至静音状态
- 2. 打开"腾讯会议" (会议室:824 8461 5333) , 进入会议室,并关闭麦克风

# CHAPTER 3: ASSEMBLY LANGUAGE FUNDAMENTALS

## **Chapter Overview**

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants
- 64-Bit Programming

## Program Template

(Template.asm)

```
; Program Template
; Description:
; Author:
; Creation Date:
; Revisions:
; Date:
; Modified by:
```

```
.386
.model flat,stdcall
.stack 4096
ExitProcess PROTO, dwExitCode:DWORD
.data
; declare variables here
. code
main PROC
   ; write your code here
   INVOKE ExitProcess, 0
main ENDP
 (insert additional procedures here)
END main
```

## Basic Elements of Assembly Language

- Integer constants
- Integer expressions
- Character and string constants
- Reserved words and identifiers
- Directives
- Instructions
  - Labels
  - Mnemonics
  - Operands
  - Comments
- Examples

## Integer Constants

- Optional leading + or sign
- binary, decimal, hexadecimal, or octal digits
- Common radix (基数后缀) characters:
  - h hexadecimal
  - d decimal
  - b binary
  - o octal

Examples: 30d, 6Ah, 42, 1101b

Hexadecimal beginning with letter: 0A5h

## Integer Expressions

Operators and precedence levels:

Operator	Name	Precedence Level
( )	parentheses	1
+,-	unary plus, minus	2
*,/	multiply, divide	3
MOD	modulus	3
+,-	add, subtract	4

Examples:

Expression	Value
16 / 5	3
-(3 + 4) * (6 - 1)	-35
-3 + 4 * 6 - 1	20
25 mod 3	1

## Character and String Constants

- Enclose character in single or double quotes
  - 'A', "x"
  - ASCII character = 1 byte
- Enclose strings in single or double quotes
  - "ABC"
  - 'xyz'
  - Each character occupies a single byte
- Embedded quotes:
  - 'Say "Goodnight," Gracie'

#### Reserved Words and Identifiers

- Reserved words cannot be used as identifiers
  - Instruction mnemonics, directives, type attributes, operators, predefined symbols
  - See MASM reference in Appendix A
- Identifiers
  - 1-247 characters, including digits
  - not case sensitive
    - -Cp
  - first character must be a letter, \_, @, ?, or \$

## Directives (伪指令)

- Commands that are recognized and acted upon by the assembler
  - Not part of the Intel instruction set
  - Used to declare code, data areas, select memory model, declare procedures, etc.
  - not case sensitive
- Different assemblers have different directives
  - NASM not the same as MASM, for example

#### Instructions

- Assembled into machine code by assembler
- Executed at runtime by the CPU
- We use the Intel IA-32 instruction set
- An instruction contains:

```
    Label (optional)
```

Mnemonic (required)

Operand (depends on the instruction)

Comment (optional)

[Label:] Mnemonic Operand(s) [; Comment]

#### Labels

- Act as place markers
  - marks the address (offset) of code and data
- Follow identifier rules
- Data label
  - must be unique
  - example: myArray (not followed by colon)
- Code label
  - target of jump and loop instructions
  - example: L1: (followed by colon)

## **Mnemonics and Operands**

- Instruction Mnemonics
  - memory aid
  - examples: MOV, ADD, SUB, MUL, INC, DEC
- Operands
  - constant
  - constant expression
  - memory (data label)
  - register

Constants and constant expressions are often called immediate values

#### Comments

- Comments are good!
  - explain the program's purpose
  - when it was written, and by whom
  - revision information
  - tricky coding techniques
  - application-specific explanations
- Single-line comments
  - begin with semicolon (;)
- Multi-line comments
  - begin with COMMENT directive and a programmerchosen character
  - end with the same programmer-chosen character

```
COMMENT!
;Here is the comment
mov ax, bx
add ax, 7
!
```

## Instruction Format Examples

- No operands
  - stc ; set Carry flag
- One operand
  - inc eax ; register
  - inc myByte ; memory
- Two operands
  - add ebx,ecx ; register, register
  - sub myByte,25 ; memory, constant
  - add eax,36 \* 25 ; register, constant-expression

#### What's Next

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants
- 64-Bit Programming

## Example: Adding and Subtracting Integers

```
TITLE Add and Subtract
                                      (AddSubAlt.asm)
; This program adds and subtracts 32-bit integers.
.386
.MODEL flat, stdcall
.STACK 4096
ExitProcess PROTO, dwExitCode:DWORD
DumpRegs PROTO
. code
main PROC
                               : EAX = 10000h
   mov eax, 10000h
   add eax,40000h
                             : EAX = 50000h
   sub eax,20000h
                               : EAX = 30000h
   call DumpRegs
   INVOKE ExitProcess, 0
main ENDP
END main
```

## **Example Output**

#### Program output, showing registers and flags:

```
EAX=00030000 EBX=7FFDF000 ECX=00000101 EDX=FFFFFFFF ESI=000000000 EDI=000000000 EBP=0012FFF0 ESP=0012FFC4 EIP=00401024 EFL=000000206 CF=0 SF=0 ZF=0 OF=0
```

## Suggested Coding Standards (1 of 2)

- Some approaches to capitalization
  - capitalize nothing
  - capitalize everything
  - capitalize all reserved words, including instruction mnemonics and register names
  - capitalize only directives and operators
- Other suggestions
  - descriptive identifier names
  - spaces surrounding arithmetic operators
  - blank lines between procedures

## Suggested Coding Standards (2 of 2)

- Indentation and spacing
  - code and data labels no indentation
  - executable instructions indent 4-5 spaces
  - comments: begin at column 40-45, aligned vertically
  - 1-3 spaces between instruction and its operands
    - ex: mov ax,bx
  - 1-2 blank lines between procedures

## Program Template (review)

```
; Program Template
; Description:
; Author:
; Creation Date:
; Revisions:
; Date:
; Modified by:
```

(Template.asm)

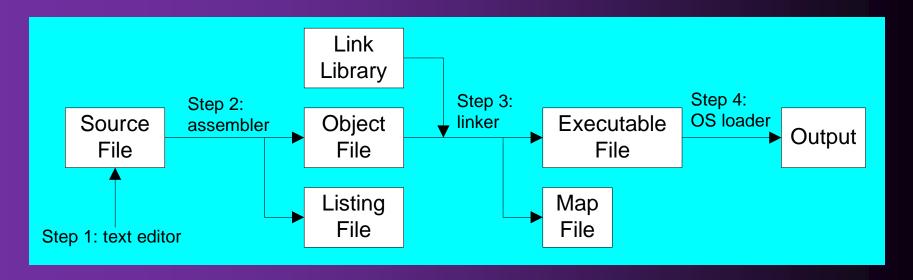
```
.386
.model flat,stdcall
.stack 4096
ExitProcess PROTO, dwExitCode:DWORD
.data
; declare variables here
. code
main PROC
   ; write your code here
   INVOKE ExitProcess, 0
main ENDP
  (insert additional procedures here)
END main
```

#### What's Next

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants

## Assemble-Link Execute Cycle

- The following diagram describes the steps from creating a source program through executing the compiled program.
- If the source code is modified, Steps 2 through 4 must be repeated.



The assembler contains a preprocessor to process directives, etc.

## Listing File

- Use it to see how your program is compiled
- Contains
  - source code
  - addresses
  - object code (machine language)
  - segment names
  - symbols (variables, procedures, and constants)
- Example: See § 3.3

#### What's Next

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants
- 64-Bit Programming

#### **Defining Data**

- Intrinsic Data Types (内部数据类型)
- Data Definition Statement
- Defining BYTE and SBYTE Data
- Defining WORD and SWORD Data
- Defining DWORD and SDWORD Data
- Defining QWORD Data
- Defining TBYTE Data
- Defining Real Number Data
- Little Endian Order
- Adding Variables to the AddSub Program
- Declaring Uninitialized Data

## Intrinsic Data Types (1 of 2)

- BYTE, SBYTE
  - 8-bit unsigned integer; 8-bit signed integer
- WORD, SWORD
  - 16-bit unsigned & signed integer
- DWORD, SDWORD
  - 32-bit unsigned & signed integer
- QWORD
  - 64-bit integer
- TBYTE
  - 80-bit integer

## Intrinsic Data Types (2 of 2)

- REAL4
  - 4-byte IEEE short real
- REAL8
  - 8-byte IEEE long real
- REAL10
  - 10-byte IEEE extended real

#### **Data Definition Statement**

- A data definition statement sets aside storage in memory for a variable.
- May optionally assign a name (label) to the data
- Syntax:

  [name] directive initializer [,initializer] . . .

  value1 BYTE 10

All initializers become binary data in memory

## Defining BYTE and SBYTE Data

Each of the following defines a single byte of storage:

- MASM does not prevent you from initializing a BYTE with a negative value, but it's considered poor style.
- If you declare a SBYTE variable, the Microsoft debugger will automatically display its value in decimal with a leading sign.

## **Defining Byte Arrays**

#### Examples that use multiple initializers:

```
list1 BYTE 10,20,30,40
list2 BYTE 10,20,30,40
        BYTE 50,60,70,80
        BYTE 81,82,83,84
list3 BYTE ?,32,41h,00100010b
list4 BYTE 0Ah,20h, 'A',22h
```

## Defining Strings (1 of 3)

- A string is implemented as an array of characters
  - For convenience, it is usually enclosed in quotation marks
  - It often will be null-terminated
- Examples:

## Defining Strings (2 of 3)

 To continue a single string across multiple lines, end each line with a comma:

```
menu BYTE "Checking Account",0dh,0ah,0dh,0ah,
   "1. Create a new account",0dh,0ah,
   "2. Open an existing account",0dh,0ah,
   "3. Credit the account",0dh,0ah,
   "4. Debit the account",0dh,0ah,
   "5. Exit",0ah,0ah,
   "Choice> ",0
```

## Defining Strings (3 of 3)

- End-of-line character sequence:
  - 0Dh = carriage return
  - OAh = line feed

```
str1 BYTE "Enter your name: ",0Dh,0Ah
    BYTE "Enter your address: ",0
newLine BYTE 0Dh,0Ah,0
```

Idea: Define all strings used by your program in the same area of the data segment.

## Using the DUP Operator

- Use DUP to allocate (create space for) an array or string. Syntax: counter DUP (argument)
- Counter and argument must be constants or constant expressions

#### Defining WORD and SWORD Data

- Define storage for 16-bit integers
  - or double characters
  - single value or multiple values

#### Defining DWORD and SDWORD Data

Storage definitions for signed and unsigned 32-bit integers:

### Defining QWORD, TBYTE, Real Data

Storage definitions for quadwords, tenbyte values, and real numbers:

```
quad1 QWORD 1234567812345678h

val1 TBYTE 100000000123456789Ah

rVal1 REAL4 -2.1

rVal2 REAL8 3.2E-260

rVal3 REAL10 4.6E+4096

ShortArray REAL4 20 DUP(0.0)
```

#### Little Endian Order

 All data types larger than a byte store their individual bytes in reverse order. The least significant byte occurs at the first (lowest) memory address.

Example:

val1 DWORD 12345678h

0000:	78
0001:	56
0002:	34
0003:	12

### Big Endian Order

 All data types larger than a byte store their individual bytes in "usual" order. The most significant byte occurs at the first (lowest) memory address.

Example:

val1 DWORD 12345678h

0000:	12
0001:	34
0002:	56
0003:	78

#### Adding Variables to AddSub

```
TITLE Add and Subtract, Version 2
                                             (AddSub2.asm)
; This program adds and subtracts 32-bit unsigned
; integers and stores the sum in a variable.
INCLUDE Irvine32.inc
.data
val1 DWORD 10000h
val2 DWORD 40000h
val3 DWORD 20000h
finalVal DWORD ?
. code
main PROC
                             ; start with 10000h
   mov eax, val1
   add eax, val2
                             ; add 40000h
                             ; subtract 20000h
   sub eax, val3
   mov finalVal, eax
                             ; store the result (30000h)
   call DumpRegs
                              ; display the registers
   exit
main ENDP
END main
```

#### **Declaring Unitialized Data**

Use the .data? directive to declare an uninitialized data segment:

.data?

Within the segment, declare variables with "?" initializers:

smallArray DWORD 10 DUP(?)

Advantage: the program's EXE file size is reduced.

#### What's Next

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants
- Real-Address Mode Programming
- 64-Bit Programming

## Symbolic Constants

- Equal-Sign Directive
- Calculating the Sizes of Arrays and Strings
- EQU Directive
- TEXTEQU Directive

## **Equal-Sign Directive**

- name = expression
  - expression is a 32-bit integer (expression or constant)
  - may be redefined
  - name is called a symbolic constant
- good programming style to use symbols

```
COUNT = 500
.
.
mov al,COUNT
```

## Calculating the Size of a Byte Array

- current location counter: \$
  - subtract address of list
  - difference is the number of bytes

```
list BYTE 10,20,30,40
ListSize = ($ - list)
```

### Calculating the Size of a Word Array

Divide total number of bytes by 2 (the size of a word)

```
list WORD 1000h,2000h,3000h,4000h
ListSize = ($ - list) / 2
```

## Calculating the Size of a Doubleword Array

Divide total number of bytes by 4 (the size of a doubleword)

```
list DWORD 1,2,3,4
ListSize = ($ - list) / 4
```

#### **EQU** Directive

- Define a symbol as either a number or text expression.
- Cannot be redefined

```
PI EQU <3.1416>
pressKey EQU <"Press any key to continue...",0>
.data
prompt BYTE pressKey
```

#### **TEXTEQU** Directive

- Define a symbol as either an integer or text expression.
- Called a text macro
- Can be redefined.

#### What's Next

- Basic Elements of Assembly Language
- Example: Adding and Subtracting Integers
- Assembling, Linking, and Running Programs
- Defining Data
- Symbolic Constants
- 64-Bit Programming

## 64-Bit Programming

- MASM supports 64-bit programming, although the following directives are not permitted:
  - INVOKE, ADDR, .model, .386, .stack
  - (Other non-permitted directives will be introduced in later chapters)

#### 64-Bit Version of AddTwoSum

```
1: ; AddTwoSum 64.asm - Chapter 3 example.
3: ExitProcess PROTO
5: .data
6: sum QWORD 0
8: .code
9: main PROC
10: mov rax,5
11: add rax,6
12: mov sum, rax
13:
14: mov ecx,0
15: call ExitProcess
16: main ENDP
17: END
```

### Things to Notice About the Previous Slide

- The following lines are not needed:
  - .386 .model flat,stdcall
  - .stack 4096
- INVOKE is not supported.
- CALL instruction cannot receive arguments
- Use 64-bit registers when possible

# CHAPTER 4: DATA TRANSFERS, ADDRESSING, AND ARITHMETIC

### **Chapter Overview**

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

#### 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 (解引用) by the assembler

#### **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
   mov ax,count
   mov eax,count
   mov ax,bl
```

#### 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.

```
0 10001111 Source
00000000 10001111 Destination
```

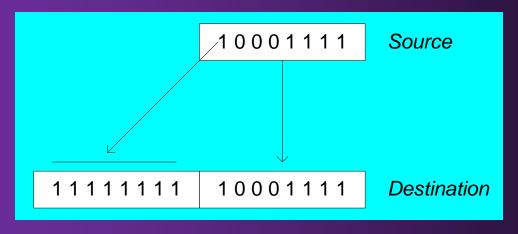
```
mov bl,10001111b

movzx ax,bl ; zero-extension
```

- The destination must be a register.
- The source cannot be immediate.

### 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.
- The source cannot be immediate.

#### **XCHG Instruction**

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

## Direct-Offset Operands (直接偏移操作数)

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.

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

#### What's Next

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

#### 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 1000000h
. code
   inc myWord
                            ; 1001h
   dec myWord
                            ; 1000h
                            ; 10000001h
   inc myDword
   mov ax,00FFh
   inc ax
                            ; AX = 0100h
   mov ax,00FFh
                            ; AX = 0000h
   inc al
```

#### 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
```

### **NEG** (negate) Instruction

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

### Implementing Arithmetic Expressions

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

```
Rval = -Xval + (Yval - Zval)
```

## 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 set when destination equals zero
  - Sign flag set when destination is negative
  - Carry flag set when unsigned value is out of range
  - Overflow flag set when signed value is out of range
- The MOV instruction never affects the flags.

## Zero Flag (ZF)

The Zero flag is set when the result of an operation produces zero in the destination operand.

#### Remember...

- 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

add cx,2

; CX = -1, SF = 1

; CX = 1, SF = 0
```

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

```
mov al,0

sub al,1

add al,2

; AL = 111111111b, SF = 1

; AL = 00000001b, SF = 0
```

## Signed and Unsigned Integers A Hardware Viewpoint

- All CPU instructions operate exactly the same on signed and unsigned integers
- The CPU cannot distinguish between signed and unsigned integers
- YOU, the programmer, are solely responsible for using the correct data type with each instruction

## 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).

## Overflow Flag (OF)

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

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.

#### 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 = 1

mov al,-2
add al,+127
; OF = 0
```

### Summary (Chap 3)

- Integer expression, character constant
- Directive interpreted by the assembler
- Instruction executes at runtime
- Code, data, and stack segments
- Source, listing, object, map, executable files
- Data definition directives:
  - BYTE, SBYTE, WORD, SWORD, DWORD, SDWORD, QWORD, TBYTE, REAL4, REAL8, and REAL10
  - DUP operator, location counter (\$)

## Summary (Chap 4)

- Data Transfer
  - MOV data transfer from source to destination
  - MOVSX, MOVZX, XCHG
- Operand types
  - direct, direct-offset, indirect, indexed
- Arithmetic
  - INC, DEC, ADD, SUB, NEG
  - Sign, Carry, Zero, Overflow flags

#### Homework

Reading Chap 3-4

Consider the topic of the team work

## Thanks!