

# x86-64

# Assembly Language

# Programming

## with

## Ubuntu



Ed Jorgensen, Ph.D.  
Version 1.1  
January 2016

Cover image

Top view of an Intel central processing unit Core i7 Skylake type core, model 6700K, released in June 2015.

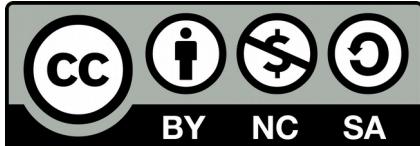
Source: Eric Gaba, [https://commons.wikimedia.org/wiki/File:Intel\\_CPU\\_Core\\_i7\\_6700K\\_Skylake\\_top.jpg](https://commons.wikimedia.org/wiki/File:Intel_CPU_Core_i7_6700K_Skylake_top.jpg)

Cover background

By Benjamint444 (Own work)

Source: [http://commons.wikimedia.org/wiki/File%3ASwirly\\_belt444.jpg](http://commons.wikimedia.org/wiki/File%3ASwirly_belt444.jpg)

Copyright © 2015, 2016, 2017, 2018, 2019 by Ed Jorger



You are free

To Share — to copy, distribute and transmit the work

To Remix — to adapt the work

Under the following conditions

Attribution — you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial — you may not use this work for commercial purposes.

Share Alike — if you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

# Table of Content

<b>1.0 Introduction.....</b>	
1.1 Prerequisites.....	1
1.2 What is Assembly Language.....	2
1.3 Why Learn Assembly Language.....	
1.3.1 Gain a Better Understanding of Architecture Issues.....	3
1.3.2 Understanding the Tool Chain.....	3
1.3.3 Improve Algorithm Development Skills.....	3
1.3.4 Improve Understanding of Functions/Procedures.....	3
1.3.5 Gain an Understanding of I/O Buffering.....	4
1.3.6 Understand Compiler Scope.....	4
1.3.7 Introduction Multi-processing Concepts.....	4
1.3.8 Introduction Interrupt Processing Concepts.....	
1.4 Additional References.....	
1.4.1 Ubuntu References.....	5
1.4.2 BASH Command Line References.....	5
1.4.3 Architecture References.....	5
1.4.4 Tool Chain References.....	
1.4.4.1 YASM References.....	6
1.4.4.2 DDD Debugger References.....	
<b>2.0 Architecture Overview.....</b>	
2.1 Architecture Overview.....	7
2.2 Data Storage Sizes.....	8
2.3 Central Processing Unit.....	
2.3.1 CPU Registers.....	
2.3.1.1 General Purpose Registers (GPRs).....	10
2.3.1.2 Stack Pointer Register (RSP).....	12
2.3.1.3 Base Pointer Register (RBP).....	12
2.3.1.4 Instruction Pointer Register (RIP).....	12
2.3.1.5 Flag Register (rFlags).....	12
2.3.1.6 XMM Registers.....	
2.3.2 Cache Memory.....	
2.4 Main Memory.....	15
2.5 Memory Layout.....	

## Table of Content

2.6	Memory Hierarchy.....	17
2.7	Exercises.....	
2.7.1	Quiz Questions.....	
<b>3.0</b>	<b>Data Representation.....</b>	
3.1	Integer Representation.....	
3.1.1	Two's Complement.....	23
3.1.2	Byte Example.....	
3.1.3	Word Example.....	
3.2	Unsigned and Signed Addition.....	
3.3	Floating-point Representation.....	
3.3.1	IEEE 32-bit Representation.....	
3.3.1.1	IEEE 32-bit Representation Examples.....	
3.3.1.1.1	Example → $-7.75_{10}$ .....	26
3.3.1.1.2	Example → $-0.125_{10}$ .....	
3.3.1.1.3	Example → $41440000_{16}$ .....	
3.3.2	IEEE 64-bit Representation.....	27
3.3.3	Not a Number (NaN).....	
3.4	Characters and Strings.....	
3.4.1	Character Representation.....	
3.4.1.1	American Standard Code for Information Interchange.....	28
3.4.1.2	Unicode.....	
3.4.2	String Representation.....	
3.5	Exercises.....	
3.5.1	Quiz Questions.....	
<b>4.0</b>	<b>Program Format.....</b>	
4.1	Comments.....	33
4.2	Numeric Values.....	33
4.3	Defining Constants.....	34
4.4	Data Section.....	34
4.5	BSS Section.....	35
4.6	Text Section.....	36
4.7	Example Program.....	
4.8	Exercises.....	
4.8.1	Quiz Questions.....	
<b>5.0</b>	<b>Tool Chain.....</b>	
5.1	Assemble/Link/Load Overview.....	41
5.2	Assembler.....	



5.2.1 Assemble Commands.....	43
5.2.2 List File.....	43
5.2.3 Two-Pass Assembler.....	43
5.2.3.1 First Pass.....	46
5.2.3.2 Second Pass.....	46
5.2.4 Assembler Directives.....	46
5.3 Linker.....	48
5.3.1 Linking Multiple Files.....	48
5.3.2 Linking Process.....	48
5.3.3 Dynamic Linking.....	48
5.4 Assemble/Link Script.....	50
5.5 Loader.....	51
5.6 Debugger.....	52
5.7 Exercises.....	52
5.7.1 Quiz Questions.....	52
<b>6.0 DDD Debugger.....</b>	
6.1 Starting DDD.....	
6.1.1 DDD Configuration Settings.....	
6.2 Program Execution with DDD.....	
6.2.1 Setting Breakpoints.....	57
6.2.2 Executing Programs.....	57
6.2.2.1 Run / Continue.....	60
6.2.2.2 Next / Step.....	60
6.2.3 Displaying Register Contents.....	60
6.2.4 DDD/GDB Commands Summary.....	60
6.2.4.1 DDD/GDB Commands, Examples.....	60
6.2.5 Displaying Stack Contents.....	65
6.2.6 Debugger Commands File (interactive).....	65
6.2.6.1 Debugger Commands File (non-interactive).....	66
6.2.6.2 Debugger Commands File (non-interactive).....	66
6.3 Exercises.....	67
6.3.1 Quiz Questions.....	67
6.3.2 Suggested Projects.....	67
<b>7.0 Instruction Set Overview.....</b>	
7.1 Notational Conventions.....	
7.1.1 Operand Notation.....	
7.2 Data Movement.....	



## Table of Contents

7.3 Addresses and Values.....	73
7.4 Conversion Instructions.....	74
7.4.1 Narrowing Conversions.....	74
7.4.2 Widening Conversions.....	74
7.4.2.1 Unsigned Conversions.....	74
7.4.2.2 Signed Conversions.....	74
7.5 Integer Arithmetic Instructions.....	83
7.5.1 Addition.....	83
7.5.1.1 Addition with Carry.....	83
7.5.2 Subtraction.....	83
7.5.3 Integer Multiplication.....	87
7.5.3.1 Unsigned Multiplication.....	87
7.5.3.2 Signed Multiplication.....	87
7.5.4 Integer Division.....	87
7.6 Logical Instructions.....	102
7.6.1 Logical Operations.....	102
7.6.2 Shift Operations.....	103
7.6.2.1 Logical Shift.....	103
7.6.2.2 Arithmetic Shift.....	103
7.6.3 Rotate Operations.....	103
7.7 Control Instructions.....	109
7.7.1 Labels.....	109
7.7.2 Unconditional Control Instructions.....	109
7.7.3 Conditional Control Instructions.....	109
7.7.3.1 Jump Out of Range.....	109
7.7.4 Iteration.....	109
7.8 Example Program, Sum of Squares.....	117
7.9 Exercises.....	118
7.9.1 Quiz Questions.....	118
7.9.2 Suggested Projects.....	118
<b>8.0 Addressing Modes.....</b>	
8.1 Addresses and Values.....	126
8.1.1 Register Mode Addressing.....	126
8.1.2 Immediate Mode Addressing.....	126
8.1.3 Memory Mode Addressing.....	126
8.2 Example Program, List Summation.....	129
8.3 Example Program, Pyramid Areas and Volumes.....	131
8.4 Exercises.....	131



8.4.1 Quiz Questions.....	136
8.4.2 Suggested Projects.....	
<b>9.0 Process Stack.....</b>	
9.1 Stack Example.....	141
9.2 Stack Instructions.....	142
9.3 Stack Implementation.....	
9.3.1 Stack Layout.....	143
9.3.2 Stack Operations.....	
9.4 Stack Example.....	147
9.5 Exercises.....	
9.5.1 Quiz Questions.....	148
9.5.2 Suggested Projects.....	
<b>10.0 Program Development.....</b>	
10.1 Understand the Problem.....	151
10.2 Create the Algorithm.....	152
10.3 Implement the Program.....	154
10.4 Test/Debug the Program.....	156
10.5 Error Terminology.....	
10.5.1 Assembler Error.....	157
10.5.2 Run-time Error.....	157
10.5.3 Logic Error.....	
10.6 Exercises.....	
10.6.1 Quiz Questions.....	158
10.6.2 Suggested Projects.....	
<b>11.0 Macros.....</b>	
11.1 Single-Line Macros.....	161
11.2 Multi-Line Macros.....	
11.2.1 Macro Definition.....	162
11.2.2 Using a Macro.....	
11.3 Macro Example.....	163
11.4 Debugging Macros.....	165
11.5 Exercises.....	
11.5.1 Quiz Questions.....	165
11.5.2 Suggested Projects.....	
<b>12.0 Functions.....</b>	
12.1 Updated Linking Instructions.....	



## Table of Content

12.2 Debugger Commands.....	169
12.2.1 Debugger Command, <i>next</i> .....	169
12.2.2 Debugger Command, <i>step</i> .....	169
12.3 Stack Dynamic Local Variables.....	168
12.4 Function Declaration.....	169
12.5 Standard Calling Convention.....	169
12.6 Linkage.....	170
12.7 Argument Transmission.....	171
12.8 Calling Convention.....	172
12.8.1 Parameter Passing.....	172
12.8.2 Register Usage.....	173
12.8.3 Call Frame.....	173
12.8.3.1 Red Zone.....	173
12.9 Example, Statistical Function 1 (leaf).....	177
12.9.1 Caller.....	177
12.9.2 Callee.....	177
12.10 Example, Statistical Function2 (non-leaf).....	179
12.10.1 Caller.....	179
12.10.2 Callee.....	179
12.11 Stack-Based Local Variables.....	183
12.12 Summary.....	186
12.13 Exercises.....	186
12.13.1 Quiz Questions.....	187
12.13.2 Suggested Projects.....	187
<b>13.0 System Services.....</b>	
13.1 Calling System Services.....	191
13.2 Newline Character.....	192
13.3 Console Output.....	192
13.3.1 Example, Console Output.....	192
13.4 Console Input.....	193
13.4.1 Example, Console Input.....	193
13.5 File Open Operations.....	202
13.5.1 File Open.....	202
13.5.2 File Open/Create.....	202
13.6 File Read.....	204
13.7 File Write.....	205
13.8 File Operations Examples.....	205
13.8.1 Example, File Write.....	205



13.8.2 Example, File Read.....	
13.9 Exercises.....	
13.9.1 Quiz Questions.....	216
13.9.2 Suggested Projects.....	
<b>14.0 Multiple Source Files.....</b>	
14.1 Extern Statement.....	219
14.2 Example, Sum and Average.....	
14.2.1 Assembly Main.....	220
14.2.2 Function Source.....	222
14.2.3 Assemble and Link.....	
14.3 Interfacing with a High-Level Language.....	
14.3.1 Example, C++ Main / Assembly Function.....	224
14.3.2 Compile, Assemble, and Link.....	
14.4 Exercises.....	
14.4.1 Quiz Questions.....	226
14.4.2 Suggested Projects.....	
<b>15.0 Stack Buffer Overflow.....</b>	
15.1 Understanding a Stack Buffer Overflow.....	230
15.2 Code to Inject.....	231
15.3 Code Injection.....	234
15.4 Code Injection Protections.....	
15.4.1 Data Stack Smashing Protector (or Canaries).....	235
15.4.2 Data Execution Prevention.....	236
15.4.3 Data Address Space Layout Randomization.....	
15.5 Exercises.....	
15.5.1 Quiz Questions.....	236
15.5.2 Suggested Projects.....	
<b>16.0 Command Line Arguments.....</b>	
16.1 Parsing Command Line Arguments.....	239
16.2 High-Level Language Example.....	240
16.3 Argument Count and Argument Vector Table.....	241
16.4 Assembly Language Example.....	242
16.5 Exercises.....	
16.5.1 Quiz Questions.....	246
16.5.2 Suggested Projects.....	
<b>17.0 Input/Output Buffering.....</b>	



## Table of Content

17.1 Why Buffer?.....	249
17.2 Buffering Algorithm.....	251
17.3 Exercises.....	
17.3.1 Quiz Questions.....	254
17.3.2 Suggested Projects.....	
<b>18.0 Floating-Point Instructions.....</b>	
18.1 Floating-Point Values.....	257
18.2 Floating-Point Registers.....	258
18.3 Data Movement.....	258
18.4 Integer / Floating-Point Conversion Instructions.....	260
18.5 Floating-Point Arithmetic Instructions.....	
18.5.1 Floating-Point Addition.....	262
18.5.2 Floating-Point Subtraction.....	263
18.5.3 Floating-Point Multiplication.....	265
18.5.4 Floating-Point Division.....	267
18.5.5 Floating-Point Square Root.....	
18.6 Floating-Point Control Instructions.....	
18.6.1 Floating-Point Comparison.....	
18.7 Floating-Point Calling Conventions.....	274
18.8 Example Program, Sum and Average.....	275
18.9 Example Program, Absolute Value.....	276
18.10 Exercises.....	
18.10.1 Quiz Questions.....	278
18.10.2 Suggested Projects.....	
<b>19.0 Parallel Processing.....</b>	
19.1 Distributed Computing.....	280
19.2 Multiprocessing.....	
19.2.1 POSIX Threads.....	281
19.2.2 Race Conditions.....	
19.3 Exercises.....	
19.3.1 Quiz Questions.....	285
19.3.2 Suggested Projects.....	
<b>20.0 Interrupts.....</b>	
20.1 Multi-user Operating System.....	
20.1.1 Interrupt Classification.....	
20.1.2 Interrupt Timing.....	
20.1.2.1 Asynchronous Interrupts.....	288



20.1.2.2 Synchronous Interrupts.....	
20.1.3 Interrupt Categories.....	
20.1.3.1 Hardware Interrupt.....	
20.1.3.1.1 Exceptions.....	
20.1.3.2 Software Interrupts.....	
20.2 Interrupt Types and Levels.....	
20.2.1 Interrupt Types.....	290
20.2.2 Privilege Levels.....	
20.3 Interrupt Processing.....	
20.3.1 Interrupt Service Routine (ISR).....	292
20.3.2 Processing Steps.....	
20.3.2.1 Suspension.....	292
20.3.2.2 Obtaining ISR Address.....	292
20.3.2.3 Jump to ISR.....	293
20.3.2.4 Suspension Execute ISR.....	293
20.3.2.5 Resumption.....	
20.4 Suspension Interrupt Processing Summary.....	294
20.5 Exercises.....	
20.5.1 Quiz Questions.....	295
20.5.2 Suggested Projects.....	
<b>21.0 Appendix A – ASCII Table.....</b>	
<b>22.0 Appendix B – Instruction Set Summary.....</b>	
22.1 Notation.....	299
22.2 Data Movement Instructions.....	300
22.3 Data Conversion instructions.....	300
22.4 Integer Arithmetic Instructions.....	301
22.5 Logical, Shift, and Rotate Instructions.....	303
22.6 Control Instructions.....	305
22.7 Stack Instructions.....	307
22.8 Function Instructions.....	307
22.9 Floating-Point Data Movement Instructions.....	307
22.10 Floating-Point Data Conversion Instructions.....	308
22.11 Floating-Point Arithmetic Instructions.....	309
22.12 Floating-Point Control Instructions.....	
<b>23.0 Appendix C – System Services.....</b>	
23.1 Return Codes.....	315
23.2 Basic System Services.....	



## Table of Content

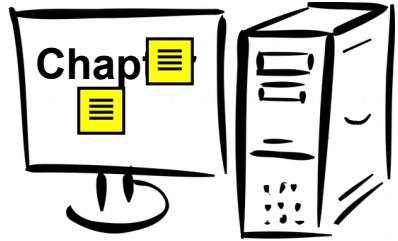
23.3	File Modes.....	317
23.4	Error Codes.....	
<b>24.0</b>	<b>Appendix D – Quiz Question Answers.....</b>	
24.1	Quiz Question Answers, Chapter 1.....	321
24.2	Quiz Question Answers, Chapter 2.....	321
24.3	Quiz Question Answers, Chapter 3.....	322
24.4	Quiz Question Answers, Chapter 4.....	324
24.5	Quiz Question Answers, Chapter 5.....	325
24.6	Quiz Question Answers, Chapter 6.....	326
24.7	Quiz Question Answers, Chapter 7.....	327
24.8	Quiz Question Answers, Chapter 8.....	330
24.9	Quiz Question Answers, Chapter 9.....	331
24.10	Quiz Question Answers, Chapter 10.....	331
24.11	Quiz Question Answers, Chapter 11.....	332
24.12	Quiz Question Answers, Chapter 12.....	332
24.13	Quiz Question Answers, Chapter 13.....	333
24.14	Quiz Question Answers, Chapter 14.....	333
24.15	Quiz Question Answers, Chapter 15.....	334
24.16	Quiz Question Answers, Chapter 16.....	334
24.17	Quiz Question Answers, Chapter 17.....	335
24.18	Quiz Question Answers, Chapter 18.....	335
24.19	Quiz Question Answers, Chapter 19.....	336
24.20	Quiz Question Answers, Chapter 20.....	
<b>25.0</b>	<b>Alphabetical Index.....</b>	

## Illustration Index

Illustration 1: Computer Architecture.....	7
Illustration 2: CPU Block Diagram.....	15
Illustration 3: Little-Endian Data Layout.....	16
Illustration 4: General Memory Layout.....	17
Illustration 5: Memory Hierarchy.....	18
Illustration 6: Overview: Assemble, Link, Load.....	42
Illustration 7: Little-Endian, Multiple Variable Data Layout.....	44
Illustration 8: Linking Multiple Files.....	49
Illustration 9: Initial Debugger Screen.....	56
Illustration 10: Debugger Screen with Breakpoint Set.....	58
Illustration 11: Debugger Screen with Green Arrow.....	59
Illustration 12: DDD Command Bar.....	60
Illustration 13: Register Window.....	61
Illustration 14: MOV Instruction Overview.....	71
Illustration 15: Integer Multiplication Overview.....	88
Illustration 16: Integer Division Overview.....	96
Illustration 17: Logical Operations.....	102
Illustration 18: Logical Shift Overview.....	104
Illustration 19: Logical Shift Operations.....	104
Illustration 20: Arithmetic Left Shift.....	106
Illustration 21: Arithmetic Right Shift.....	106
Illustration 22: Process Memory Layout.....	144
Illustration 23: Process Memory Layout Example.....	145
Illustration 24: Stack Frame Layout.....	175
Illustration 25: Stack Frame Layout with Red Zone.....	176
Illustration 26: Stack Call Frame Example.....	230
Illustration 27: Stack Call Frame Corruption.....	235
Illustration 28: Argument Vector Layout.....	242
Illustration 29: Privilege Levels.....	291
Illustration 30: Interrupt Processing Overview.....	

**Table of Cont**

*If you give someone a program, you will frustrate them for a day; if you teach them to program, you will frustrate them for a lifetime.*



# 1.0 Introduction

The purpose of this text is to provide a reference for University level assembly language and systems programming courses. Specifically, this text addresses the x86 instruction set for the popular x86-64 class of processors using the Ubuntu 64-bit Operating System (OS). While the provided code and various examples should work under any Linux-based 64-bit OS, they have only been tested under Ubuntu 14.04 LTS (64-bit).

The x86-64 is a Complex Instruction Set Computing (CISC<sup>2</sup>) CPU design. This refers to the internal processor design philosophy. CISC processors typically include a wide variety of instructions (sometimes overlapping), varying instruction sizes, and a wide range of addressing modes. The term was retroactively coined in contrast to Reduced Instruction Set Computer (RISC<sup>3</sup>).

## 1.1 Prerequisites

It must be noted that the text is not geared toward learning how to program. It is assumed that the reader has already become proficient in a high-level programming language. Specifically, the text is generally geared toward a compiled, C-based high-level language such as C, C++, or Java. Many of the explanations and examples assume the reader is already familiar with programming concepts such as declarations, arithmetic operations, control structures, iteration, function calls, functions, indirection (i.e., pointers), and variable scoping issues.

Additionally, the reader should be comfortable using a Linux-based operating system including using the command line. If the reader is new to Linux, the Additional References section has links to some useful documentation.

1 For more information, refer to: <http://en.wikipedia.org/wiki/X86>

2 For more information, refer to: [http://en.wikipedia.org/wiki/Complex\\_instruction\\_set\\_computer](http://en.wikipedia.org/wiki/Complex_instruction_set_computer)

3 For more information, refer to: [http://en.wikipedia.org/wiki/Reduced\\_instruction\\_set\\_computer](http://en.wikipedia.org/wiki/Reduced_instruction_set_computer)

## 1.2 What is Assembly Language

The typical question asked by students is 'why learn assembly?'. Before addressing that question, let's clarify what exactly assembly language is.

Assembly language is machine specific. For example, code written for an x86-64 processor will not run on a different processor such as a RISC processor (popular in tablets and smart-phones).

Assembly language is a "low-level" language and provides the basic instructional interface to the computer processor. Assembly language is as close to the processor as you can get as a programmer. Programs written in a high-level language are translated into assembly language in order for the processor to execute the program. The high-level language is an abstraction between the language and the actual processor instructions. As such, the idea that "assembly is dead" is nonsense.

Assembly language gives you direct control of the system's resources. This involves setting processor registers, accessing memory locations, and interfacing with other hardware elements. This requires a significantly deeper understanding of exactly how the processor and memory work.

## 1.3 Why Learn Assembly Language

The goal of this text is to provide a comprehensive introduction to programming in assembly language. The reasons for learning assembly language are more about understanding how a computer works instead of developing large programs. Since assembly language is machine specific, the lack of portability is very limiting for programming projects.

The process of actually learning assembly language involves writing non-trivial programs to perform specific low-level actions including arithmetic operations, function calls, using stack-dynamic local variables, and operating system interaction for activities such as input/output. Just looking at small assembly language programs will not be enough.

In the long run, learning the underlying principles, including assembly language, is what makes the difference between a coding technician unable to cope with changing languages and a computer scientist who is able to adapt to the ever-changing technologies.

The following sections provide some detail on the various, more specific reasons for learning assembly language.

### 1.3.1 Gain a Better Understanding of Architecture Issues

Learning and spending some time working at the assembly language level provides a richer understanding of the underlying computer architecture. This includes the basic instruction set, processor registers, memory addressing, hardware interfacing, and Input/Output. Since ultimately all programs execute at this level, knowing the capabilities of assembly language provides useful insights into what is possible, what is easy, and what might be more difficult or slower.

### 1.3.2 Understanding the Tool Chain

The tool chain is the name for the process of taking code written by a human and converting it into something that the computer can directly execute. This includes the compiler, or assembler in our case, the linker, the loader, and the debugger. In reference to compiling, beginning programmers are told “just do this” with little explanation of the complexity involved in the process. Working at the low-level can help provide the basis for understanding and appreciating the details of the tool chain.

### 1.3.3 Improve Algorithm Development Skills

Working with assembly language and writing low-level programs helps programmers improve algorithm development skills by practicing with a language that requires more thought and more attention to detail. In the highly unlikely event that a program does not work the first time, debugging assembly language also provides practice debugging and requires a more nuanced approach since just adding a bunch of output statements is more difficult at the assembly language level. This typically involves a more comprehensive use of a debugger which is a useful skill for any programmer.

### 1.3.4 Improve Understanding of Functions/Procedures

Working with assembly language provides a greatly improved understanding of how function/procedure calls work. This includes the contents and structure of the function call frame, also referred to as the activation record. Depending on the specific instance, the activation record might include stack-based arguments, preserved registers, and/or stack dynamic local variables. There are some significant implementation and security implications regarding stack dynamic local variables that are best understood working at a low-level. Due to the security implications, it would be appropriate to remind readers to always use their powers for good. Additionally, use of the stack and the associated call frame is the basis for recursion and understanding the fairly straightforward implementation of recursive functions.

### 1.3.5 Gain an Understanding of I/O Buffering

In a high-level language, input/output instructions and the associated buffering operations can appear magical. Working at the assembly language level and performing some low-level input/output operations provides a more detailed understanding of how input/output and buffering really works. This includes the differences between interactive input/output, file input/output, and the associated operating system services.

### 1.3.6 Understand Compiler Structure

Programming with assembly language, after having already learned a high-level language, helps ensure programmers understand the scope and capabilities of a compiler. Specifically, this means learning what the compiler does and does not do in relation to the computer architecture.

### 1.3.7 Introduction Multi-processing Concepts

This text will also provide a brief introduction to multi-processing concepts. The general concepts of distributed and multi-core programming are presented with the focus being placed on shared memory, threaded processing. It is the author's belief that truly understanding the subtle issues associated with threading such as shared memory and race conditions is most easily understood at the low-level.

### 1.3.8 Introduction Interrupt Processing Concepts

The underlying fundamental mechanism in which modern multi-user computers work is based on interrupts. Working at a low-level is the best place to provide an introduction to the basic concepts associated with interrupt handling, interrupt service handles, and vector interrupts.

## 1.4 Additional References

Some key references for additional information are noted in the following sections. These references provide much more extensive and detailed information.

If any of these locations change, a web search will be able to find the new location.

#### 1.4.1 Ubuntu Reference

There is significant documentation available for the Ubuntu OS. The principal user guide is as follows:

- [Ubuntu Community Wiki](#)
- [Getting Started with Ubuntu 10.04 LTS](#)

In addition, there are many other sites dedicated to providing help using Ubuntu (or other Linux-based OS's).

#### 1.4.2 BASH Command Line Reference

BASH is the default shell for Ubuntu. The reader should be familiar with basic command line operations. Some additional references are as follows:

- [Linux Command Line](#) (on-line Tutorial and t
- [An Introduction to the Linux Command Shell For Beginners](#) (

In addition, there are many other sites dedicated to providing information regarding the BASH command shell.

#### 1.4.3 Architecture Reference

Some key references published by Intel provide a detailed technical description of the architecture and programming environment of Intel processors supporting IA-32 and Intel 64 Architectures.

- [Intel® 64 and IA-32 Architectures Software Developer's Manual: Building Block Architectures](#)
- [Intel 64 and IA-32 Architectures Software Developer's Manual: Instruction Set Reference](#)
- [Intel 64 and IA-32 Architectures Software Developer's Manual: System Programming Guide](#)

If the embedded links do not work, an Internet search can help find the new location.

#### 1.4.4 Tool Chain Reference

The tool chain includes the assembler, linker, loader, and debugger. Chapter 5, Tool Chain, provides an overview of the tool chain being used in this text. The following references provide more detailed information and documentation.



#### 1.4.4.1 YASM Reference



The YASM assembler is an open source assembler commonly available on Linux-based systems. The YASM references are as follows:

- [Yasm Web Site](#)
- [Yasm Documentation](#)

Additional information regarding YASM may be available at a number of assembly language sites and can be found through an Internet search.



#### 1.4.4.2 DDD Debugger Reference



The DDD debugger is an open source debugger capable of supporting assembly language.

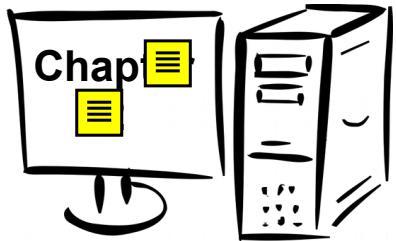


- [DDD Web Site](#)
- [DDD Documentation](#)

Additional information regarding DDD may be at a number of assembly language sites and can be found through an Internet search.



*Warning, keyboard not found. Press enter to continue.*



## 2.0 Architecture Overview

This chapter presents a basic, general overview of the x86-64 architecture. For a more detailed explanation, refer to the additional references noted in Chapter 1, Introduction.

### 2.1 Architecture Overview

The basic components of a computer include a Central Processing Unit (CPU), Primary Storage or Random Access Memory (RAM), Secondary Storage, Input/Output devices (e.g., screen, keyboard, mouse), and an interconnection referred to as the Bus.

A very basic diagram of the computer architecture is as follows:

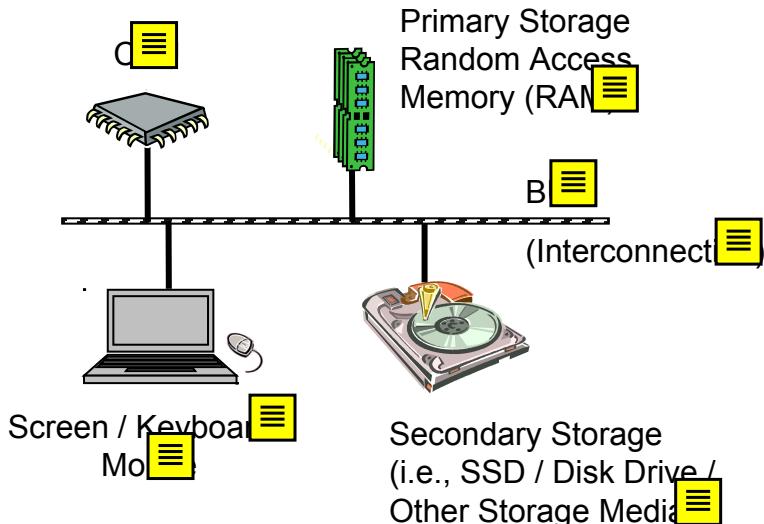


Illustration 1: Computer Architecture



The architecture is typically referred to as the Von Neumann Architecture<sup>4</sup>, or the Princeton architecture, and was described in 1945 by the mathematician and physicist John von Neumann.



Programs and data are typically stored on secondary storage (e.g., disk drive or solid state drive). When a program is executed, it must be copied from secondary storage into the primary storage or main memory (RAM). The CPU executes the program from primary storage or RAM.



Primary storage or main memory is also referred to as volatile memory since when power is removed, the information is not retained and thus lost. Secondary storage is referred to as non-volatile memory since the information is retained when powered off.



For example, consider storing a term paper on secondary storage (i.e., disk). When the user starts to write or edit the term paper, it is copied from the secondary storage medium into primary storage (i.e., RAM or main memory). When done, the updated version is typically stored back to the secondary storage (i.e., disk). If you have ever lost power while editing a document (assuming no battery or uninterruptible power supply), losing the unsaved work will certainly clarify the difference between volatile and non-volatile memory.



## 2.2 Data Storage Sizes



The x86-64 architecture supports a specific set of data storage size elements, all based on powers of two. The supported storage sizes are as follows:



Storage	Size (bits)	Size (bytes)
Byte	8-bits	1 byte
Word	16-bits	2 bytes
Double-word	32-bits	4 bytes
Quadword	64-bits	8 bytes
Double quadword	128-bits	16 bytes

Lists or arrays (sets of memory) can be reserved in any of these types.



These storage sizes have a direct correlation to variable declarations in high-level languages (e.g., C, C++, Java, etc.).



---

<sup>4</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Von\\_Neumann\\_architecture](http://en.wikipedia.org/wiki/Von_Neumann_architecture)

For example, C/C++ declarations are mapped as follows:

C/C++ Declaration	Storage	Size (bits)	Size (bytes)
char	Byte	8-bits	1 byte
short	Word	16-bits	2 bytes
int	Double-word	32-bits	4 bytes
unsigned int	Double-word	32-bits	4 bytes
long <sup>5</sup>	Quadword	64-bits	8 bytes
long long	Quadword	64-bits	8 bytes
char *	Quadword	64-bits	8 bytes
int *	Quadword	64-bits	8 bytes
float	Double-word	32-bits	4 bytes
double	Quadword	64-bits	8 bytes

The asterisk indicates an address variable. For example, `int *` means the address of an integer. Other high-level languages typically have similar mappings.

## 2.3 Central Processing Unit

The Central Processing Unit<sup>6</sup> (CPU) is typically referred to as the “brains” of the computer since that is where the actual calculations are performed. The CPU is housed in a single chip, sometimes called a processor, chip, or die<sup>7</sup>. The cover image shows one such CPU.

The CPU chip includes a number of functional units, including the Arithmetic Logic Unit<sup>8</sup> (ALU) which is the part of the chip that actually performs the arithmetic and logical calculations. In order to support the ALU, processor registers<sup>9</sup> and cache memory are also included “on the die” (term for inside the chip). The CPU registers and cache memory are described in subsequent sections.

It should be noted that the internal design of a modern processor is quite complex. This section provides a very simplified, high-level view of some key functional units within a CPU. Refer to the footnotes or additional references for more information.

5 Note, the 'long' type declaration is compiler dependent. Type shown is for gcc and g++ compilers.

6 For more information, refer to: [http://en.wikipedia.org/wiki/Central\\_processing\\_unit](http://en.wikipedia.org/wiki/Central_processing_unit)

7 For more information, refer to: [http://en.wikipedia.org/wiki/Die\\_\(integrated\\_circuit\)](http://en.wikipedia.org/wiki/Die_(integrated_circuit))

8 For more information, refer to: [http://en.wikipedia.org/wiki/Arithmetic\\_logic\\_unit](http://en.wikipedia.org/wiki/Arithmetic_logic_unit)

9 For more information, refer to: [http://en.wikipedia.org/wiki/Processor\\_register](http://en.wikipedia.org/wiki/Processor_register)

10 For more information, refer to: [http://en.wikipedia.org/wiki/Cache\\_\(computer\\_memory\)](http://en.wikipedia.org/wiki/Cache_(computer_memory))

## 2.3.1 CPU Registers

A CPU register, or just register, is a temporary storage or working location built into the CPU itself (separate from memory). Computations are typically performed by the CPU using registers.

### 2.3.1.1 General Purpose Registers (GPRs)

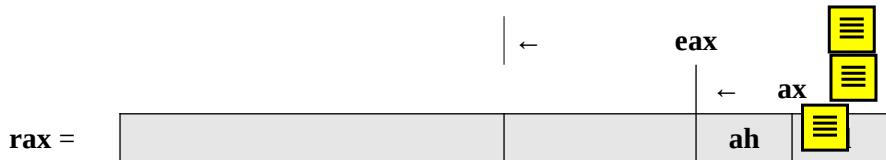
There are sixteen, 64-bit General Purpose Registers (GPRs). The GPRs are described in the following table. A GPR register can be accessed with all 64-bits or some portion or subset accessed.

64-bit register	Lowest 32-bit	Lowest 16-bit	Lowest 8-bit
<b>rax</b>	<b>eax</b>	<b>ax</b>	
<b>rbx</b>	<b>ebx</b>	<b>bx</b>	
<b>rcx</b>	<b>ecx</b>	<b>cx</b>	
<b>rdx</b>	<b>edx</b>	<b>dx</b>	
<b>rsi</b>	<b>esi</b>	<b>si</b>	
<b>rdi</b>	<b>edi</b>	<b>di</b>	
<b>rbp</b>	<b>ebp</b>	<b>bp</b>	
<b>rsp</b>	<b>esp</b>	<b>sp</b>	
<b>r8</b>	<b>r8d</b>	<b>r8w</b>	
<b>r9</b>	<b>r9d</b>	<b>r9w</b>	
<b>r10</b>	<b>r10d</b>	<b>r10w</b>	
<b>r11</b>	<b>r11d</b>	<b>r11w</b>	
<b>r12</b>	<b>r12d</b>	<b>r12w</b>	
<b>r13</b>	<b>r13d</b>	<b>r13w</b>	
<b>r14</b>	<b>r14d</b>	<b>r14w</b>	
<b>r15</b>	<b>r15d</b>	<b>r15w</b>	

Additionally, some of the GPR registers are used for dedicated purposes as described in the later sections.

When using data element sizes less than 64-bits (i.e., 32-bit, 16-bit, or 8-bit), the lower portion of the register can be accessed by using a different register name as shown in the table.

For example, when accessing the lower portions of the 64-bit **rax** register, the layout is as follows:



As shown in the diagram, the first four registers, **rax**, **rbx**, **rcx**, and **rdx** also allow the bits 8-15 to be accessed with the **ah**, **bh**, **ch**, and **dh** register names. With the exception of **ah**, these are provided for legacy support and will not be used in this text.

The ability to access portions of the register means that, if the quadword **rax** register is set to 50,000,000,000<sub>10</sub> (fifty billion), the **rax** register would contain the following value in hex.

**rax = 0000 000B A43B 7**

If a subsequent operation sets the word **ax** register to 50,000<sub>10</sub> (fifty thousand, which is C350<sub>16</sub>), the **rax** register would contain the following value in hex.

**rax = 0000 000B A43B C**

In this case, when the lower 16-bit **ax** portion of the 64-bit **rax** register is set, the upper 48-bits are unaffected. Note the change in AX (from 7400<sub>16</sub> to C350<sub>16</sub>).

If a subsequent operation sets the byte sized **al** register to 50<sub>10</sub> (fifty, which is 32<sub>16</sub>), the **rax** register would contain the following value in hex.

**rax = 0000 000B A43B C**

When the lower 8-bit **al** portion of the 64-bit **rax** register is set, the upper 56-bits are unaffected. Note the change in AL (from 50<sub>16</sub> to 32<sub>16</sub>).

For 32-bit register operations, the upper 32-bits is cleared (set to zero). Generally, this is not an issue since operations on 32-bit registers do not use the upper 32-bits of the register. For unsigned values, this can be useful to convert from 32-bits to 64-bits. However, this will not work for signed conversions from 32-bit to 64-bit values. Specifically, it will potentially provide incorrect results for negative values. Refer to Chapter 3, Data Representation for additional information regarding the representation of signed values.



### 2.3.1.2 Stack Pointer Register (RSP)



One of the CPU registers, **rsp**, is used to point to the current top of the stack. The **rsp** register should not be used for data or other uses. Additional information regarding the stack and stack operations is provided in Chapter 9, Process Stack.



### 2.3.1.3 Base Pointer Register (RBP)



One of the CPU registers, **rbp**, is used as a base pointer during function calls. The **rbp** register should not be used for data or other uses. Additional information regarding the functions and function calls is provided in Chapter 12, Functions.



### 2.3.1.4 Instruction Pointer Register (RIP)



In addition to the GPRs, there is a special register, **rip**, which is used by the CPU to point to the **next instruction to be executed**. Specifically, since the **rip** points to the next instruction, that means the instruction being pointed to by **rip**, and shown in the debugger, has not yet been executed. This is an important distinction which can be confusing when reviewing code in a debugger.



### 2.3.1.5 Flag Register (RFlags)



The flag register, **rFlags**, is used for status and CPU control information. The **rFlag** register is updated by the CPU after each instruction and not directly accessible by programs. This register stores status information about the instruction that was just executed. Of the 64-bits in the **rFlag** register, many are reserved for future use.



The following table shows some of the status bits in the flag register.



Name	Symbol	Bit	
Carry	CF	0	Used to indicate if the previous operation resulted in a carry.
Parity	PF	2	Used to indicate if the last byte has an even number of 1's (i.e., even parity).
Adjust	AF	4	Used to support Binary Coded Decimal operations.
Zero	ZF	6	Used to indicate if the previous operation resulted in a zero result.



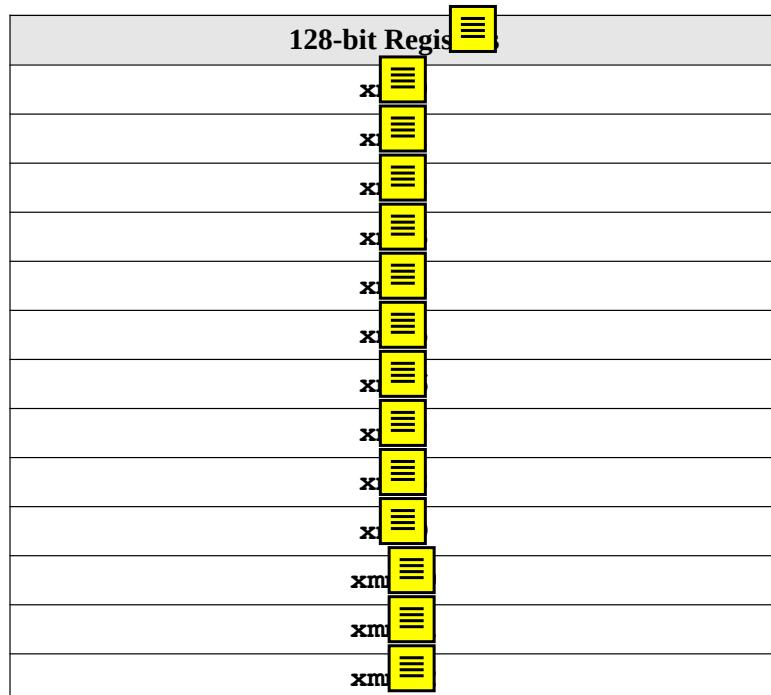
Sign	SF	7	Used to indicate if the result of the previous operation resulted in a 1 in the most significant bit (indicating negative in the context of signed data). 
Direction	DF	10	Used to specify the direction (increment or decrement) for some string operations. 
Overflow	OF	11	Used to indicate if the previous operation resulted in an overflow. 

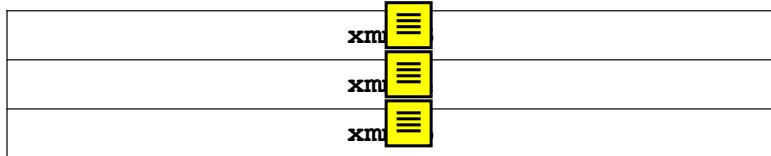
There are a number of additional bits not specified in this text. More information can be obtained from the additional references noted in Chapter 1, Introduction. 

### 2.3.1.6 XMM Registers

There are a set of dedicated registers used to support 64-bit and 32-bit floating-point operations and Single Instruction Multiple Data (SIMD) instructions. The SIMD instructions allow a single instruction to be applied simultaneously to multiple data items. Used effectively, this can result in a significant performance increase. Typical applications include some graphics processing and digital signal processing. 

The XMM registers are as follows: 





Note, some of the more recent X86-64 processors support 256-bit XMM registers. This will not be an issue for the programs in this text.

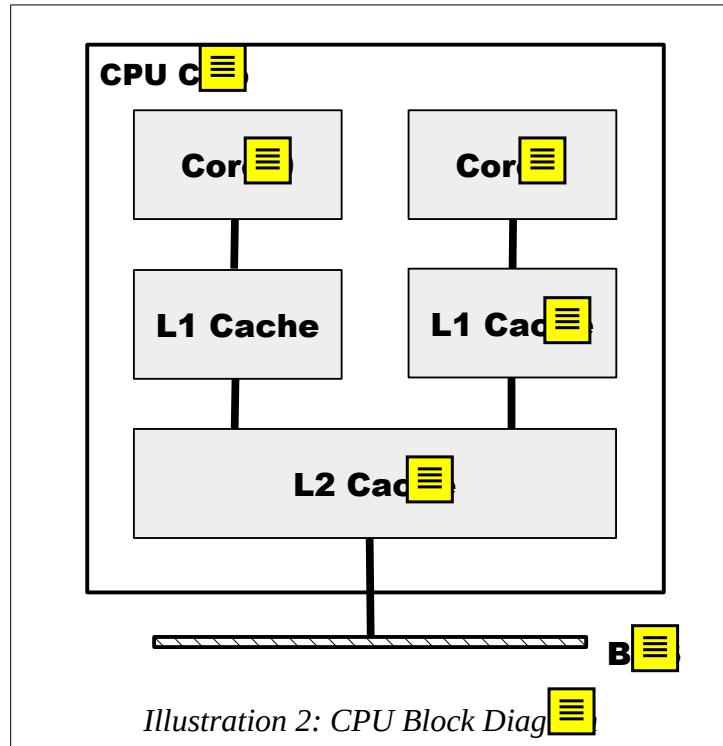
Additionally, the XMM registers are used to support the Streaming SIMD Extensions (SSE). The SSE instructions are out of the scope of this text. More information can be obtained from the Intel references (as noted in Chapter 1, Introduction).

### 2.3.2 Cache Memory

Cache memory is a small subset of the primary storage or RAM located in the CPU chip. If a memory location is accessed, a copy of the value is placed in the cache. Subsequent accesses to that memory location that occur in quick succession are retrieved from the cache location (internal to the CPU chip). A memory read involves sending the address via the bus to the memory controller, which will obtain the value at the requested memory location, and send it back through the bus. Comparatively, if a value is in cache, it would be much faster to access that value.

A cache hit occurs when the requested data can be found in a cache, while a cache miss occurs when it cannot. Cache hits are served by reading data from the cache, which is faster than reading from main memory. The more requests that can be served from cache, the faster the system will typically perform. Successive generations of CPU chips have increased cache memory and improved cache mapping strategies in order to improve overall performance.

A block diagram of a typical CPU chip configuration is as follows:



Current chip designs typically include an L1 cache per core and a shared L2 cache. Many of the newer CPU chips will have an additional L3 cache.

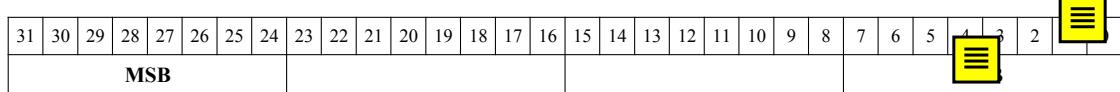
As can be noted from the diagram, all memory accesses travel through each level of cache. As such, there is a potential for multiple, duplicate copies of the value (CPU register, L1 cache, L2 cache, and main memory). This complication is managed by the CPU and is not something the programmer can change. Understanding the cache and associated performance gain is useful in understanding how a computer works.

## 2.4 Main Memory

Memory can be viewed as a series of bytes, one after another. That is, memory is *byte addressable*. This means each memory address holds one byte of information. To store a double-word, four bytes are required which use four memory addresses.

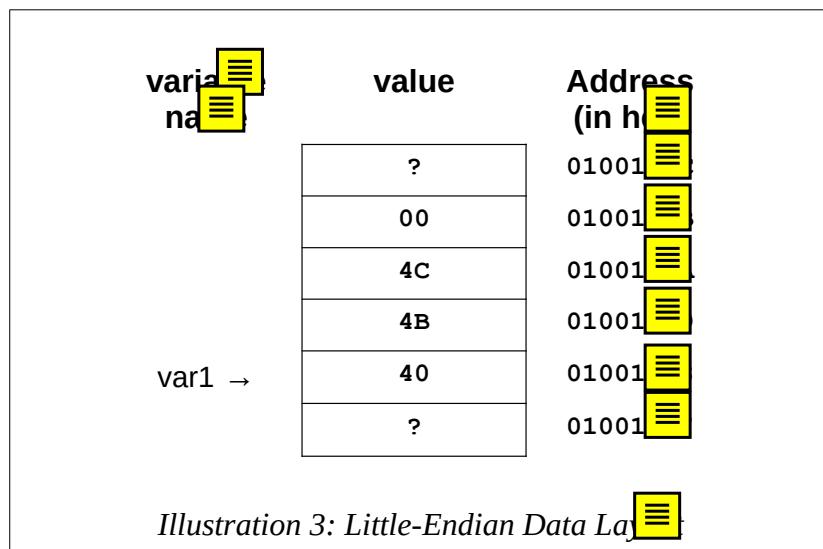
Additionally, architecture is **little-endian**. This means that the Least Significant Byte (LSB) is stored in the lowest memory address. The Most Significant Byte (MSB) is stored in the highest memory location.

For a double-word (32-bits), the MSB and LSB are allocated as shown below:



For example, assuming the value of,  $5,000,000_{10}$  ( $004C4B40_{16}$ ), is to be placed in a double-word variable named **var1**.

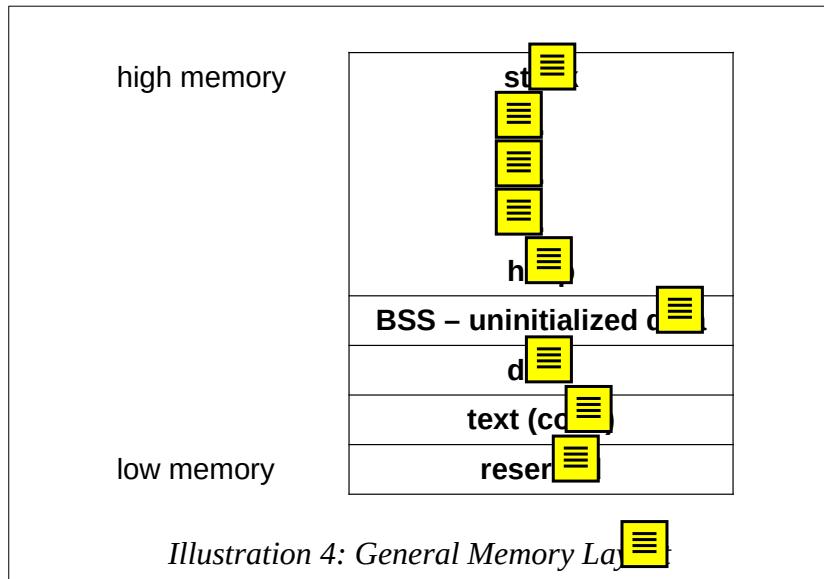
For a little-endian architecture, the memory picture would be as follows:



Based on the little-endian architecture, the LSB is stored in the lowest memory address and the MSB is stored in the highest memory location.

## 2.5 Memory Lay

The general memory layout for a program is as sho



The reserved section is not available to user programs. The text (or code) section is where the machine language<sup>11</sup> (i.e., the 1's and 0's that represent the code) is stored. The data section is where the initialized data is stored. This includes declared variables that have been provided an initial value at assemble-time. The uninitialized data section, typically called BSS section, is where declared variables that have not been provided an initial value are stored. If accessed before being set, the value will not be meaningful. The heap is where dynamically allocated data will be stored (if requested). The stack starts in high memory and grows downward.

Later sections will provide additional detail for the text and data secti

## 2.6 Memory Hierar

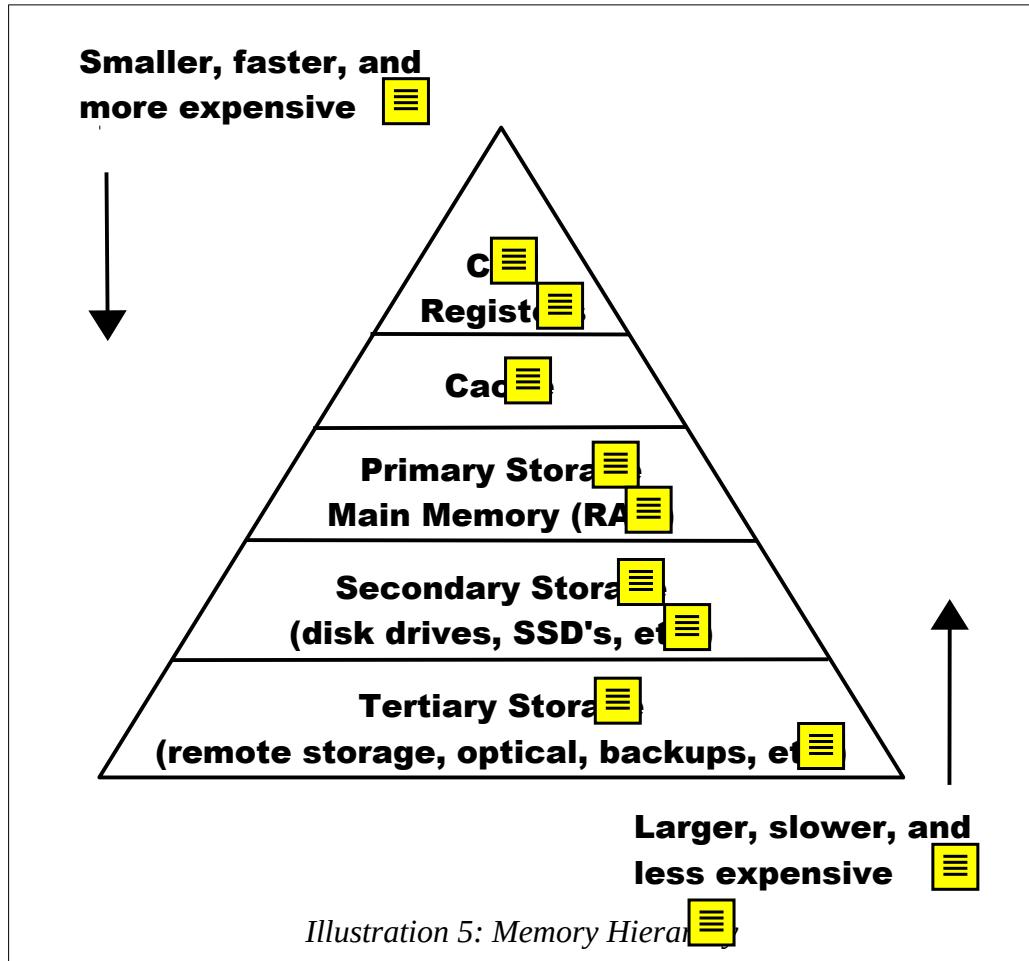
In order to fully understand the various different memory levels and associated usage, it is useful to review the memory hierarchy<sup>12</sup>. In general terms, faster memory is more expensive and slower memory blocks are less expensive. The CPU registers are small, fast, and expensive. Secondary storage devices such as disk drives and Solid S

<sup>11</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Machine\\_code](http://en.wikipedia.org/wiki/Machine_code)

<sup>12</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Memory\\_hierarchy](http://en.wikipedia.org/wiki/Memory_hierarchy)

Drives (SSD's) are larger, slower, and less expensive. The overall goal is to balance performance with cost.

An overview of the memory hierarchy is as follows:



Where the top of the triangle represents the fastest, smallest, and most expensive memory. As we move down levels, the memory becomes slower, larger, and less expensive. The goal is to use an effective balance between the small, fast, expensive memory and the large, slower, and cheaper memory.

Some typical performance and size characteristics are as follows:

Memory Unit	Example Size	Typical Speed
Registers	16, 64-bit registers	~1 nanosecond
Cache Memory	4 - 8+ Megabytes (L1 and L2)	~5-60 nanoseconds
Primary Storage (i.e., main memory)	2 – 32+ Gigabytes <sup>15</sup>	~100-150 nanoseconds
Secondary Storage (i.e., disk, SSD's, etc.)	500 Gigabytes – 4+ Terabytes <sup>16</sup>	~3-15 milliseconds <sup>17</sup>

Based on this table, a primary storage access at 100 nanoseconds ( $100 \times 10^{-9}$ ) is 30,000 times faster than a secondary storage access, at 3 milliseconds ( $3 \times 10^{-3}$ ).

The typical speeds improve over time (and these are already out of date). The key point is the relative difference between each memory unit is significant. This difference between the memory units applies even as newer, faster SSDs are being utilized.

## 2.7 Exercises

Below are some questions based on this chapter:

### 2.7.1 Quiz Questions

Below are some quiz questions:

- 1) Draw a picture of the Von Neumann Architecture.
- 2) What architecture component connects the memory to the CPU?
- 3) Where are programs stored when the computer is turned off?
- 4) Where must programs be located when they are executing?
- 5) How does cache memory help overall performance?
- 6) How many bytes does a C++ integer declared with the declaration `int` use?
- 7) On the Intel X86-64 architecture, how many **bytes** can be stored at each address?

13 For more information, refer to: <http://en.wikipedia.org/wiki/Nanosecond>

14 For more information, refer to: <http://en.wikipedia.org/wiki/Megabyte>

15 For more information, refer to: <http://en.wikipedia.org/wiki/Gigabyte>

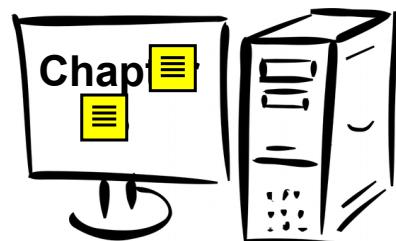
16 For more information, refer to: <http://en.wikipedia.org/wiki/Terabyte>

17 For more information, refer to: <http://en.wikipedia.org/wiki/Millisecond>

- 8) Given the 32-bit hex  $004C4B40_{16}$  what is ?
1. Least Significant Byte (L) 
  2. Most Significant Byte (M) 
- 9) Given the 32-bit hex  $004C4B40_{16}$ , show the little-endian memory layout showing each byte in memory .
- 10) Draw a picture of the layout for the **rax** register .
- 11) How many bits does each of the following represent? 
1. 
  2. 
  3. 
  4. 
  5. 
  6. 
  7. 
  8. **r1** 
- 12) Which register points to the next instruction to be executed .
- 13) Which register points to the current top of the stack .
- 14) If **al** is set to  $05_{16}$  and **ax** is set to  $0007_{16}$ , **eax** is set to  $00000020_{16}$ , and **rax** is set to  $0000000000000000_{16}$ , and show the final complete contents of the complete **rax** register.   
  
  

- 15) If the **rax** register is set to  $81,985,529,216,486,895_{10}$  ( $123456789ABCDE123456789ABCDE_{16}$ ), what are the contents of the following registers in hex? 
1. 
  2. 
  3. 
  4. 

*There are 10 types of people in the world;  
those that understand binary and those that  
don't.*



## 3.0 Data Representation

Data representation refers to how information is stored within the computer. There is a specific method for storing integers which is different than storing floating-point values which is different than storing characters. This chapter presents a brief summary of the integer, floating-point, and ASCII representation schemes.

It is assumed the reader is already generally familiar with binary, decimal, and hex numbering systems.

It should be noted that if not specified, a number is in base-10. Additionally, a number preceded by 0x is a hex value. For example,  $19 = 19_{10} = 13_{16} = 0x13$ .

### 3.1 Integer Representation

Representing integer numbers refers to how the computer stores or represents a number in memory. The computer represents numbers in binary (1's and 0's). However, the computer has a limited amount of space that can be used for each number or variable. This directly impacts the size, or range, of the number that can be represented. For example, a byte (8-bits) can be used to represent  $2^8$  or 256 different numbers. Those 256 different numbers can be *unsigned* (all positive) in which case we can represent any number between 0 and 255 (inclusive). If we choose *signed* (positive and negative values), then we can represent any number between -128 and +127 (inclusive).

If that range is not large enough to handle the intended values, a larger size must be used. For example, a word (16-bits) can be used to represent  $2^{16}$  or 65,536 different values, and a double-word (32-bits) can be used to represent  $2^{32}$  or 4,294,967,296 different numbers. So, if you wanted to store a value of 100,000 then a double-word would be required.

As you may recall from C, C++, or Java, an integer declaration (e.g., `int <variable>`) is a single double-word which can be used to represent values between  $(-2,147,483,648)$  and  $+2^{31} - 1 (+2,147,483,647)$ .

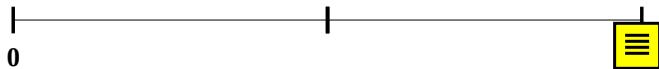
The following table shows the ranges associated with typical sizes:

Size	Size	Unsigned Range	Signed Range
Bytes (8-bits)	$2^8$	0 to 255	-128 to +127
Words (16-bits)	$2^{16}$	0 to 65,535	-32,768 to +32,767
Double-words (32-bits)	$2^{32}$	0 to 4,294,967,295	-2,147,483,648 to +2,147,483,647
Quadword	$2^{64}$	0 to $2^{64} - 1$	$-(2^{63})$ to $2^{63} - 1$
Double quadword	$2^{128}$	0 to $2^{128} - 1$	$-(2^{127})$ to $2^{127} - 1$

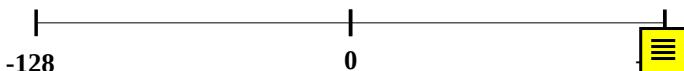
In order to determine if a value can be represented, you will need to know the size of the storage element (byte, word, double-word, quadword, etc.) being used and if the values are signed or unsigned.

- For representing *unsigned* values within the range of a given storage size, standard binary is used.
- For representing *signed* values within the range, **two's complement** is used. Specifically, the two's complement encoding process applies to the values in the negative range. For values within the positive range, standard binary is used.

For example, the unsigned byte range can be represented using a number line as follows:



For example, the signed byte range can also be represented using a number line as follows:



The same concept applies to halfwords and words which have larger ranges.

Since unsigned values have a different, positive only, range than signed values, there is overlap between the values. This can be very confusing when examining variables in memory (with the debugger).

For example, when the unsigned and signed values are within the overlapping positive range (0 to +127):

- A signed byte representation of  $12_{10}$  is  $0x0C$
- An unsigned byte representation of  $-12_{10}$  is also  $0x0C$

When the unsigned and signed values are outside the overlapping range:

- A signed byte representation of  $-15_{10}$  is  $0xE1$
- An unsigned byte representation of  $241_{10}$  is also  $0xE1$

This overlap can cause confusion unless the data types are clearly and correctly defined.

### 3.1.1 Two's Complement

The following describes how to find the two's complement representation for negative values (not positive values).

To take the two's complement of a number:

1. take the one's complement (negate)
2. add 1 (in binary)

The same process is used to encode a decimal value into two's complement and from two's complement back to decimal. The following sections provide some examples.

### 3.1.2 Byte Examples

For example, to find the byte size (8-bits), two's complement representation of -9 and 12.

9 ( $8+1$ ) =	00001001
Step 1	11110110
Step 2	11110111
-9 (in hex) =	F7

12 ( $8+4$ ) =	00001
Step 1:	11110
	11110
-12 (in hex) =	

Note, all bits for the given size, byte in this example, must be specified.

### 3.1.3 Word Examples

To find the word size (16-bits), two's complement representation of -18 and 40.

18 (16+2) =	<b>00000000000010010</b>
Step 1	<b>1111111111101101</b>
Step 2	<b>1111111111101110</b>
-18 (hex) =	<b>0xFFEE</b>

40 (32+8) =	<b>0000000000001010</b>
Step 1	<b>1111111111101010</b>
Step 2	<b>111111111110111</b>
-40 (hex) =	<b>0xF</b>

Note, all bits for the given size, words in these examples, must be specified.

### 3.2 Unsigned and Signed Addition

As previously noted, the unsigned and signed representations may provide different interpretations for the final value being represented. However, the addition and subtraction operations are the same. For example:

241	<b>11110001</b>
+ 7	<b>00000111</b>
248	<b>11111000</b>
248 =	<b>F8</b>

-15	<b>11110</b>
+ 7	<b>00000</b>
-8	<b>11111</b>
-8 =	

The final result of 0xF8 may be interpreted as 248 for unsigned representation and -8 for a signed representation. Additionally,  $0xF8_{16}$  is the  $^{\circ}$  (degree symbol) in the ASCII table.

As such, it is very important to have a clear definition of the sizes (byte, halfword, word etc.) and types (signed, unsigned) of data for the operations being performed.

### 3.3 Floating-point Representation

The representation issues for floating-point numbers are more complex. There are a series of floating-point representations for various ranges of the value. For simplicity we will look primarily at the IEEE 754 32-bit floating-point standard.

### 3.3.1 IEEE 32-bit Representation

The IEEE 754 32-bit floating-point standard is defined as follows:

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
s	biased exponent																												

Where  $s$  is the sign (0 => positive and 1 => negative). More formally, this can be written as;

$$N = (s \times 1.F) \times 2^E$$

When representing floating-point values, the first step is to convert floating-point value into binary. The following table provides a brief reminder of how binary handles fractional components:

	$2^3$	$2^2$	$2^1$	$2^0$	.	$2^{-1}$	$2^{-2}$	$2^{-3}$
...	8	4	2	1	.	1/2	1/4	1/8
	0	0	0	0	.	0	0	0

For example,  $100.101_2$  would be  $4.625_{10}$ . For repeating decimals, calculating the binary value can be time consuming. However, there is a limit since computers have finite storage sizes (32-bits in this example).

The next step is to show the value in normalized scientific notation in binary. This means that the number should have a single, non-zero leading digit to the left of the decimal point. For example,  $8.125_{10}$  is  $1000.001_2$  (or  $1000.001_2 \times 2^0$ ) and in binary normalized scientific notation that would be written as  $1.000001 \times 2^3$  (since the decimal point was moved three places to the left). Of course, if the number was  $0.125_{10}$  the binary would be  $0.001_2$  (or  $0.001_2 \times 2^0$ ) and the normalized scientific notation would be  $1.0 \times 2^{-3}$  (since the decimal point was moved three places to the right). The numbers after the leading 1, **not** including the leading 1, are stored left-justified in the fraction portion of the double-word.

The next step is to calculate the *biased exponent*, which is the exponent from the normalized scientific notation plus the bias. The bias for the IEEE 754 32-bit floating-point standard is  $127_{10}$ . The result should be converted to a byte (8-bits) and stored in the biased exponent portion of the word.

Note, converting from the IEEE 754 32-bit floating-point representation to the decimal value is done in reverse, however leading 1 must be added back (as it is not stored in the word). Additionally, the bias is subtracted (instead of added).

**3.3.1.1 IEEE 32-bit Representation Examples**

This section presents several examples of encoding and decoding floating-point representation for reference.

**3.3.1.1.1 Example → -7.75**

For example, to find the IEEE 754 32-bit floating-point representation for -7.75

**Example 1: -7.75**

- determine sign
  - convert to binary
  - normalized scientific notation
  - compute biased exponent
    - and convert to binary
  - write components in binary
- sign exponent mantissa  
1 10000001 11110000000000000000000000
- convert to hex (split into groups)
- 1100000011111000000000000000000000  
1100 0000 1111 1000 0000 0000 0000 0000  
C 0 F 8 0 0 0 0
- final result: C0F8 00

$$\begin{aligned}
 -7.75 &\Rightarrow 1 \text{ (since negative)} \\
 -7.75 &= -0.11111111_2 \\
 &= 1.11111111_2 \\
 2_{10} + 127_{10} &= 129_{10} \\
 &= 10000001_2
 \end{aligned}$$

**3.3.1.1.2 Example → -0.125**

For example, to find the IEEE 754 32-bit floating-point representation for -0.125

**Example 2: -0.125**

- determine sign
  - convert to binary
  - normalized scientific notation
  - compute biased exponent
    - and convert to binary
  - write components in binary
- sign exponent mantissa  
1 01111100 00000000000000000000000000
- convert to hex (split into groups)
- 10111110000000000000000000000000000  
1011 1110 0000 0000 0000 0000 0000 0000  
B E 0 0 0 0 0 0
- final result: BE00 00

$$\begin{aligned}
 -0.125 &\Rightarrow 1 \text{ (since negative)} \\
 -0.125 &= -0.10000000_2 \\
 &= 1.00000000_2 \\
 -3_{10} + 127_{10} &= 124_{10} \\
 &= 01111100_2
 \end{aligned}$$

### 3.3.1.1.3 Example → 414400

For example, given the IEEE 754 32-bit floating-point representation  $41440000_{16}$  find the decimal value:

#### Example 3: 414400

- convert to binary
- split into components
- determine exponent
  - and remove bias
- determine sign
- write result

0100 0001 0100 0100 0000 0000 0000 0  
0 10000010 1000100000000000000000000  
 $10000010_2 = 1100.01$   
 $130_{10} - 127 = 3$   
 $0 \Rightarrow \text{pos}$   
 $+1.10001 \times 2^3 = +1100.01 = +1$

### 3.3.2 IEEE 64-bit Representation

The IEEE 754 64-bit floating-point standard is defined as follows:

63	62		52	51		
s		biased exponent			frac	

The representation process is the same, however the format allows for an 11-bit biased exponent (which supports large and smaller values). The 11-bit biased exponent uses a bias of  $\pm 1023$ .

### 3.3.3 Not a Number (NaN)

When a value is interpreted as a floating-point value and it does not conform to the defined standard (either for 32-bit or 64-bit), then it cannot be used as a floating-point value. This might occur if an integer representation is treated as a floating-point representation or a floating-point arithmetic operation (add, subtract, multiply, or divide) results in a value that is too large or too small to be represented. The incorrect format or unrepresentable number is referred to as a NaN which is an abbreviation for *not a number*.

## 3.4 Characters and Strings

In addition to numeric data, symbolic data is often required. Symbolic or non-numeric data might include an important message such as “Hello World”<sup>18</sup> a common greeting for first programs. Such symbols are well understood by English language speakers.

<sup>18</sup> For more information, refer to: [http://en.wikipedia.org/wiki/%22Hello,\\_World!%22\\_program](http://en.wikipedia.org/wiki/%22Hello,_World!%22_program)



Computer memory is designed to store and retrieve numbers. Consequently, the symbols are represented by assigning numeric values to each symbol or character.



### 3.4.1 Character Representation



In a computer, a character<sup>19</sup> is a unit of information that corresponds to a symbol such as a letter in the alphabet. Examples of characters include letters, numerical digits, common punctuation marks (such as "." or "!"'), and whitespace. The general concept also includes control characters, which do not correspond to symbols in a particular language, but to other information used to process text. Examples of control characters include carriage return or tab.



#### 3.4.1.1 American Standard Code for Information Interchange



Characters are represented using the American Standard Code for Information Interchange (ASCII<sup>20</sup>). Based on the ASCII table, each character and control character is assigned a numeric value. When using ASCII, the character displayed is based on the assigned numeric value. This only works if everyone agrees on common values, which is the purpose of the ASCII table. For example, the letter "A" is defined as  $65_{10}$  (0x41). The 0x41 is stored in computer memory, and when displayed to the console, the letter "A" is shown. Refer to Appendix A for the complete ASCII table.



Additionally, numeric symbols can be represented in ASCII. For example, "9" is represented as  $57_{10}$  (0x39) in computer memory. The "9" can be displayed as output to the console. If sent to the console, the integer value  $9_{10}$  (0x09) would be interpreted as an ASCII value which in the case would be a tab.



It is very important to understand the difference between characters (such as "2") and integers (such as  $2_{10}$ ). Characters can be displayed to the console, but cannot be used for calculations. Integers can be used for calculations, but cannot be displayed to the console (without changing the representation).



A character is typically stored in a byte (8-bits) of space. This works well since memory is byte addressable.



19 For more information, refer to: [http://en.wikipedia.org/wiki/Character\\_\(computing\)](http://en.wikipedia.org/wiki/Character_(computing))

20 For more information, refer to: <http://en.wikipedia.org/wiki/ASCII>



### 3.4.1.2 Unicode

It should be noted that Unicode<sup>21</sup> is a current standard that includes support for different languages. The Unicode Standard provides series of different encoding schemes (UTF-8, UTF-16, UTF-32, etc.) in order to provide a unique number for every character, no matter what platform, device, application or language. In the most common encoding scheme, UTF-8, the ASCII English text looks exactly the same in UTF-8 as it did in ASCII. Additional bytes are used for other characters as needed. Details regarding Unicode representation are not addressed in this text.

### 3.4.2 String Representation

A string<sup>22</sup> is a series of ASCII characters, typically terminated with a NULL. The NULL is a non-printable ASCII control character. Since it is not printable, it can be used to mark the end of a string.

For example, the string “Hello” would be represented as follows:

Character	“H”	“e”	“l”	“l”	“o”	NULL
ASCII Value (decimal)	72	101	108	108	111	
ASCII Value (hex)	0x48	0x65	0x6C	0x6C	0x6F	

A string may consist partially or completely of numeric symbols. For example, the string “19653” would be represented as follows:

Character	“1”	“9”	“6”	“5”	“3”	NULL
ASCII Value (decimal)	49	57	54	53	51	
ASCII Value (hex)	0x31	0x39	0x36	0x35	0x33	

Again, it is very important to understand the difference between the string “19653” (using 6 bytes) and the single integer  $19,653_{10}$  (which can be stored in a single word which is 2 bytes).

## 3.5 Exercises

Below are some questions based on this chapter.

<sup>21</sup> For more information, refer to: <http://en.wikipedia.org/wiki/Unicode>

<sup>22</sup> For more information, refer to: [http://en.wikipedia.org/wiki/String\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/String_(computer_science))

### 3.5.1 Quiz Questions

Below are some quiz questions:

1) Provide the range for each of the following:

1. signed byte
2. unsigned byte
3. signed word
4. unsigned word
5. signed double-word
6. unsigned double-word

2) Provide the decimal values of the following binary numbers:

1. 00001
2. 00010
3. 00011
4. 00101

3) Provide the hex, **byte** size, two's complement values of the following decimal values. Note, two hex digits expected.

1. -1
2. +1
3. -2
4. +2

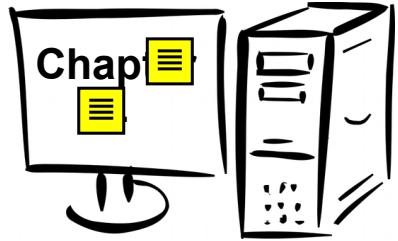
4) Provide the hex, **word** size, two's complement values of the following decimal values. Note, four hex digits expected.

1. -1
2. +1
3. -3
4. +3

- 5) Provide the hex, **double-word** size, two's complement values of the following decimal values. Note, eight hex digits expected.
1. -1
  2. -2
  3. -3
  4. -20
- 6) Provide the decimal values of the following hex, double-word sized, two's complement values.
1. FFFFFFFF
  2. FFFFFFFE
  3. FFFFFFFF
  4. FFFFFFFF
- 7) Which of the following decimal values has an **exact** representation in binary?
- 1.
  - 2.
  - 3.
  - 4.
  - 5.
- 8) Provide the decimal representation of the following IEEE 32-bit floating-point values.
1. 0xC14400
  2. 0x414400
  3. 0xC0D000
  4. 0xC0F000

- 9) Provide the IEEE 32-bit floating-point representation of the following floating-point values.
1. +11.2
  2. -17.12
  3. +21.87
  4. -0.7
- 10) What is the ASCII code, in hex, for each of the following characters?
1. ‘W’
  2. ‘1’
  3. ‘Y’
  4. ‘E’
  5. ‘o’
- 11) What are the ASCII values, in hex, for each of the following strings?
1. “World”
  2. “12345”
  3. “Yellow”

*I would love to change the world, but they won't give me the source code.*



## 4.0 Program Form

This chapter summarizes the formatting requirements for assembly language programs. The formatting requirements are specific to the **yasm** assembler. Other assemblers may be slightly different. A complete assembly language program is presented to demonstrate the appropriate program formatting.

A properly formatted assembly source file consists of several main parts:

- Data section where initialized data is declared and defined
- BSS section where uninitialized data is declared
- Text section where code is placed

The following sections summarize the basic formatting requirements. Only the basic formatting and assembler syntax are presented. For additional information, refer to the **yasm** reference manual (as noted in Chapter 1, Introduction).

### 4.1 Comments

The semicolon (;) is used to note program comments. Comments (using the ; ) may be placed anywhere, including after an instruction. Any characters after the ; are ignored by the assembler. This can be used to explain steps taken in the code or to comment out sections of code.

### 4.2 Numeric Values

Number values may be specified in decimal, hex, or octal.

When specifying hex, or base-16 values, they are preceded with a **0x**. For example, to specify 127 as hex, it would be **0x7f**.

When specifying octal, or-base-8 values, they are followed by a **q**. For example, to specify 511 as octal, it would be **777q**.

## Chapter 4.0 ◀ Program For [☰]

The default radix (base) is decimal, so no special notation is required for decimal (base 10) numbers.

### 4.3 Defining Constants [☰]

Constants are defined with `equ`. The general form is:

`<name> equ <val>`

The value of a constant cannot be changed during program execution.

The constants are substituted for their defined values during the assembly process. As such, a constant is not assigned a memory location. This makes the constant more flexible since it is not assigned a specific type/size (byte, word, double-word, etc.). The values are subject to the range limitations of the intended use. For example, the following constant,

`SIZE equ 10`

could be used as a word or a double-word, but not a byte.

### 4.4 Data Section [☰]

The initialized data must be declared in the "section .data" section. There must be a space after the word 'section'. All initialized variables and constants are placed in this section. Variable names must start with a letter, followed by letters or numbers, including some special characters (such as the underscore, "\_"). Variable definitions must include the name, the data type, and the initial value for the variable.

The general form is:

`<variableName> <dataType> <initialValue>`

Refer to the following sections for a series of examples using various data types.

The supported data types are as follows:

Declaration	
<code>db</code>	8-bit variable
<code>dw</code>	16-bit variable

<b>dd</b>	32-bit variable(s) → integer
<b>dq</b>	64-bit variable(s) → floating point
<b>ddq</b>	128-bit variable(s) → integer
<b>dt</b>	128-bit variable(s) → floating point

These are the primary assembler directives for initialized data declarations. Other directives are referenced in different sections.

Initialized arrays are defined with comma separated values.

Some simple examples include:

<b>bVar</b>	<b>db</b>	10	; byte variable
<b>cVar</b>	<b>db</b>	"H"	; single character
<b>strng</b>	<b>db</b>	"Hello World"	; string
<b>wVar</b>	<b>dw</b>	5000	; 16-bit variable
<b>dVar</b>	<b>dd</b>	50000	; 32-bit variable
<b>arr</b>	<b>dd</b>	100, 200, 300	; 3 element array
<b>flt1</b>	<b>dd</b>	3.14159	; 32-bit float
<b>qVar</b>	<b>dq</b>	1000000000	; 64-bit variable

The value specified must be able to fit in the specified data type. For example, if the value of a byte sized variables is defined as 500, it would generate an assembler error.

## 4.5 BSS Section

Uninitialized data is declared in the "section .bss" section. There must be a space after the word 'section'. All uninitialized variables are declared in this section. Variable names start with a letter followed by letters or numbers including some special characters (such as the underscore, "\_"). Variable definitions must include the name, the data type, and the count.

The general form is:

**<variableName>**      **<resType>**      **<count>**

Refer to the following sections for a series of examples using various data types.

The supported data types are as follows:

Declar	
<b>resb</b>	8-bit variab
<b>resw</b>	16-bit variab
<b>resd</b>	32-bit variab
<b>resq</b>	64-bit variab
<b>resdq</b>	128-bit variab

These are the primary assembler directives for uninitialized data declarations. Other directives are referenced in different sections.

Some simple examples incl

```
bArr      resb      10      ; 10 element byte arr
wArr      resw      50      ; 50 element word arr
dArr      resd      100     ; 100 element double arr
qArr      resq      200     ; 200 element quad arr
```

The allocated array is not initialized to any specific va

## 4.6 Text Sect

The code is placed in the "section .text" section. There must be a space after the word 'section'. The instructions are specified one per line and each must be a valid instruction with the appropriate required operands.

The text section will include some headers or labels that define the initial program entry point. For example, assuming a basic program using the standard system linker, the following declarations must be included.

```
global _start
_start:
```

No special label or directives are required to terminate the program. However, a system service should be used to inform the operating system that the program should be terminated and the resources, such as memory, recovered and re-utilized. Refer to the example program in the following section.

## 4.7 Example Program

A very simple assembly language program is presented to demonstrate the appropriate program formatting.

```
; Simple example demonstrating basic program format and layout

; Ed Jorgensen
; July 18, 2018

; *****
; Some basic data declarations

section    .data

; -----
; Define constants

EXIT_SUCCESS    equ      0          ; successful operation
SYS_exit        equ      60         ; call code for termination

; -----
; Byte (8-bit) variable declarations

bVar1           db       17
bVar2           db       0
bResult         db       0

; -----
; Word (16-bit) variable declarations

wVar1           dw       17000
wVar2           dw       9000
wResult         dw       0

; -----
; Double-word (32-bit) variable declarations

dVar1           dd       17000000
dVar2           dd       90000000
dResult         dd       0
```

## Chapter 4.0 ◀ Program Format

```
; -----  
; quadword (64-bit) variable declaration  
  
qVar1      dq      1700000000  
qVar2      dq      90000000  
qResult    dq      0  
  
*****  
;  
; Code Section  
  
section      .text  
global _start  
_start:  
  
; Performs a series of very basic addition operations  
; to demonstrate basic program format.  
  
;  
; Byte example  
; bResult = bVar1 + bV  
  
    mov     al, byte [bVar1]  
    add     al, byte [bVar2]  
    mov     byte [bResult],  
  
;  
; Word example  
; wResult = wVar1 + wV  
  
    mov     ax, word [wVar1]  
    add     ax, word [wVar2]  
    mov     word [wResult],  
  
;  
; Double-word example  
; dResult = dVar1 + dV  
  
    mov     eax, dword [dVar1]  
    add     eax, dword [dVar2]  
    mov     dword [dResult],
```

```

; -----
; Quadword example
; qResult = qVar1 + qV[=]
;           mov      rax, qword [qVar1]
;           add      rax, qword [qVar2]
;           mov      qword [qResult], [=]

; ****
; Done, terminate program.[=]

last:
    mov      rax, SYS_exit      ; Call code for exit
    mov      rdi, EXIT_SUCCESS   ; Exit program with success
    syscall

```

This example program will be referenced and further explained in the following chapters.

## 4.8 Exercises

Below are some questions based on this chapter.

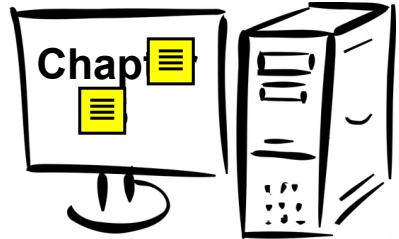
### 4.8.1 Quiz Questions

Below are some quiz questions.

- What is the name of the assembler being used in this chapter?
- How are comments marked in an assembly language program?
- What is the name of the section where the initialized data declarations are placed?
- What is the name of the section where the uninitialized data declarations are placed?
- What is the name of the section where the code is placed?
- What is the data declaration for each of the following variables with the given values?
  - byte sized variable **bNum** set to 1
  - word sized variable **wNum** set to 10,290
  - double-word sized variable **dwNum** set to 2,126,012,900
  - quadword sized variable **qwNum** set to 10,000,000,000

- 7) What is the uninitialized data declaration for each of the following?
1. byte sized array named **bArr** with 100 elements
  2. word sized array named **wArr** with 3000 elements
  3. double-word sized array named **dwArr** with 200 elements
  4. quadword sized array named **qArr** with 5000 elements
- 8) What are the required declarations to signify the start of a program (in the **section**)?

*There are two ways to write error-free programs; only the third works.*



## 5.0 Tool Chain

In general, the set of programming tools used to create a program is referred to as the **tool chain**<sup>23</sup>. For the purposes of this text, the tool chain consists of the following:

- Assem
- Link
- Load
- Debug

While there are many options for the tool chain, this text uses a fairly standard set of open source tools that work well together and fully support the x86 64-bit environment.

Each of these programming tools is explained in the following sections.

### 5.1 Assemble/Link/Load Overview

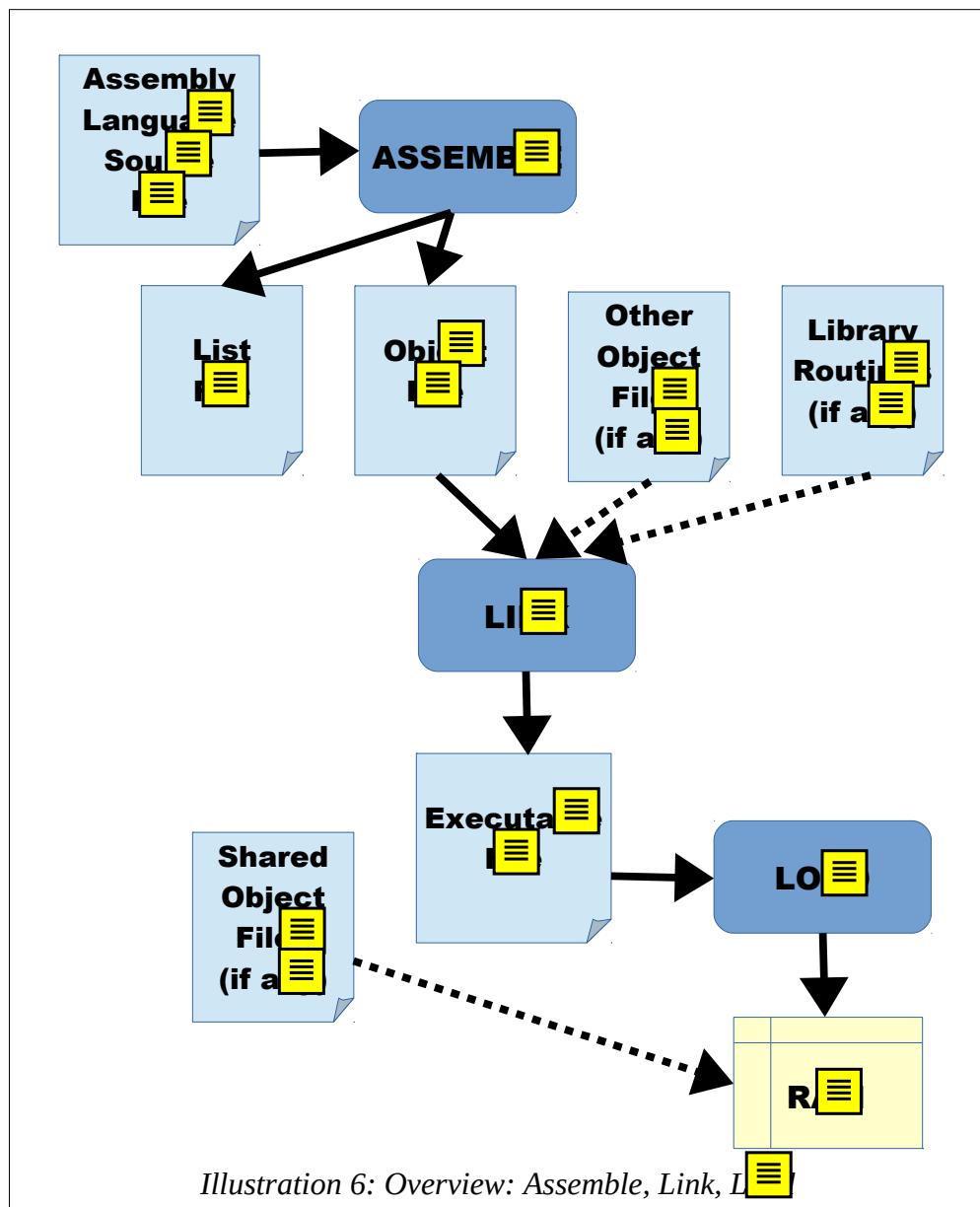
In broad terms, the assemble, link, and load process is how programmer written source files are converted into an executable program.

The human readable source file is converted into an object file by the assembler. In the most basic form, the object file is converted into an executable file by the linker. The loader will load the executable file into memory.

---

<sup>23</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Tool\\_chain](http://en.wikipedia.org/wiki/Tool_chain)

An overview of the process is provided in the following diagram:



The assemble, link, and load steps are described in more detail in the following section.

## 5.2 Assembler

The assembler<sup>24</sup> is a program that will read an assembly language input file and convert the code into a machine language binary file. The input file is an assembly language source file containing assembly language instructions in human readable form. The machine language output is referred to as an object file. As part of this process, the comments are removed, and the variable names and labels are converted into appropriate addresses (as required by the CPU during execution).

The assembler used in this text is the **yasm**<sup>25</sup> assembler. Links to the **yasm** web site and documentation can be found in Chapter 1, Introduction

### 5.2.1 Assemble Command

The appropriate **yasm** assembler command for reading the assembly language source file, such as the example from the previous chapter, is as follows:

```
yasm -g dwarf2 -f elf64 example.asm -l example.lst
```

Note, the **-l** is a dash lower-case letter L (which is easily confused with the number 1).

The **-g dwarf2**<sup>26</sup> option is used to inform the assembler to include debugging information in the final object file. This increases the size of the object file, but is necessary to allow effective debugging. The **-f elf64** informs the assembler to create the object file in the **ELF64**<sup>27</sup> format which is appropriate for 64-bit, Linux-based systems. The **example.asm** is the name of the assembly language source file for input. The **-l example.lst** (dash lower-case letter L) informs the assembler to create a list file named **example.lst**.

If an error occurs during the assembly process, it must be resolved before continuing to the link step.

### 5.2.2 List

In addition, the assembler is optionally capable of creating a list file. The list file shows the line number, the relative address, the machine language version of the instruction (including variable references), and the original source line. The list file can be useful when debugging.

For example, a fragment from the list file data section, from the example program in the previous chapter is as follows:

<sup>24</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Assembler\\_\(computing\)#Assembler](http://en.wikipedia.org/wiki/Assembler_(computing)#Assembler)

<sup>25</sup> For more information, refer to: <https://en.wikipedia.org/wiki/Yasm>

<sup>26</sup> For more information, refer to: <https://en.wikipedia.org/wiki/DWARF>

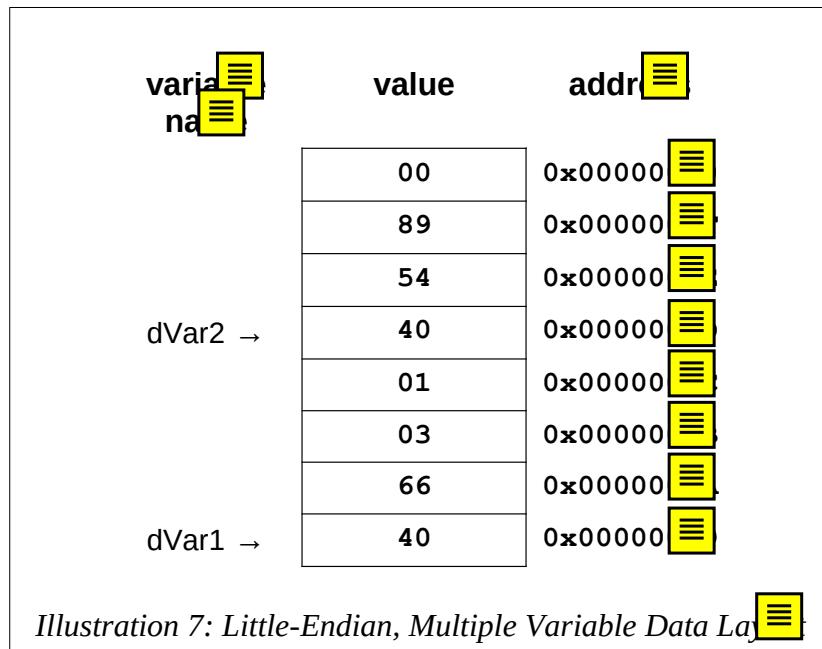
<sup>27</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Executable\\_and\\_Linkable\\_Format](http://en.wikipedia.org/wiki/Executable_and_Linkable_Format)

36 00000009 40660301 37 0000000D 40548900 38 00000011 00000000	dVar1 dd 17000000 dVar2 dd 900000A dResult dd 0
--	---

On the first line, the **36** is the line number. The next number, **0x00000009**, is the relative address in the data area of where that variable will be stored. Since *dVar1* is a double-word, which requires four bytes, the address for the next variable is **0x0000000D**. The *dVar1* variable uses 4 bytes as addresses **0x00000009**, **0x0000000A**, **0x0000000B**, and **0x0000000C**. The rest of the line is the data declaration as typed in the original assembly language source file.

The **0x40660301** is the value, in hex, as placed in memory. The  $17,000,000_{10}$  is **0x01036640**. Recalling that the architecture is little-endian, the least significant byte (**0x40**) is placed in the lowest memory address. As such, the **0x40** is placed in relative address **0x00000009**, the next byte, **0x66**, is placed in address **0x0000000A** and so forth. This can be confusing as at first glance the number may appear backwards or garbled (depending on how it is viewed).

To help visualize, the memory picture would be as follows:



For example, a fragment of the list file text section, excerpted from the example program in the previous chapter is as follows:

```
95
96 0000005A 48C7C03C000000
97 00000061 48C7C300000000
98 00000068 0F05
                                last:
                                mov      rax, SYS_exit
                                mov      rdi, EXIT_SUCCESS
                                syscall
```

Again, the numbers to the left are the line numbers. The next number, **0x0000005A**, is the relative address of where the line of code will be placed.

The next number, **0x48C7C03C000000**, is the machine language version of the instruction, in hex, that the CPU reads and understands. The rest of the line is the original assembly language source instruction.

The label, **last:**, does not have a machine language instruction since the label is used to reference a specific address and is not an executable instruction.

### 5.2.3 Two-Pass Assembly

The assembler<sup>28</sup> will read the source file and convert each assembly language instruction, typed by the programmer, into a set of 1's and 0's that the CPU knows to be that instruction. The 1's and 0's are referred to as machine language. There is a one-to-one correspondence between the assembly language instructions and the binary machine language. This relationship means that machine language, in the form of an executable file can be converted back into human readable assembly language. Of course, the comments, variable names, and label names are missing, so the resulting code can be very difficult to read.

As the assembler reads each line of assembly language, it generates machine code for that instruction. This will work well for instructions that do not perform jumps. However, for instructions that might change the control flow (e.g., IF statements, unconditional jumps), the assembler is not able to convert the instruction. For example, given the following code fragment:

```
mov      rax
jmp      skipRest
skipRe
```

This is referred to as a forward reference. If the assembler reads the assembly file one line at a time, it has not read the line where *skipRest* is defined. In fact, it does not even know for sure if *skipRest* is defined at all.

This situation can be resolved by reading the assembly source file twice. The first

<sup>28</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Assembly\\_language#Assembler](http://en.wikipedia.org/wiki/Assembly_language#Assembler).

process is referred to as a two-pass assembler. The steps required for each pass are detailed in the following sections.

### 5.2.3.1 First

The steps taken on the first pass vary based on the design of the specific assembler. However, some of the basic operations performed on the first pass include the following:

- Create symbol table
- Expand macro
- Evaluate constant expression

A macro is a program element that is expanded into a set of programmer predefined instructions. For more information, refer to Chapter 11, Macros.

A constant expression is an expression composed entirely of constants. Since the expression is constants only, it can be fully evaluated at assemble-time. For example, assuming the constant BUFF is defined, the following instruction contains a constant expression;

```
mov      rax, BUF
```

This type of constant expression is used commonly in large or complex programs.

Addresses are assigned to all statements in the program. The symbol table is a listing or table of all the program symbols, variable names and program labels, and their respective addresses in the program.

As appropriate, some assembler directives are processed in the first pass.

### 5.2.3.2 Second

The steps taken on the second pass vary based on the design of the specific assembler. However, some of the basic operations performed on the second pass include the following:

- Final generation of code
- Creation of list file (if requested)
- Create object file

The term code generation refers to the conversion of the programmer provided assembly language instruction into the CPU executable machine language instruction. Due to the one-to-one correspondence, this can be done for instructions that do not use symbols on either the first or second pass.

It should be noted that, based on the assembler design, much of the code generation might be done on the first pass or all done on the second pass. Either way, the final generation is performed on the second pass. This will require using the symbol table to check program symbols and obtain the appropriate addresses from the table.

The list file, while optional, can be useful for debugging. If requested, it would be generated on the second pass.

If there are no errors, the final object file is created on the second pass.

#### 5.2.4 Assembler Directives

Assembler directives are instructions to the assembler that direct the assembler to do something. This might be formatting or layout. These directives are not translated into instructions for the CPU.

### 5.3 Linker

The linker<sup>29</sup>, sometimes referred to as linkage editor, will combine one or more object files into a single executable file. Additionally, any routines from user or system libraries are included as necessary. The GNU gold linker, **ld**<sup>30</sup>, is used. The appropriate linker command for the example program from the previous chapter is as follows:

```
ld -g -o example example.o
```

Note, the **-o** is a dash lower-case letter O, which can be confused with the number zero.

The **-g** option is used to inform the linker to include debugging information in the final executable file. This increases the size of the executable file, but is necessary to allow effective debugging. The **-o example** specifies to create the executable file named *example* (with no extension). If the **-o <fileName>** option is omitted, the output file is named *a.out* (by default). The *example.o* is the name of the input object file read by the linker. It should be noted that the executable file could be named anything and does not need to have the same name as any of the input object files.

<sup>29</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Linker\\_\(computing\)](http://en.wikipedia.org/wiki/Linker_(computing))

<sup>30</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Gold\\_\(linker\)](http://en.wikipedia.org/wiki/Gold_(linker))

### 5.3.1 Linking Multiple F

In programming, large problems are typically solved by breaking them into smaller problems. The smaller problems can be addressed individually, possibly by different programmers.

Additional input object files, if any, would be listed, in order, separated with a space. For example, if there are two object files, *main.o* and *funcs.o* the link command to create an executable file name *example*, with debugging information included, would be as follows:

```
ld -g -o example main.o func
```

This would typically be required for larger or more complex programs.

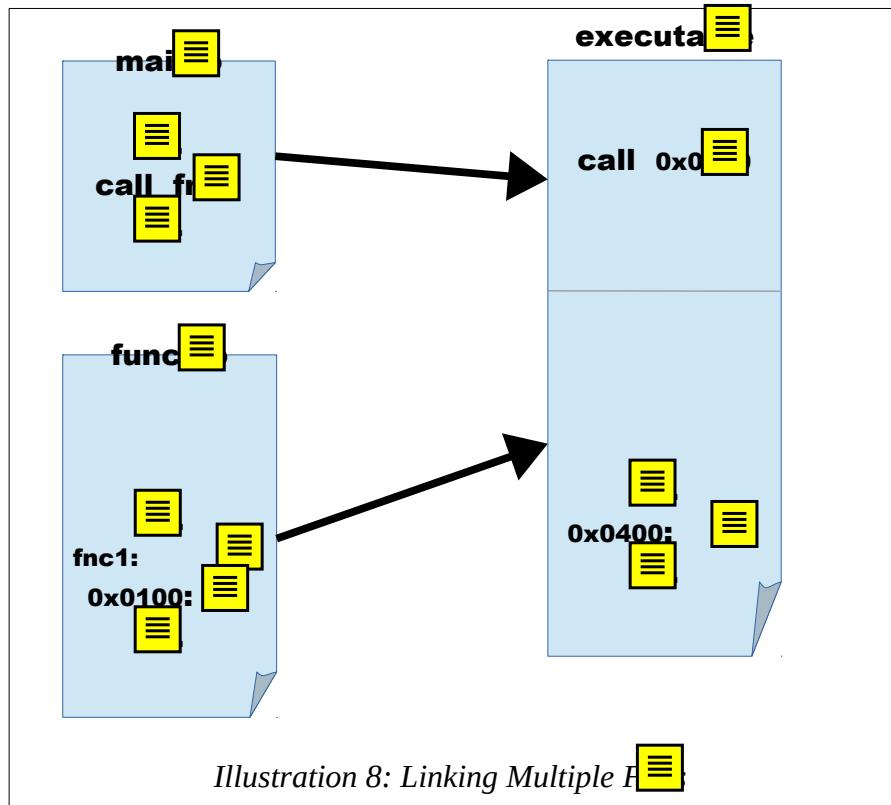
When using functions located in a different, external source file, any function or functions not in the current source file must be declared as **extern**. Variables, such as global variables, in other source files can be accessed by using the **extern** statement as well, however data is typically transferred as arguments of the function call.

### 5.3.2 Linking Programs

Linking is the fundamental process of combining the smaller solutions into a single executable unit. If any user or system library routines are used, the linker will include the appropriate routines. The object files and library routines are combined into a single executable module. The machine language code is copied from each object file into a single executable.

As part of combining the object files, the linker must adjust the relocatable addresses as necessary. Assuming there are two source files, the main and a secondary source file containing some functions, both of which have been assembled into object files *main.o* and *funcs.o*. When each file is assembled, the calls to routines outside the file being assembled are declared with the external assembler directive. The code is not available for an external reference and such references are marked as external in the object file. The list file will show an **R** for such relocatable addresses. The linker must satisfy the external references. Additionally, the final location of the external references must be placed in the code.

For example, if the *main.o* object file calls a function in the *funcs.o* file, the linker must update the call with the appropriate address as shown in the following illustration.



Here, the function `fnc1` is external to the `main.o` object file and is marked with an R. The actual function `fnc1` is in the `funcs.o` file, which starts its relative addressing from `0x0` (in the text section) since it does not know about the main code. When the object files are combined, the original relative address of `fnc1` (shown as `0x0100:`) is changed to its final address in executable file (shown as `0x0400:`). The linker must insert this final address into the call statement in the `main` (shown as `call 0x0400:`) in order to complete the linking process and ensure the function call will work correctly.

This will occur with all relocatable addresses for both code and data.

### 5.3.3 Dynamic Linking

The Linux operating system supports dynamic linking<sup>31</sup>, which allows for postponing the resolution of some symbols until a program is being executed. The actual instructions are not placed in executable file and instead, if needed, resolved and accessed at run-time.

---

<sup>31</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Dynamic\\_linking](http://en.wikipedia.org/wiki/Dynamic_linking)

While more complex, this approach offers two advantages:

- Often-used libraries (e.g., the standard system libraries) can be stored in only one location, not duplicated in every single binary.
- If a bug in a library function is corrected, all programs using it dynamically will benefit from the correction (at the next execution). Otherwise, programs that utilize this function by static linking would have to be re-linked before the correction is applied.

There are also disadvantages:

- An incompatible updated library will break executable's that depended on the behavior of the previous version of the library.
- A program, together with the libraries it uses, might be certified (e.g. as to correctness, documentation requirements, or performance) as a package, but not if components can be replaced.

In Linux/Unix, the dynamically linked object files typically have **.so** (shared object) extension. In Windows, they have a **.dll** (dynamically linked library) extension. Further details of dynamic linking are outside the scope of this text.

## 5.4 Assemble/Link Script

When programming, it is often necessary to type the assemble and link commands many times with various different programs. Instead of typing the assemble (**yasm**) and link (**ld**) commands each time, it is possible to place them in a file, called a script file. Then, the script file can be executed which will just execute the commands that were entered in the file. While not required, using a script file can save time and make things easier when working on a program.

A simple example bash<sup>32</sup> assemble/link script is as follows:

```
#!/bin/bash
# Simple assemble/link script

if [ -z $1 ]; then
    echo "Usage: ./asm64 <asmMainFile> (no extension)"
    exit

```

---

<sup>32</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Bash\\_\(Unix\\_shell\)](http://en.wikipedia.org/wiki/Bash_(Unix_shell))

```

# Verify no extensions were entered
if [ ! -e "$1.asm" ]; then
    echo "Error, $1.asm not found."
    echo "Note, do not enter file extensions"
    exit
# Compile, assemble, and link
yasm -Worphan-labels -g dwarf2 -f elf64 $1.asm -l $1.list
ld -g -o $1 $1.o

```

The above script should be placed in a file. For this example, the file will be named ***asm64*** and placed in the current working directory (where the source files are located).

Once created, execute privilege will need to be added to the script file as follows:

```
chmod +x asm64
```

This will only need to be done once for each script.

The script file will read the source file name from the command line. For example, to use the script file to assemble the example from the previous chapter (named ***example.asm***), type the following:

```
./asm64 example
```

The ".asm" extension on the ***example.asm*** file should not be included (since it is added in the script). The script file will assemble and link the source file, creating the list file, object file, and executable file.

Use of this, or any script file, is optional. The name of the script file can be changed as desired.

## 5.5 Loading

The loader<sup>33</sup> is a part of the operating system that will load the program from secondary storage into primary storage (i.e., main memory). In broad terms, the loader will attempt to find, and if found, read a properly formatted executable file, create a new process, and load the code into memory and mark the program as ready for execution. The operating system scheduler will make the decisions about which process is executed and when the process is executed.

---

33 For more information, refer to: [http://en.wikipedia.org/wiki/Loader\\_\(computing\)](http://en.wikipedia.org/wiki/Loader_(computing))

The loader is implicitly invoked by typing the program name. For example, on the previous example program, named *example*, the Linux command would be:

```
./example
```

which will execute the file named *example* created via the previous steps (assemble and link). Since the example program does not perform any output, nothing will be displayed to the console. As such, a debugger can be used to check the results.

## 5.6 Debug

The debugger<sup>34</sup> is used to control execution of a program. This allows for testing and debugging activities to be performed.

In the previous example, the program computed a series of calculations, but did not output any of the results. The debugger can be used to check the results. The executable file is created with the assemble and link command previously described and must include the -g option.

The debugger used is the GNU DDD debugger which provides a visual front-end for the GNU command line debugger, **gdb**. The DDD web site and documentation are noted in the references section of Chapter 1, Introduction.

Due to the complexity and importance of the debugger, a separate chapter for the debugging is provided.

## 5.7 Exercise

Below are some questions based on this chapter.

### 5.7.1 Quiz Questions

Below are some quiz questions.

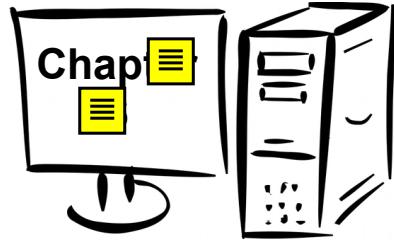
- 1) What is the relationship between assembly language and machine language?
- 2) What actions are performed on the first pass of the assembler?
- 3) What actions are performed on the second pass of the assembler?
- 4) What actions are performed by the linker?
- 5) What actions are performed by the loader?

<sup>34</sup> For more information, refer to: <http://en.wikipedia.org/wiki/Debugger>.

- 6) Provide an example of a *constant expression* 
- 7) Draw a diagram of the entire assemble, link, and load process 
- 8) When is a shared object file linked with a program? 
- 9) What is contained in the symbol table (two things)? 



*My software never has bugs. It just develops random features.*



## 6.0 DDD Debugging

A debugger allows the user to control execution of a program, examine variables, other memory (i.e., stack space), and display program output (if any). The open source GNU Data Display Debugger (DDD<sup>35</sup>) is a visual front-end to the GNU Debugger (GDB<sup>36</sup>) and is widely available. Other debuggers can easily be used if desired.

Only the basic debugger commands are addressed in this chapter. The DDD debugger has many more features and options not covered here. As you gain experience, it would be worth reviewing the DDD documentation, referenced in Chapter 1, to learn more about additional features in order to help improve overall debugging efficiency.

DDD functionality can be extended using various plug-ins. The plug-ins are not required and will not be addressed in this Chapter.

This chapter addresses using the GNU DDD debugger as a tool. The logical process of how to debug a program is not addressed in this chapter.

### 6.1 Starting DDD

The **ddd** debugger is started with the executable file. The program must be assembled and linked with the correct options (as noted in the previous chapter). For example, using the previous sample program, *example*, the command would be:

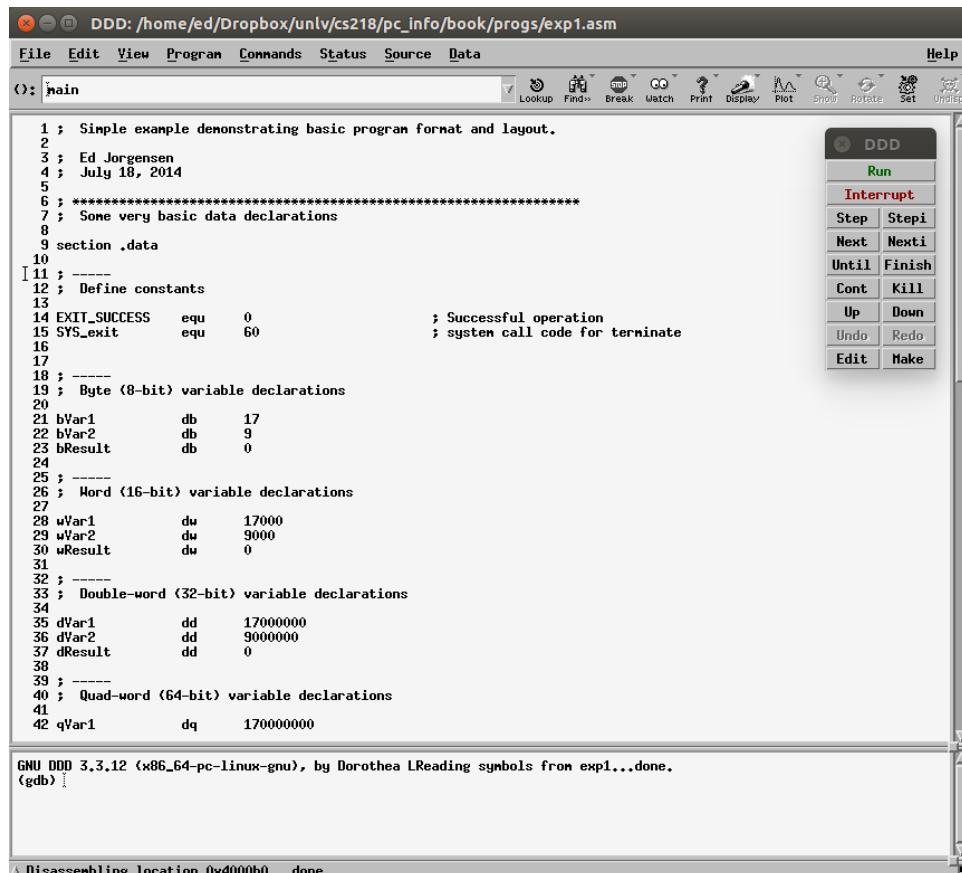
```
ddd example
```

Upon starting DDD/GDB, something similar to the screen, shown below, should be displayed (with the appropriate source code displayed).

<sup>35</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Data\\_Display\\_Debugger](http://en.wikipedia.org/wiki/Data_Display_Debugger)

<sup>36</sup> For more information, refer to: [http://en.wikipedia.org/wiki/GNU\\_Debugger](http://en.wikipedia.org/wiki/GNU_Debugger)

## Chapter 6.0 ◀ DDD Debug



The screenshot shows the DDD (Debug) interface. The main window displays assembly code for a file named `exp1.asm`. The code includes declarations for constants, byte, word, double-word, and quad-word variables. A vertical line on the left margin indicates the current instruction being executed. The status bar at the bottom shows "Disassembling location 0x4000b0...done." The menu bar includes File, Edit, View, Program, Commands, Status, Source, Data, Help, and a toolbar with various icons. On the right, there is a sidebar with buttons for DDD, Run, Interrupt, Step, Next, Until, Cont, Up, Undo, and Edit.

```
1 ; Simple example demonstrating basic program format and layout.
2
3 ; Ed Jorgensen
4 ; July 18, 2014
5
6 ; ****
7 ; Some very basic data declarations
8
9 section .data
10
11 ; -----
12 ; Define constants
13
14 EXIT_SUCCESS    equ    0          ; Successful operation
15 SYS_exit        equ    60         ; system call code for terminate
16
17
18 ; -----
19 ; Byte (8-bit) variable declarations
20
21 bVar1           db     17
22 bVar2           db     9
23 bResult         db     0
24
25 ; -----
26 ; Word (16-bit) variable declarations
27
28 wVar1           dw     17000
29 wVar2           dw     9000
30 wResult         dw     0
31
32 ; -----
33 ; Double-word (32-bit) variable declarations
34
35 dVar1           dd     17000000
36 dVar2           dd     9000000
37 dResult         dd     0
38
39 ; -----
40 ; Quad-word (64-bit) variable declarations
41
42 qVar1           dq     1700000000000000
```

GNU DDD 3.3.12 (x86\_64-pc-linux-gnu), by Dorothea LReading symbols from exp1...done.  
(gdb)

Disassembling location 0x4000b0...done.

Illustration 9: Initial Debugger Scr

If the code is not displayed in a similar manner as shown above, the assemble and link steps should be verified. Specifically, the `-g` qualifier must be included in both the assemble and link steps.

Built in help is available by clicking on the Help menu item (upper right-hand corner). The DDD and GDB manuals are available from the virtual library link on the class web page. To exit DDD/GDB, select **File → Exit** (from the top menu bar).

### 6.1.1 DDD Configuration Settings

Some additional DDD/GDB configuration settings suggestions include:

**Edit → Preferences → General → Suppress X Warnings**

**Edit → Preferences → Source → Display Source Line Numbers**

These are not required, but can make using the debugger easier. If set, the options will be saved and remembered for successive uses of the debugger (on the same machine).

## 6.2 Program Execution with DDD

To execute the program, click on the **Run** button from the command tool menu (shown below). Alternately, you can type **run** at the (gdb) prompt (bottom GDB console window). However, this will execute the program entirely and, when done, the results will be reset (and lost).

### 6.2.1 Setting Breakpoints

In order to control program execution, it will be necessary to set a breakpoint (execution pause location) to pause the program at a user selected location. This can be done by selecting the source location (line to stop at). For this example, we will stop at line 9.

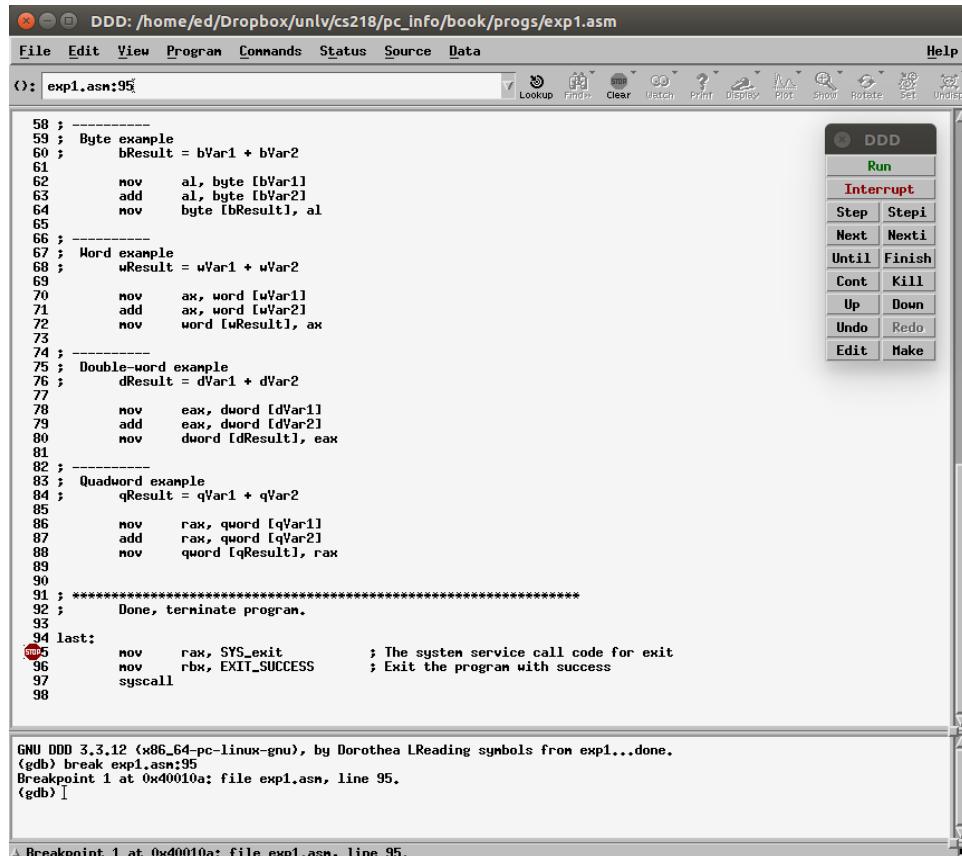
The breakpoint can be done one of three ways:

- Right click on the line number and select: *Set Breakpoint*
- In the GDB Command Console, at the (gdb) prompt, type: **break**
- In the GDB Command Console, at the (gdb) prompt, type: **b**

In the following example, line 94 is a label with no instruction. If a breakpoint is set on this label, it will stop at the next executable instruction (line 95 in this example).

## Chapter 6.0 ◀ DDD Debug

When set correctly, the “stop” icon will appear to the left of line number (as shown in the diagram).



The screenshot shows the DDD (Debian Debugger) interface. The main window displays assembly code for a file named `exp1.asm`. The code includes various examples of arithmetic operations (byte, word, double-word, quadword) and a termination section. A red circle highlights the line number 95, indicating a breakpoint has been set. The bottom window shows the GDB command-line interface with the following text:  
GNU DDD 3.3.12 (x86\_64-pc-linux-gnu), by Dorothea LReading symbols from exp1...done.  
(gdb) break exp1.asm:95  
Breakpoint 1 at 0x40010a: file exp1.asm, line 95.  
(gdb) [ ]

Illustration 10: Debugger Screen with Breakpoint

DDD/GDB commands can be typed inside the bottom window (at the **(gdb)** prompt) at any time. Multiple breakpoints can be set if desired.

### 6.2.2 Executing Program

Once the debugger is started, in order to effectively use the debugger, an initial breakpoint must be set.

Once the breakpoint is set, the run command can be performed via clicking **Run** menu window or typing **run** at the (gdb) prompt. The program will execute up to, *but not including* the statement with the green arrow.

The screenshot shows the DDD debugger interface. The main window displays assembly code for `exp1.asm`. A green arrow points to the instruction at line 95, which is a `mov` instruction. The code includes examples for byte, word, double-word, and quadword operations. The bottom window shows the gdb command-line interface with the following session:

```
Breakpoint 1 at 0x40010a: file exp1.asm, line 95.
(gdb) run
Starting program: /home/ed/Dropbox/unlv/cs218/pc_info/book/progs/exp1
Breakpoint 1, last () at exp1.asm:95
(gdb) [redacted]
Disassembling location 0x40010a...done.
```

Illustration 11: Debugger Screen with Green Arrow

The breakpoint is indicated with the stop sign on the left and the current location is indicated with a green arrow (see example above). Specifically, the green arrow points to the *next instruction to be executed*. That is, the statement pointed to by the green arrow has **not** yet been executed.



### 6.2.2.1 Run / Cont



As needed, additional breakpoints can be set. However, click the **Run** command will re-start execution from the beginning and stop at the initial breakpoint.



Illustration 12: DDD Command



After the initial **Run** command, to continue to the next breakpoint, the continue command must be used (by clicking **Cont** menu window or typing **cont** at the (gdb) prompt). Single lines can also be executed one line at a time by typing the step or next commands (via clicking **Step** or **Next** menu window or typing **step** or **next** at the (gdb) prompt).



### 6.2.2.2 Next / S



The **next** command will execute to the next instruction. This includes executing an entire function if necessary. The **step** command will execute one step, stepping into functions if necessary. For a single, non-function instruction, there is no difference between the **next** and **step** commands.



### 6.2.3 Displaying Register Content



The simplest method to see the contents of the registers is to use the registers window. The registers window is not displayed by default, but can be viewed by selecting **Status → Registers** (from the top menu bar). When displayed, the register window will show register contents by register name (left column), in both hex (middle column) and unsigned decimal (right column). Since the right column will display the unsig



value of the entire register it can be confusing when the data is signed (as it will be displayed as unsigned). Additionally, for some registers, such as **rbp** and **rsp**, both columns are shown in hex (since they are typically used for addresses). The examine memory command will as described in the following sections allows a more specific control over what format (e.g., signed, unsigned, hex) in which to display values.



Illustration 13: Register Window

Depending on the machine and screen resolution, the register window may need to be resized to view the entire contents.

The third column of the register window generally shows the decimal quadword representation except for some special purpose registers (**rbp** and **rsp**). The signed quadword decimal representation may not always be meaningful. For example, if unsigned data is being used (such as addresses), the signed representation would be incorrect. Additionally, when character data is used, the signed representation would not be meaningful.

By default, only the integer registers are displayed. Clicking on the “All registers” box will add the floating-point registers to the display. Viewing will require scrolling down within the register window.

## 6.2.4 DDD/GDB Commands Summary

The following table provides a small subset of the most common DDD commands. When typed, most commands may be abbreviated. For example, **quit** can be abbreviated as **q**. The command and the abbreviation are shown in the table.

Command	Description
<b>quit   q</b>	Quit the debugger
<b>break &lt;label/address&gt;</b>   <b>b &lt;label/address&gt;</b>	Set a break point (stop point) at <label> or <address>.
<b>run &lt;args&gt;   r &lt;args&gt;</b>	Execute the program (to the first breakpoint)
<b>continue   c</b>	Continue execution (to the next breakpoint)
<b>continue &lt;n&gt;   c &lt;n&gt;</b>	Continue execution (to the next breakpoint), skipping <b>n</b> -1 crossing of the breakpoint. This is can be used to quickly get to the <b>n</b> <sup>th</sup> iteration of a loop.
<b>step   s</b>	Step into next instruction (i.e., steps into function/procedure calls).
<b>next   n</b>	Next instruction (steps through function/procedure calls)
<b>F3</b>	Re-start program (and stop at first breakpoint)
<b>where</b>	Current activation (call depth)
<b>x/&lt;n&gt;&lt;f&gt;&lt;u&gt; \$rsp</b>	Examine contents of the stack

Command	Description
<code>x/&lt;n&gt;&lt;f&gt;&lt;u&gt; &amp;&lt;variable&gt;</code>	Examine memory location <variable> <n> number of locations to display, 1 is default. <f> format: d – decimal (sign) [ ] x – [ ] u – decimal (unsigned) [ ] c – character [ ] s – string [ ] f – floating-point [ ] <u> unit size: b – byte (8-bit) [ ] h – halfword (16-bit) [ ] w – word (32-bit) [ ] g – giant (64-bit) [ ]
<code>source &lt;filename&gt;</code>	Read commands from file <filename> [ ]
<code>set logging file &lt;filename&gt;</code>	Set logging file to <filename>, default is <i>gdb.txt</i> . [ ]
<code>set logging on</code>	Turn logging (to a file) [ ]
<code>set logging off</code>	Turn logging (to a file) [ ]
<code>set logging overwrite</code>	When logging (to a file) is turned on, overwrite previous log file (if any). [ ]

More information can be obtained via the built-in help facility or from the documentation on the **ddd** website (referenced from Chapter 1). [ ]

#### 6.2.4.1 DDD/GDB Commands, Exam [ ]

For example, given the below data declarati [ ]

```
bnum1      db      [ ]  
wnum2      dw      -2 [ ]  
dnum3      dd      100 [ ]
```

## Chapter 6.0 ◀ DDD Debug

qnum	dq	1234567
class	db	"Assembly"
twopi	dd	6

Assuming *signed data*, the commands to examine memory commands would be as follows:

```
x/db &bnum1  
x/dh &wnum2  
x/dw &dnum3  
x/dg &qnum  
x/s &class  
x/f &twopi
```

If an inappropriate memory dump command is used (i.e., incorrect size), *there is no error message* and the debugger will display what was requested (even if it does not make sense). Examining variables will require use of the appropriate memory dump command based on the data declarations. Additional options can be accessed across the menu at the top of the screen.

To display an array in DDD, the basic examine memory command is u

```
x/<n><f><u> &<variable>
```

For example, assuming the declaration

```
list1 dd 100001, -100002, 100003, 100004, 100005
```

The examine memory commands would be as follows

```
x/5dw &list1
```

Where the **5** is the array length. The **d** indicates signed data (**u** would have been unsigned data). The **w** indicates 32-bit sized data (which is what the **dd**, define double, definition declares in the source file). The **&list1** refers to the address of the variable. Note, the address points to the first element (and only the first element). As such, it is possible to display less or more elements than are actually declared in the array.

The basic examine memory command can be used with a memory address directly (as opposed to a variable name). For example:

```
x/dw 0x600
```

Addresses are typically displayed in hexadecimal, so a **0x** would be required in order

enter the hexadecimal address directly as `sho`

## 6.2.5 Displaying Stack Content

There are some occasions when displaying the contents of the stack may be useful. The stack is normally comprised of 64-bit, unsigned elements. The examine memory command is used, however the address is in the `rsp` register (not a variable name). The examine memory command to display the current top of the stack would be as follows

```
x/ug $
```

The examine memory command to display the top 6 items on the stack would be as follows:

```
x/6ug $
```

Due to the stack implementation, the first item shown will always be current top of the stack.

## 6.2.6 Debugger Commands File (interactive)

Since the data display commands must be correct (since there is no error), it can be tedious. To help reduce errors, the correct execution and display command can be stored in a text file. The debugger can then read the commands from the file (instead of typing them by hand). While the results are typically displayed to the screen, the results can be redirected to an output file. This can be useful for easy review.

For example, some typical debugger commands to set the breakpoint, run the program, display some variables, and redirect the output to a log file might be as follows:

```
#-----
# Debugger Input Script
#-----
echo \n\n
break last
run
set pagination off
set logging file out.txt
set logging overwrite
set logging on
set prompt
echo ----- \n
echo display variables \n
echo \n
x/100dw &list
```

```
x/dw &length  
echo \n  
x/dw &listMin  
x/dw &listMid  
x/dw &listMax  
x/dw &listSum  
x/dw &listAve  
echo \n \n  
set logging off  
quit
```

Note 1; this example assumes a label '**last**' is defined in the source program (as is done on the example program).

Note 2; this example exits the debugger. If that is not desired, the '**quit**' command can be removed. When exiting from the input file, the debugger may request user confirmation of the exit (yes or no).

These commands should be placed in a file (such as *gdbIn.txt*), so they can be read from within the debugger.

### 6.2.6.1 Debugger Commands File (non-interactive)

The debugger command to read a file is "source <filename>". For example, if the command file is named *gdbIn.txt*,

```
(gdb) source gdbIn.
```

Based on the above commands, the output will be placed in the file *out.txt*. The output file name can be changed as desired.

Each program will require a custom set of input command based on the specific variables and associated sizes in that program. The debugger input commands file will only be useful when the program is fairly close to working. Program crashes and other more significant errors will require interactive debugging to determine the specific error or errors.

### 6.2.6.2 Debugger Commands File (non-interactive)

It is possible to obtain the output file directly without an interactive DDD session. The following command, entered at the command line, will execute the command in the input file on the given program, create the output file, and exit the program.

```
gdb <gdbIn.txt p
```

Which will create the output file (as specified in the *gdbIn.txt* input file) and exit the debugger. Once the input file is created, this is the fastest option for obtaining the final output file for a working program. Again, this would only be useful if the program is working or very close to working correctly.



## 6.3 Exercises

Below are some quiz questions based on this chapter.

### 6.3.1 Quiz Questions

Below are some quiz questions.

- 1) How is the debugger started (from the command line)?
- 2) What option is required during the assemble and link step in order to ensure the program be easily debugged?
- 3) What does the **run** command do specifically?
- 4) What does the **continue** command do specifically?
- 5) How is the register window displayed?
- 6) There are three columns in the register window. The first shows the register name. What do the other two columns show?
- 7) Once the debugger is started, how can the user exit?
- 8) Describe how a breakpoint is set (multiple ways).
- 9) What is the debugger command to read debugger commands from a file?
- 10) When the DDD shows a green arrow pointing to an instruction, what does it mean?
- 11) Provide the debugger command to display each of the following variables:  
decimals:
  1. **bVar1** (byte sized variable)
  2. **wVar1** (word sized variable)
  3. **dVar1** (double-word sized variable)
  4. **qVar1** (quadword sized variable)
  5. **bArr1** (30 element array of bytes)
  6. **wArr1** (50 element array of words)
  7. **dArr1** (75 element array of double-words)



12) Provide the debugger command to display each of the following variable in hexadecimal form

1. **bVar1** (byte sized variable)
2. **wVar1** (word sized variable)
3. **dVar1** (double-word sized variable)
4. **qVar1** (quadword sized variable)
5. **bArr1** (30 element array of bytes)
6. **wArr1** (50 element array of words)
7. **dArr1** (75 element array of double-words)

13) What is the debugger command to display the value at the current top of stack?

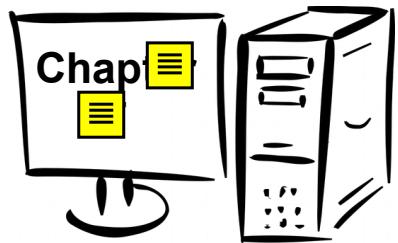
14) What is the debugger command to display five (5) values at the current top of stack?

### 6.3.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Type in the example program from Chapter 4, Program Format. Assemble and link the program as described in Chapter 5, Tool Chain. Execute the debugger as noted in this chapter. Set a breakpoint at the label last and execute the program (to that breakpoint). Interactively verify that the calculations performed resulted in the correct values. This will require typing the appropriate debugger examination memory commands (based on the variable size).
- 2) After completing the previous problem, create a debugger input file that will send the output to a text file, set a breakpoint, execute the program, and display the results for each variable (based on the appropriate variable size). Execute the debugger and read the source file. Review the input file worked correctly and that the program calculations are correct based on the results shown in the output file.
- 3) Create an assemble and link script file, as described in Chapter 5, Tool Chain. Use the script to assemble and link the program. Ensure that the script correctly assembles and links.

*Why are math books sad?  
Because they have so many problems.*



## 7.0 Instruction Set Overview

This chapter provides a basic overview for a simple subset of the x86-64 instruction set focusing on the integer operations. This will cover only the subset of instructions required for the topics and programs discussed within the scope of this text. This will exclude some of the more advanced instructions and restricted mode instructions. For a complete listing of all processor instructions, refer to the references listed in Chapter 12.

The instructions are presented in the following order:

- Data Movements
- Conversion Instructions
- Arithmetic Instructions
- Logical Instructions
- Control Instructions

The instructions for function calls are discussed in the chapter in Chapter 12, Functions.

A complete listing of the instructions covered in this text is located in Appendix B for reference.

### 7.1 Notational Conventions

This section summarizes the notation used within this text which is fairly common in the technical literature. In general, an instruction will consist of the instruction or operation itself (i.e., add, sub, mul, etc.) and the **operands**. The operands refer to where the data (to be operated on) is coming from and/or where the result is to be placed.



### 7.1.1 Operand Notation



The following table summarizes the notational conventions used in the remainder of the document.



Operand Notation	Description
<b>&lt;reg&gt;</b>	Register operand. The operand must be a register.
<b>&lt;reg8&gt;, &lt;reg16&gt;, &lt;reg32&gt;, &lt;reg64&gt;</b>	Register operand with specific size requirement. For example, <b>reg8</b> means a byte sized register (e.g., <b>al</b> , <b>bl</b> , etc.) only and <b>reg32</b> means a double-word sized register (e.g., <b>eax</b> , <b>ebx</b> , etc.) only.
<b>&lt;dest&gt;</b>	Destination operand. The operand may be a register or memory. Since it is a destination operand, the contents will be overwritten with the new result (based on the specific instruction).
<b>&lt;RXdest&gt;</b>	Floating-point destination register operand. The operand must be a floating-point register. Since it is a destination operand, the contents will be overwritten with the new result (based on the specific instruction).
<b>&lt;src&gt;</b>	Source operand. Operand value is unchanged after the instruction.
<b>&lt;imm&gt;</b>	Immediate value. May be specified in decimal, hex, octal or binary.
<b>&lt;mem&gt;</b>	Memory location. May be a variable name or an indirect reference (i.e., a memory address).
<b>&lt;op&gt; or &lt;operand&gt;</b>	Operand, register or memory.
<b>&lt;op8&gt;, &lt;op16&gt;, &lt;op32&gt;, &lt;op64&gt;</b>	Operand, register or memory, with specific size requirement. For example, <b>op8</b> means a byte sized operand only and <b>reg32</b> means a double-word sized operand only.
<b>&lt;label&gt;</b>	Program label.

By default, the immediate values are decimal or base-10. Hexadecimal or base-16 immediate values may be used but must be preceded with a **0x** to indicate the value is hex. For example,  $15_{10}$  could be entered in hex as **0x0F**.



Refer to Chapter 8, Addressing Modes for more information regarding memory locations and indirection.

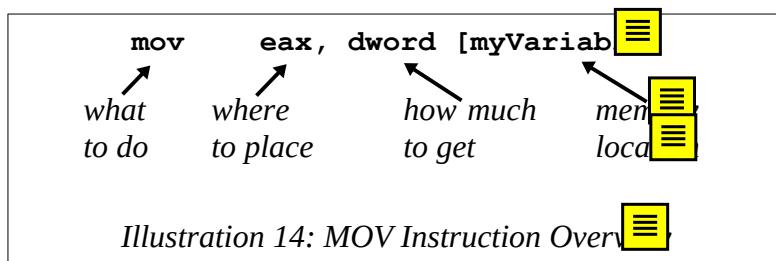
## 7.2 Data Movement

Typically, data must be moved into a CPU register from RAM in order to be operated upon. Once the calculations are completed, the result may be copied from the register and placed into a variable. There are a number of simple formulas in the example program that perform these steps. This basic data movement operation is performed with the move instruction.

The general form of the move instruction:

```
mov <dest>, <source>
```

The source operand is copied from the source operand into the destination operand. The value of the source operand is unchanged. The destination and source operand must be of the same size (both bytes, both words, etc.). The destination operand cannot be an immediate. Both operands cannot be memory. If a memory to memory operation is required, two instructions must be used.



When the destination register operand is of double-word size and the source operand is of double-word size, the upper-order double-word of the quadword register is set to zero. This only applies when the destination operand is a double-word sized integer register.

Specifically, if the following operations are performed:

```
mov eax, 100          ; eax = 0x000000  
mov rcx, -1           ; rcx = 0xffffffffffff  
mov ecx, eax          ; ecx = 0x000000
```

Initially, the **rcx** register is set to -1 (which is all 0xF's). When the positive number from the **eax** register ( $100_{10}$ ) is moved into the **rcx** register, the upper-order portion of the quadword register **rcx** is set to 0 over-writing the 1's from the previous instruction.



The move instruction is summarized as follows:



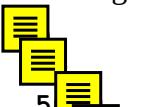
Instruction	Explanation
<code>mov &lt;dest&gt;, &lt;src&gt;</code>	<p>Copy source operand to the destination operand.</p> <p><i>Note 1</i>, both operands cannot be memory.</p> <p><i>Note 2</i>, destination operands cannot be an immediate.</p> <p><i>Note 3</i>, for double-word destination and source operand, the upper-order portion of the quadword register is set to 0.</p>
Examples:	<code>mov ax, 42</code> <code>mov cl, byte [bVar]</code> <code>mov dword [dVar],</code> <code>mov qword [qVar],</code>

A more complete list of the instructions is located in Appendix E.



For example, assuming the following data declarations:



<code>dValue</code>	<code>dd</code>	
<code>bNum</code>	<code>db</code>	
<code>wNum</code>	<code>dw</code>	
<code>dNum</code>	<code>dd</code>	
<code>qNum</code>	<code>dq</code>	
<code>bAns</code>	<code>db</code>	
<code>wAns</code>	<code>dw</code>	
<code>dAns</code>	<code>dd</code>	
<code>qAns</code>	<code>dq</code>	

To perform, the basic operations:



```

dValue = 27
bAns = bNum
wAns = wNum
dAns = dNum
qAns = qNum
    
```

The following instructions could be used:



```

mov     dword [dValue], 27 ; dValue = 27
    
```



```

mov    al, byte [bN] ; bAns = b
mov    byte [bAns], al

mov    ax, word [wN] ; wAns = w
mov    word [wAns], ax

mov    eax, dword [dN] ; dAns = d
mov    dword [dAns], eax

mov    rax, qword [qN] ; qAns = q
mov    qword [qAns], rax

```

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) can be omitted as the other operand will clearly define the size. In the text it will be included for consistency and good programming practice.

### 7.3 Addresses and Val

The only way to access memory is with the brackets ([]'s). Omitting the brackets will not access memory and instead obtain the address of the item. For example:

```

mov    rax, qword [var1] ; value of var1 in
mov    rax, var1          ; address of var1 in

```

Since omitting the brackets is not an error, the assembler will not generate error messages or warnings. This can lead to confusion.

In addition, the address of a variable can be obtained with the load effective address or **lea**, instruction. The load effective address instruction is summarized as follows:

Instruction	Explanation
<b>lea &lt;reg64&gt;, &lt;mem&gt;</b>	Place address of <mem> into reg
Examples:	<pre>         lea    rcx, byte [bvar1]         lea    rsi, dword [dvar1]       </pre>

A more complete list of the instructions is located in Appendix E.

Additional information and extensive examples are presented in Chapter 8, Addressing Modes.



## 7.4 Conversion Instructions



It is sometimes necessary to convert from one size to another size. For example, a byte might need to be converted to a double-word for some calculations in a formula. The process used for conversions depends on the size and type of the operand. The following sections summarize how conversions are performed.



### 7.4.1 Narrowing Conversions



Narrowing conversions are converting from a larger type to a smaller type (i.e., word to byte or double-word to word).



No special instructions are needed for narrowing conversions. The lower portion of the memory location or register may be accessed directly. For example, if the value of 50 (0x32) is placed in the **rax** register, the **al** register may be accessed directly to obtain the value as follows:

```
mov    rax, [bVal]  
mov    byte [bVal], al
```



This example is reasonable since the value of 50 will fit in a byte value. However, if the value of 500 (0x1f4) is placed in the **rax** register, the **al** register can still be accessed.



```
mov    rax, [bVal]  
mov    byte [bVal], al
```



In this example, the **bVal** variable will contain 0xf4 which may lead to incorrect results. The programmer is responsible for ensuring that narrowing conversions are performed appropriately. Unlike a compiler, no warnings or error messages will be generated.



### 7.4.2 Widening Conversions



Widening conversions are from a smaller type to a larger type (e.g., byte to word or word to double-word). Since the size is being expanded, the upper-order bits must be set based on the sign of the original value. As such, the data type, signed or unsigned, must be known and the appropriate process or instructions must be used.



#### 7.4.2.1 Unsigned Conversions



For unsigned widening conversions, the upper part of the memory location or register must be set to zero. Since an unsigned value can only be positive, the upper-order bits can only be zero. For example, to convert the byte value of 50 in the **al** register, to a quadword value in **rbx**, the following operations can be performed.



```

mov      al,
mov      rbx
mov      bl,

```

Since the **rbx** register was set to 0 and then the lower 8-bits were set to the value from (50 in this example), the entire 64-bit **rbx** register is now 50.

This general process can be performed on memory or other registers. It is the programmer's responsibility to ensure that the values are appropriate for the data sizes being used.

An unsigned conversion from a smaller size to a larger size can also be performed with a special move instruction, as follows:

```
movzx    <dest>, <s
```

Which will fill the upper-order bits with zero. The **movzx** instruction does not allow a quadword destination operand with a double-word source operand. As previously noted, a **mov** instruction with a double-word register destination operand with a double-word source operand will zero the upper-order double-word of the quadword destination register.

A summary of the instructions that perform the unsigned widening conversion are as follows:

Instruction	Explanation
<b>movzx</b> <dest>, <src> <b>movzx</b> <reg16>, <op8> <b>movzx</b> <reg32>, <op8> <b>movzx</b> <reg32>, <op16> <b>movzx</b> <reg64>, <op8> <b>movzx</b> <reg64>, <op16>	Unsigned widening conversion. <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operands cannot be an immediate. <i>Note 3</i> , immediate values not allowed.
Examples:	<b>movzx</b> cx, byte [bVar] <b>movzx</b> dx, al <b>movzx</b> ebx, word [wVar] <b>movzx</b> ebx, cx <b>movzx</b> rbx, cl <b>movzx</b> rbx, cx

A more complete list of the instructions is located in Appendix E.



### 7.4.2.2 Signed Conversions



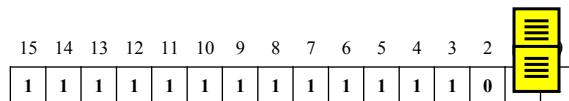
For signed widening conversions, the upper-order bits must be set to either 0's or 1's depending on if the original value was positive or negative.



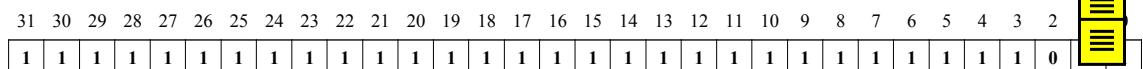
This is performed by a sign-extend operation. Specifically, the upper-order bit of the original value indicates if the value is positive (with a 0) or negative (with a 1). The upper-order bit of the original value is extended into the higher bits of the new, widened value.



For example, given that the **ax** register is set to -7 (0xffff9), the bits would be set as follows:



Since the value is negative, the upper-order bit (bit 15) is a 1. To convert the word value in the **ax** register into a double-word value in the **eax** register, the upper-order bit (1 in this example) is extended or copied into the entire upper-order word (bits 31–16) resulting in the following:



There are a series of dedicated instructions used to convert signed values in the **A** register from a smaller size into a larger size. These instructions work only on the **A** register, sometimes using the **D** register for the result. For example, the **cwd** instruction will convert a signed value in the **ax** register into a double-word value in the **dx** (upper-order portion) and **ax** (lower-order portion) registers. This is typically by convention written as **dx:ax**. The **cwde** instruction will convert a signed value in the **ax** register into a double-word value in the **eax** register.



A more generalized signed conversion from a smaller size to a larger size can also be performed with some special move instructions, as follows:



```
movsx      <dest>, <s>
movsxd    <dest>, <s>
```



Which will perform the sign extension operation on the source argument. The **movsx** instruction is the general form and the **movsxd** instruction is used to allow a quadword destination operand with a double-word source operand.



A summary of the instructions that perform the signed widening conversion are as follows:

Instruction	Explanation
<b>cbw</b>	Convert byte in <b>al</b> into word in <b>ax</b> . <i>Note</i> , only works for <b>al</b> to <b>ax</b> register.
Examples:	
<b>cwd</b>	Convert word in <b>ax</b> into double-word in <b>dx:ax</b> . <i>Note</i> , only works for <b>ax</b> to <b>dx:ax</b> registers.
Examples:	
<b>cwde</b>	Convert word in <b>ax</b> into double-word in <b>eax</b> . <i>Note</i> , only works for <b>ax</b> to <b>eax</b> register.
Examples:	
<b>cdq</b>	Convert double-word in <b>eax</b> into quadword in <b>edx:eax</b> . <i>Note</i> , only works for <b>eax</b> to <b>edx:eax</b> registers.
Examples:	
<b>cdqe</b>	Convert double-word in <b>eax</b> into quadword in <b>rax</b> . <i>Note</i> , only works for <b>rax</b> register.
Examples:	
<b>cqo</b>	Convert quadword in <b>rax</b> into word in double-quadword in <b>rdx:rax</b> . <i>Note</i> , only works for <b>rax</b> to <b>rdx:rax</b> registers.
Examples:	

Instruction	Explanation
<b>movsx</b> <dest>, <src> <b>movsx</b> <reg16>, <op8> <b>movsx</b> <reg32>, <op8> <b>movsx</b> <reg32>, <op16> <b>movsx</b> <reg64>, <op8> <b>movsx</b> <reg64>, <op16> <b>movsxd</b> <reg64>, <op32>	Signed widening conversion (via sign extension). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operands cannot be an immediate. <i>Note 3</i> , immediate values not allowed. <i>Note 4</i> , special instruction ( <i>movsxd</i> ) required for 32-bit to 64-bit signed extension.
Examples:	<b>movsx</b> cx, byte [bVar] <b>movsx</b> dx, al <b>movsx</b> ebx, word [wVar] <b>movsx</b> ebx, cx <b>movsxd</b> rbx, dword [dVar]

A more complete list of the instructions is located in Appendix E.

## 7.5 Integer Arithmetic Instructions

The integer arithmetic instructions perform arithmetic operations such as addition, subtraction, multiplication, and division on integer values. The following sections present the basic integer arithmetic operations.

### 7.5.1 Addition

The general form of the integer addition instruction is as follows:

**add**    <dest>, <source>

Where operation performs the following:

$$<\text{dest}> = <\text{dest}> + <\text{source}>$$

Specifically, the source and destination operands are added and the result is placed in the destination operand (over-writing the previous contents). The value of the source operand is unchanged. The destination and source operand must be of the same size (both bytes, both words, etc.). The destination operand cannot be an immediate. Both operands, cannot be memory. If a memory to memory addition operation is required, two instructions must be used.

For example, assuming the following data declarati

bNum1	db	
bNum2	db	
bAns	db	
wNum1	dw	
wNum2	dw	
wAns	dw	
dNum1	dd	
dNum2	dd	
dAns	dd	
qNum1	dq	
qNum2	dq	
qAns	dq	

To perform the basic operations

```
bAns = bNum1 + bNum2  
wAns = wNum1 + wNum2  
dAns = dNum1 + dNum2  
qAns = qNum1 + qN
```

The following instructions could be u

```
; bAns = bNum1 + bNum2  
mov al, byte [bNum1]  
add al, byte [bNum2]  
mov byte [bAns], al  
  
; wAns = wNum1 + wNum2  
mov ax, word [wNum1]  
add ax, word [wNum2]  
mov word [wAns], ax  
  
; dAns = dNum1 + dNum2  
mov eax, dword [dNum1]  
add eax, dword [dNum2]  
mov dword [dAns], eax  
  
; qAns = qNum1 + qNum2  
mov rax, qword [qNum1]  
add rax, qword [qNum2]  
mov qword [qAns], rax
```

## Chapter 7.0 ◀ Instruction Set Overview

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) can be omitted as the other operand will clearly define the size. It is included for consistency and good programming practice.

In addition to the basic add instruction, there is an increment instruction that will add one to the specified operand. The general form of the increment instruction is as follows:

`inc <operation>`

Where operation is as follows:

`<operation> = <operand>`

The result is exactly the same as using the add instruction (and adding one). When using a memory operand, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) is required to clearly define the size.

For example, assuming the following data declarations:

bNum	db	4
wNum	dw	42
dNum	dd	42000
qNum	dq	

To perform, the basic operations:

```
rax = rax + 1  
bNum = bNum + 1  
wNum = wNum + 1  
dNum = dNum + 1  
qNum = qNum + 1
```

The following instructions could be used:

inc rax
inc byte [bNum]
inc word [wNum]
inc dword [dNum]
inc qword [qNum]

; rax = rax
; bNum = bNum
; wNum = wNum
; dNum = dNum
; qNum = qNum

The addition instruction operates the same on signed and unsigned data. It is the programmer's responsibility to ensure that the data types and sizes are appropriate for the operations being performed.

The integer addition instructions are summarized as follows:

Instruction	Explanation
<code>add &lt;dest&gt;, &lt;src&gt;</code>	Add two operands, ( <code>&lt;dest&gt; + &lt;src&gt;</code> ) and place the result in <code>&lt;dest&gt;</code> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<code>add cx, word [wVar]</code> <code>add rax, 42</code> <code>add dword [dVar],</code> [ ] <code>add qword [qVar],</code> [ ]
<code>inc &lt;operand&gt;</code>	Increment <code>&lt;operand&gt;</code> by 1. <i>Note</i> , <code>&lt;operand&gt;</code> cannot be an immediate.
Examples:	<code>inc word [wVar]</code> <code>inc rax</code> <code>inc dword [dVar],</code> [ ] <code>inc qword [qVar],</code> [ ]

A more complete list of the instructions is located in Appendix E.

### 7.5.1.1 Addition with Carry

The add with carry is a special add instruction that will include a carry from a previous addition operation. This is useful when adding very large numbers, specifically numbers larger than the register size of the machine.

Using a carry in addition is fairly standard. For example, consider the following operation.

$$\begin{array}{r} 17 \\ + 25 \\ \hline \end{array}$$

As you may recall, the least significant digits (7 and 5) are added first. The result of 12 is noted as a 2 with a 1 carry. The most significant digits (1 and 2) are added along with the previous carry (1 in this example) resulting in a 4.

As such, two addition operations are required. Since there is no carry possible with the least significant portion, a regular addition instruction is used. The second addition operation would need to include a possible carry from the previous operation and must be done with an add with carry instruction. Additionally, the add with carry must immediately follow the initial addition operation to ensure that the **rFlag** register is not altered by an unrelated instruction (thus possibly altering the carry bit).

For assembly language programs the Least Significant Quadword (LSQ) is added with the **add** instruction and then immediately the Most Significant Quadword (MSQ) is added with the **adc** which will add the quadwords and include a carry from the previous addition operation.

The general form of the integer add with carry instruction is as follows:

**adc**    <dest>, <s>

Where operation performs the following:

<dest> = <dest> + <src> + <carryB>

Specifically, the source and destination operands along with the carry bit are added and the result is placed in the destination operand (over-writing the previous value). The carry bit is part of the **rFlag** register. The value of the source operand is unchanged. The destination and source operand must be of the same size (both bytes, both words, etc.). The destination operand cannot be an immediate. Both operands, cannot be memory. If a memory to memory addition operation is required, two instructions must be used.

For example, given the following declarations:

<b>dquad1</b>	<b>ddq</b>	0x1A00000000000000	[ ]
<b>dquad2</b>	<b>ddq</b>	0x2C00000000000000	[ ]
<b>dqSum</b>	<b>ddq</b>		[ ]

Each of the variables **dquad1**, **dquad2**, and **dqSum** are 128-bits and thus will exceed the machine 64-bit register size. However, two 64-bit registers can be used for each of the 128-bit values. This requires two move instructions, one for each 64-bit register. For example,

```
mov    rax, qword [dquad1]
mov    rdx, qword [dquad1]
```

The first move to the **rax** register accesses the first 64-bits of the 128-bit variable. The second move to the **rdx** register access the next 64-bits of the 128-bit variable. This is accomplished by using the variable starting address, **dquad1** and adding 8 bytes, thus skipping the first 64-bits (or 8 bytes) and accessing the next 64-bits.

If the LSQ's are added and then the MSQ's are added including any carry, the 128-bit result can be correctly obtained. For example,

```

mov    rax, qword [dquad1]
mov    rdx, qword [dquad1+8]

add    rax, qword [dquad2]
adc    rdx, qword [dquad2+8]

mov    qword [dqSum], rax
mov    qword [dqSum+8], rdx

```

Initially, the LSQ of **dquad1** is placed in **rax** and the MSQ is placed in **rdx**. Then the **add** instruction will add the 64-bit **rax** with the LSQ of **dquad2** and, in this example, provide a carry of 1 with the result in **rax**. Then the **rdx** is added with the MSQ of **dquad2** along with the carry via the **adc** instruction and the result placed in **rdx**.

The integer add with carry instruction is summarized as folld

Instruction	Explanation
<b>adc &lt;dest&gt;, &lt;src&gt;</b>	Add two operands, ( <b>&lt;dest&gt;</b> + <b>&lt;src&gt;</b> ) and any previous carry (stored in the carry bit in the <b>rFlag</b> register) and place the result in <b>&lt;dest&gt;</b> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<b>adc rcx, qword [dvva1]</b> <b>adc rax, 42</b>

A more complete list of the instructions is located in Appendix

## 7.5.2 Subtract

The general form of the integer subtraction instruction is as folld

```
sub    <dest>, <src>
```

## Chapter 7.0 ◀ Instruction Set Overview

Where operation performs the following:

$\text{<dest>} = \text{<dest>} - \text{<s...>}$

Specifically, the source operand is subtracted from the destination operand and the result is placed in the destination operand (over-writing the previous value). The value of the source operand is unchanged. The destination and source operand must be of the same size (both bytes, both words, etc.). The destination operand cannot be an immediate. Both operands, cannot be memory. If a memory to memory subtraction operation is required, two instructions must be used.

For example, assuming the following data declarations:

bNum1	db	
bNum2	db	
bAns	db	
wNum1	dw	1
wNum2	dw	4
wAns	dw	
dNum1	dd	73
dNum2	dd	12
dAns	dd	
qNum1	dq	73000
qNum2	dq	73000
qAns	dd	

To perform, the basic operations:

```
bAns = bNum1 - bNum2  
wAns = wNum1 - wNum2  
dAns = dNum1 - dNum2  
qAns = qNum1 - qNum2
```

The following instructions could be used:

```
; bAns = bNum1 - bNum2  
mov al, byte [bNum1]  
sub al, byte [bNum2]  
mov byte [bAns], al  
  
; wAns = wNum1 - wNum2  
mov ax, word [wNum1]
```

```

sub      ax, word [wNum]
mov      word [wAns],   

;  

; dAns = dNum1 - dNum2
mov      eax, dword [dNum1]
sub      eax, dword [dNum2]
mov      dword [dAns],   

;  

; qAns = qNum1 - qNum2
mov      rax, qword [qNum1]
sub      rax, qword [qNum2]
mov      qword [qAns], 

```

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) can be omitted as the other operand will clearly define the size. It is included for consistency and good programming practices. 

In addition to the basic subtract instruction, there is a decrement instruction that will subtract one from the specified operand. The general form of the decrement instruction is as follows: 

```
dec    <operand>
```

Where operation performs the following: 

$$<\text{operation}> = <\text{operand}> - 1$$

The result is exactly the same as using the subtract instruction (and subtracting one). When using a memory operand, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) is required to clearly define the size. 

For example, assuming the following data declarations: 

bNum	db	
wNum	dw	4 
dNum	dd	42 
qNum	dq	42000 

To perform, the basic operations: 

```

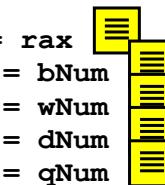
rax = rax - 1
bNum = bNum - 1
wNum = wNum - 1
dNum = dNum - 1
qNum = qNum - 1

```

## Chapter 7.0 ◀ Instruction Set Overview

The following instructions could be used:

```
dec    rax          ; rax = rax
dec    byte [bNum]   ; bNum = bNum
dec    word [wNum]   ; wNum = wNum
dec    dword [dNum]  ; dNum = dNum
dec    qword [qNum]  ; qNum = qNum
```



The subtraction instructions operate the same on signed and unsigned data. It is the programmer's responsibility to ensure that the data types and sizes are appropriate for the operations being performed.

The integer subtraction instructions are summarized as follows:

Instruction	Explanation
<b>sub &lt;dest&gt;, &lt;src&gt;</b>	Subtract two operands, ( <b>&lt;dest&gt;</b> - <b>&lt;src&gt;</b> ) and place the result in <b>&lt;dest&gt;</b> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<b>sub cx, word [wVar]</b> <b>sub rax, 42</b> <b>sub dword [dVar],</b> <b>sub qword [qVar],</b>
<b>dec &lt;operand&gt;</b>	Decrement <b>&lt;operand&gt;</b> by 1. <i>Note</i> , <b>&lt;operand&gt;</b> cannot be an immediate.
Examples:	<b>dec word [wVar]</b> <b>dec rax</b> <b>dec dword [dVar]</b> <b>dec qword [qVar]</b>

A more complete list of the instructions is located in Appendix E.

### 7.5.3 Integer Multiplication

The multiply instruction multiplies two integer operands. Mathematically, there are special rules for handling multiplication of signed values. As such, different instructions are used for unsigned multiplication (**mul**) and signed multiplication (**imul**).

Multiplication typically produces double sized results. That is, multiplying two  $n$ -bit values produces a  $2n$ -bit result. Multiplying two 8-bit numbers will produce a 16-bit result. Similarly, multiplication of two 16-bit numbers will produce a 32-bit result, multiplication of two 32-bit numbers will produce a 64-bit result, and multiplication of two 64-bit numbers will produce a 128-bit result.

There are many variants for the multiply instruction. For the signed multiply, some forms will truncate the result to the size of the original operands. It is the programmer's responsibility to ensure that the values used will work for the specific instructions selected.

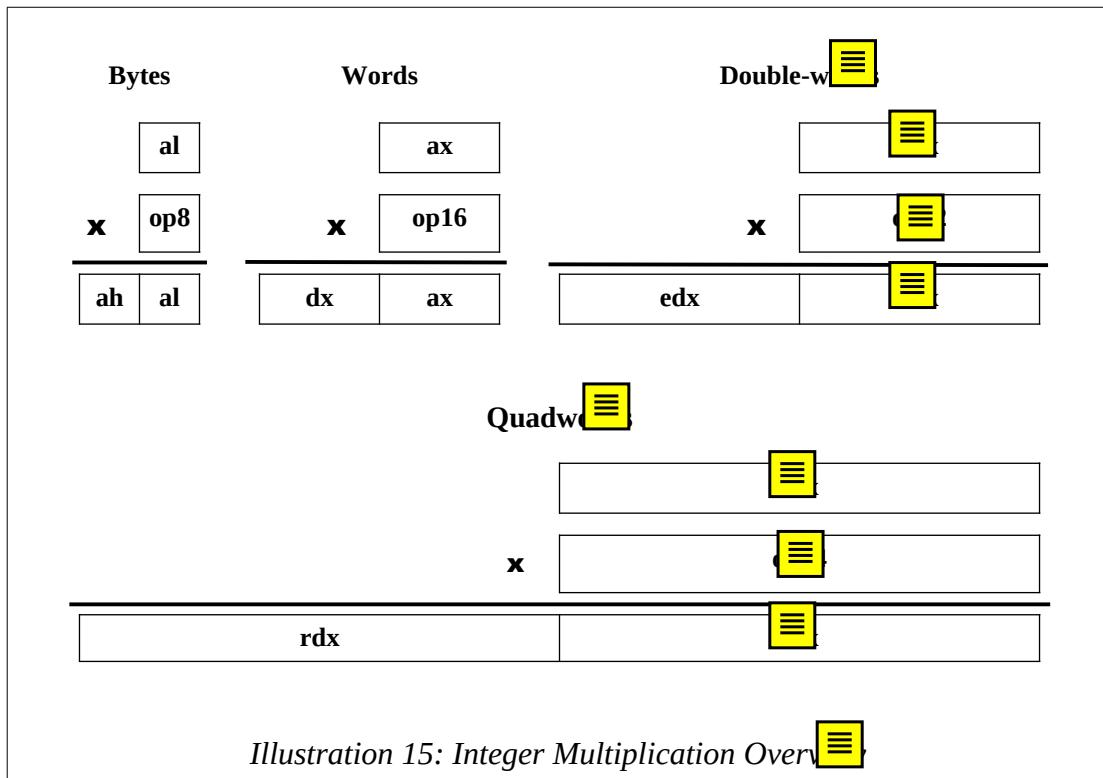
#### 7.5.3.1 Unsigned Multiplication

The general form of the unsigned multiplication is as follows:

**mul**      <s>

Where the source operand must be a register or memory location. An immediate operand is not allowed.

For the single operand multiply instruction, the **A** register (**al/ax/eax/rax**) must be used for one of the operands (**al** for 8-bits, **ax** for 16-bits, **eax** for 32-bits, and **rax** for 64-bit). The other operand can be a memory location or register, but not an immediate. Additionally, the result will be placed in the **A** and possibly **D** registers, based on the sizes being multiplied. The following table shows the various options for the byte, word, double-word, and quadword unsigned multiplications.



As shown in the chart, for most cases the integer multiply uses a combination of the **A** and **D** registers. This can be very confusing.

For example, when multiplying a **rax** (64-bits) times a quadword operand (64-bits), the multiplication instruction provides a double quadword result (128-bit). This can be useful and important when dealing with very large numbers. Since the 64-bit architecture only has 64-bit registers, the 128-bit result is, and must be, placed in two different quadword (64-bit) registers, **rdx** for the upper-order result and **rax** for the lower-order result, which is typically written as **rdx:rax** (by convention).

However, this use of two registers is applied to smaller sizes as well. For example, the result of multiplying **ax** (16-bits) times a word operand (also 16-bits) provides a double-word (32-bit) result. However, the result is not placed in **eax** (which might be easier), it is placed in two registers, **dx** for the upper-order result (16-bits) and **ax** for the lower-order result (16-bits), typically written as **dx:ax** (by convention). Since the double-word (32-bit) result is in two different registers, two moves may be required to save the res

This pairing of registers, even when not required, is due to legacy support for previous earlier versions of the architecture. While this helps ensure backwards compatibility, it can be quite confusing.

For example, assuming the following data declarati

bNumA	db	
bNumB	db	
wAns	dw	
wAns1	dw	
wNumA	dw	
wNumB	dw	
dAns2	dd	
dNumA	dd	
dNumB	dd	
qAns3	dq	
qNumA	dq	
qNumB	dq	
dqAns4	ddq	

To perform, the basic operations

```
wAns = bNumA^2 ; bNumA squared
bAns1 = bNumA * bNumB
wAns1 = bNumA * bNumB
wAns2 = wNumA * wNumB
dAns2 = wNumA * wNumB

dAns3 = dNumA * dNumB
qAns3 = dNumA * dNumB

qAns4 = qNumA * qNumB
dqAns4 = qNumA * qNumB
```

The following instructions could be u

```
; wAns = bNumA^2 or bNumA squared
mov    al, byte [bNumA]
mul    al, al ; result in al
mov    word [wAns], al
```

```

; wAns1 = bNumA * bNumB
mov    al, byte [bNumA]           ; result in al
mul    byte [bNumB]
mov    word [wAns1], al

; dAns2 = wNumA * wNumB
mov    ax, word [wNumA]
mul    word [wNumB]
mov    word [dAns2], ax           ; result in dx
mov    word [dAns2+2], al

; qAns3 = dNumA * dNumB
mov    eax, dword [dNumA]
mul    dword [dNumB]
mov    dword [qAns3], eax         ; result in edx:eax
mov    dword [qAns3+4], al

; dqAns4 = qNumA * qNumB
mov    rax, qword [qNumA]
mul    qword [qNumB]
mov    qword [dqAns4], rax         ; result in rdx:rax
mov    qword [dqAns4+8], al

```

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) is required to clearly define the size.

The integer unsigned multiplication instruction is summarized as follows:

Instruction	Explanation
<b>mul &lt;s&gt;</b> <b>mul &lt;op8&gt;</b> <b>mul &lt;op16&gt;</b> <b>mul &lt;op32&gt;</b> <b>mul &lt;op64&gt;</b>	Multiply A register ( <b>al</b> , <b>ax</b> , <b>eax</b> , or <b>rax</b> ) times the <b>&lt;src&gt;</b> operand. Byte: <b>ax</b> = <b>al</b> * <b>&lt;src&gt;</b> Word: <b>dx:ax</b> = <b>ax</b> * <b>&lt;src&gt;</b> Double: <b>edx:eax</b> = <b>eax</b> * <b>&lt;src&gt;</b> Quad: <b>rdx:rax</b> = <b>rax</b> * <b>&lt;src&gt;</b> <i>Note, &lt;src&gt; operand cannot be an immediate value.</i>
Examples:	<b>mul word [wVar]</b> <b>mul al</b> <b>mul dword [dVar]</b> <b>mul qword [qVar]</b>

A more complete list of the instructions is located in Appendix 

### 7.5.3.2 Signed Multiplication

The signed multiplication allows a wider range of operands and operand sizes. The general forms of the signed multiplication are as follows: 

```
imul    <source>  
imul    <dest>, <src/immediate>  
imul    <dest>, <src>, <immediate>
```

In all cases, the destination operand must be a register. For the multiple operand multiply instruction, byte operands are not supported. 

When using a **single** operand multiply instruction, the **imul** is the same layout as the **mul** (as previously presented). However, the operands are interpreted only as signed 

When two operands are used, the destination operand and the source operand are multiplied and the result placed in the destination operand (over-writing the previous value). 

Specifically, the action performed is 

$$<\text{dest}> = <\text{dest}> * <\text{src}/\text{imm}>$$

For two operands, the **<src/imm>** operand may be a register, memory location, or immediate value. The size of the immediate value is limited to the size of the source operand, up to a double-word size (32-bit), even for quadword (64-bit) multiplications. The final result is truncated to the size of the destination operand. A byte sized destination operand is not supported. 

When three operands are used, two operands are multiplied and the result placed in the destination operand. Specifically, the action performed is: 

$$<\text{dest}> = <\text{src}> * <\text{immediate}>$$

For three operands, the **<src>** operand must be a register or memory location, but not an immediate. The **<imm>** operand must be an immediate value. The size of the immediate value is limited to the size of the source operand, up to a double-word size (32-bit), even for quadword multiplications. The final result is truncated to the size of the destination operand. A byte sized destination operand is not supported. 

It should be noted that when the multiply instruction provides a larger type, the original type may be used. For this to work, the values multiplied must fit into the smaller size which limits the range of the data. For example, when two double-words are multiplied and a quadword result is provided, the least significant double-word (of the quadw 



will contain the answer if the values are sufficiently small which is often the case. This is typically done in high-level languages when an **int** (32-bit integer) variable is multiplied by another **int** variable and assigned to an **int** variable.



For example, assuming the following data declarations:

wNumA	dw	1	
wNumB	dw	2	
wAns1	dw		
wAns2	dw		
dNumA	dd	42	
dNumB	dd	13	
dAns1	dd		
dAns2	dd		
qNumA	dq	120	
qNumB	dq	230	
qAns1	dq		
qAns2	dq		

To perform the basic operations:

```
wAns1 = wNumA * -13
wAns2 = wNumA * wN
```



```
dAns1 = dNumA * 113
dAns2 = dNumA * dN
```



```
qAns1 = qNumA * 7096
qAns2 = qNumA * qN
```



The following instructions could be used:

```
; wAns1 = wNumA * -13
mov ax, word [wNu]
imul ax, -13
mov word [wAns1],
```



; result in



```
; wAns2 = wNumA * wNumB
mov ax, word [wNu]
imul ax, word [wNu]
mov word [wAns2],
```



; result in



```

; dAns1 = dNumA * 113
mov    eax, dword [dNu1]
imul   eax, 113
mov    dword [dAns1], 2 ; result in 3

; dAns2 = dNumA * dNumB
mov    eax, dword [dNu1]
imul   eax, dword [dNu2]
mov    dword [dAns2], 3 ; result in 4

; qAns1 = qNumA * 7096
mov    rax, qword [qNu1]
imul   rax, 7096
mov    qword [qAns1], 2 ; result in 3

; qAns2 = qNumA * qNumB
mov    rax, qword [qNu1]
imul   rax, qword [qNu2]
mov    qword [qAns2], 3 ; result in 4

```

Another way to perform the multiplication:

```
qAns1 = qNumA * 7 1
```

Would be as follows:

```

; qAns1 = qNumA * 7096
mov    rcx, qword [qNu1]
imul   rbx, rcx, 7096
mov    qword [qAns1], 2 ; result in 3

```

This example shows the three-operand multiply instruction using different registers:

In these examples, the multiplication result is truncated to the size of the destination operand. For a full-sized result, the single operand instruction should be used (as fully described in the section regarding unsigned multiplication).

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) may not be required to clearly define the size.

The integer signed multiplication instruction is summarized as follows:

Instruction	Explanation
<pre>imul &lt;src&gt; imul &lt;dest&gt;, &lt;src/imm32&gt; imul &lt;dest&gt;, &lt;src&gt;, &lt;imm&gt;</pre> <pre>imul &lt;op8&gt; imul &lt;op16&gt; imul &lt;op32&gt; imul &lt;op64&gt; imul &lt;reg16&gt;, &lt;op16/imm&gt; imul &lt;reg32&gt;, &lt;op32/imm&gt; imul &lt;reg64&gt;, &lt;op64/imm&gt; imul &lt;reg16&gt;, &lt;op16&gt;, &lt;imm&gt; imul &lt;reg32&gt;, &lt;op32&gt;, &lt;imm&gt; imul &lt;reg64&gt;, &lt;op64&gt;, &lt;imm&gt;</pre>	<p>Signed multiply instruction</p> <p>For single operand:</p> <ul style="list-style-type: none"> <li>Byte: <math>ax = al * &lt;src&gt;</math></li> <li>Word: <math>dx:ax = ax * &lt;src&gt;</math></li> <li>Double: <math>edx:eax = eax * &lt;src&gt;</math></li> <li>Quad: <math>rdx:rax = rax * &lt;src&gt;</math></li> </ul> <p>Note, &lt;src&gt; operand cannot be an immediate.</p> <p>For two operands:</p> <ul style="list-style-type: none"> <li><math>&lt;\text{reg16}&gt; = &lt;\text{reg16}&gt; * &lt;\text{op16/imm}&gt;</math></li> <li><math>&lt;\text{reg32}&gt; = &lt;\text{reg32}&gt; * &lt;\text{op32/imm}&gt;</math></li> <li><math>&lt;\text{reg64}&gt; = &lt;\text{reg64}&gt; * &lt;\text{op64/imm}&gt;</math></li> </ul> <p>For three operands:</p> <ul style="list-style-type: none"> <li><math>&lt;\text{reg16}&gt; = &lt;\text{op16}&gt; * &lt;\text{imm}&gt;</math></li> <li><math>&lt;\text{reg32}&gt; = &lt;\text{op32}&gt; * &lt;\text{imm}&gt;</math></li> <li><math>&lt;\text{reg64}&gt; = &lt;\text{op64}&gt; * &lt;\text{imm}&gt;</math></li> </ul>
Examples:	<pre>imul ax, 17 imul al imul ebx, dword [dVar] imul rbx, dword [dVar], 791 imul rcx, qword [qVar] imul qword [qVar]</pre>

A more complete list of the instructions is located in Appendix E.

#### 7.5.4 Integer Division

The division instruction divides two integer operands. Mathematically, there are special rules for handling division of signed values. As such, different instructions are used for unsigned division (**div**) and signed division (**idiv**).

Recall that  $\frac{\text{dividend}}{\text{divisor}} = \text{quotient}$

Division requires that the dividend must be a larger size than the divisor. In order to divide by an 8-bit divisor, the dividend must be 16-bits (i.e., the larger size). Similarly, a 16-bit divisor requires a 32-bit dividend. And, a 32-bit divisor requires a 64-bit dividend.

Like the multiplication, for most cases the integer division uses a combination of the **A** and **D** registers. This pairing of registers is due to legacy support for previous earlier versions of the architecture. While this helps ensure backwards compatibility, it can be quite confusing.

Further, the **A**, and possibly the **D** register, must be used in combination for the dividend.

- Byte Divide:
- Word Divide:
- Double-word divide:
- Quadword Divide:

**ax** for 16-bit  
**dx:ax** for 32-bit  
**edx:eax** for 64-bit  
**rdx:rax** for 128-bit

Setting the dividend (top operand) correctly is a key source of problems. For the word, double-word, and quadword division operations, the dividend requires both the **D** register (for the upper-order portion) and **A** (for the lower-order portion).

Setting these correctly depends on the data type. If a previous multiplication was performed, the **D** and **A** registers may already be set correctly. Otherwise, a data item may need to be converted from its current size to a larger size with the upper-order portion being placed in the **D** register. For unsigned data, the upper portion will always be zero. For signed data, the existing data must be sign extended as noted in a previous section, *Signed Conversions*.

The divisor can be a memory location or register, but not an immediate. Additionally, the result will be placed in the **A** register (**al/ax/eax/rax**) and the remainder in either the **ah**, **dx**, **edx**, or **rdx** register. Refer to the *Integer Division Overview* table to see the layout more clearly.

The use of a larger size operand for the dividend matches the single operand multiplication. For simple divisions, an appropriate conversion may be required in order to ensure the dividend is set correctly. For unsigned divisions, the upper-order part of the dividend can be set to zero. For signed divisions, the upper-order part of the dividend can be set with an applicable conversion instruction.

As always, division by zero will crash the program and damage the space-time continuum. So, try not to divide by zero.



The following tables provide an overview of the divide instruction for bytes, words, double-words, and quadwords.

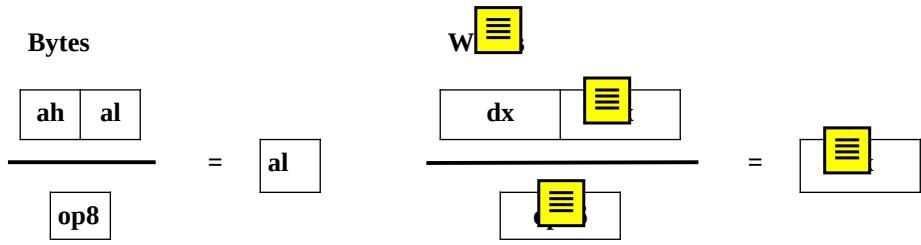
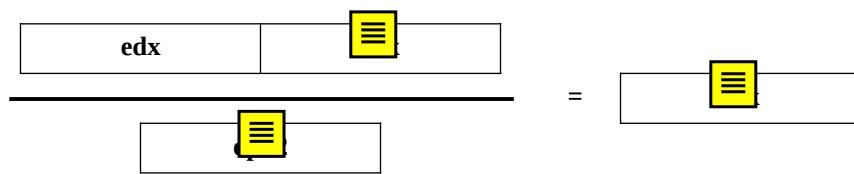
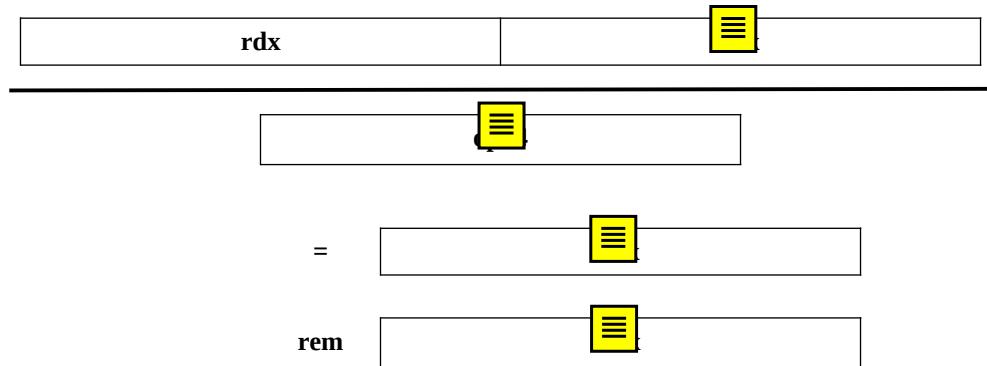
**Bytes****Double-w****Quadw**

Illustration 16: Integer Division Overview



The signed and unsigned division instructions operate in the same manner. However, the range values that can be divided is different. The programmer is responsible for ensuring that the values being divided are appropriate for the operand sizes being used.

The general forms of the unsigned and signed division are as follows:

**div**      <src>

; unsigned division  
; signed division

The source operand and destination operands (A and D registers) are described in the preceding table.

For example, assuming the following data declarations:

bNumA	db	
bNumB	db	
bNumC	db	
bAns1	db	
bAns2	db	
bRem2	db	
bAns3	db	
wNumA	dw	4
wNumB	dw	1
wNumC	dw	
wAns1	dw	
wAns2	dw	
wRem2	dw	
wAns3	dw	
dNumA	dd	42
dNumB	dd	-3
dNumC	dd	
dAns1	dd	
dAns2	dd	
dRem2	dd	
dAns3	dd	
qNumA	dq	730
qNumB	dq	-12
qNumC	dq	1
qAns1	dq	
qAns2	dq	
qRem2	dq	
qAns3	dq	

To perform, the basic operations

```
bAns1 = bNumA / 3
bAns2 = bNumA / bNumB
bRem2 = bNumA % bNumB
bAns3 = (bNumA * bNumC) / bNumB

wAns1 = wNumA / 5
wAns2 = wNumA / wNumB
wRem2 = wNumA % wNumB
wAns3 = (wNumA * wNumC) / wNumB

dAns = dNumA / 7
dAns3 = dNumA * dNumB
dRem1 = dNumA % dNumB
dAns3 = (dNumA * dNumC) / dNumB

qAns = qNumA / 9
qAns4 = qNumA * qNumB
qRem1 = qNumA % qNumB
qAns3 = (qNumA * qNumC) / qNumB
```

```
; unsigned int
; unsigned int
; % is modu
; unsigned int
; unsigned int
; % is modu
; unsigned int
; unsigned int
; % is modu
; unsigned int
; signed int
; signed int
; % is modu
; signed int
; signed int
; % is modu
; signed int
```

The following instructions could be used

```
; -----
; example byte operations, unsigned int

; bAns1 = bNumA / 3 (unsigned int)
mov al, byte [bNumA]
mov ah, 0
mov bl, 1
div bl
mov byte [bAns1], al ; al = ax

; bAns2 = bNumA / bNumB (unsigned int)
mov ax, 0
mov al, byte [bNumA]
div byte [bNumB] ; al = ax / bNumB
mov byte [bAns2], al
mov byte [bRem2], ah ; ah = ax % bNumB

; bAns3 = (bNumA * bNumC) / bNumB (unsigned int)
mov al, byte [bNumA]
```

```

mul    byte [bNumC] ; result in al
div    byte [bNumB] ; al = ax / bN
mov    byte [bAns3], al

; -----
; example word operations, unsigned

; wAns1 = wNumA / 5 (unsigned)
mov    ax, word [wNumA]
mov    dx, word [wNumB]
mov    bx, word [wAns1], al
div    bx
; ax = dx:ax

; wAns2 = wNumA / wNumB (unsigned)
mov    dx, 0
mov    ax, word [wNumA]
div    word [wNumB]
mov    word [wAns2], al
mov    word [wRem2], al
; ax = dx:ax / wN

; wAns3 = (wNumA * wNumC) / wNumB (unsigned)
mov    ax, word [wNumA]
mul    word [wNumC]
div    word [wNumB]
; result in dx
; ax = dx:ax / wN
mov    word [wAns3], al

; -----
; example double-word operations, signed

; dAns1 = dNumA / 7 (signed)
mov    eax, dword [dNumA]
cdq
mov    ebx, dword [dAns1], al
idiv  ebx
; eax → edx:eax
; eax = edx:eax

; dAns2 = dNumA / dNumB (signed)
mov    eax, dword [dNumA]
cdq
idiv  dword [dNumB]
mov    dword [dAns2], al
mov    dword [dRem2], edx
; eax → edx:eax
; eax = edx:eax/dN
; edx = edx:eax%dN

```

```

; dAns3 = (dNumA * dNumC) / dNumB (signed)
mov    eax, dword [dNumA]
imul   dword [dNumC]
idiv   dword [dNumB]
mov    dword [dAns3],  
; result in edx:  
; eax = edx:eax/dNb   

;  

; example quadword operations, signed
; qAns1 = qNumA / 9 (signed)
mov    rax, qword [qNumA] 
cqq
mov    rbx 
idiv   rbx
mov    qword [qAns1],  
; rax → rdx:   

; eax = edx:eax   

; qAns2 = qNumA / qNumB (signed)
mov    rax, qword [qNumA] 
cqq
idiv   qword [qNumB] 
mov    qword [qAns2],  
mov    qword [qRem2], rdx 
; rax → rdx:   

; rax = rdx:rax/qNb   

; rdx = rdx:rax%qNb   

; qAns3 = (qNumA * qNumC) / qNumB (signed)
mov    rax, qword [qNumA]
imul   qword [qNumC]
idiv   qword [qNumB]
mov    qword [qAns3],  
; result in rdx:  
; rax = rdx:rax/qNb 

```

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) is required to clearly define the size. 

The integer division instructions are summarized as follows:

Instruction	Explanation
<code>div &lt;src&gt;</code> <code>div &lt;op8&gt;</code> <code>div &lt;op16&gt;</code> <code>div &lt;op32&gt;</code> <code>div &lt;op64&gt;</code>	Unsigned divide A/D register (ax, dx:ax, dx:eax, or rdx:rax) by the <src> operand. Byte: <b>al</b> = ax / <src>, rem in ah Word: <b>ax</b> = dx:ax / <src>, rem in dx Double: <b>eax</b> = eax / <src>, rem in edx Quad: <b>rax</b> = rax / <src>, rem in rdx Note, <src> operand cannot be an immediate.
Examples:	<code>div word [wVvar]</code> <code>div bl</code> <code>div dword [dVvar]</code> <code>div qword [qVvar]</code>
<code>idiv &lt;src&gt;</code> <code>idiv &lt;op8&gt;</code> <code>idiv &lt;op16&gt;</code> <code>idiv &lt;op32&gt;</code> <code>idiv &lt;op64&gt;</code>	Signed divide A/D register (ax, dx:ax, dx:eax, or rdx:rax) by the <src> operand. Byte: <b>al</b> = ax / <src>, rem in ah Word: <b>ax</b> = dx:ax / <src>, rem in dx Double: <b>eax</b> = eax / <src>, rem in edx Quad: <b>rax</b> = rax / <src>, rem in rdx Note, <src> operand cannot be an immediate.
Examples:	<code>idiv word [wVvar]</code> <code>idiv bl</code> <code>idiv dword [dVvar]</code> <code>idiv qword [qVvar]</code>

A more complete list of the instructions is located in Appendix E.

## 7.6 Logical Instructions

This section summarizes some of the more common logical instructions that may be useful when programming.



### 7.6.1 Logical Operations



As you should recall, below are the truth tables for the basic logical operations:

<b>and</b> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td></tr> </table>	0	1	0	1	0	0	1	1	0	0	0	1	<b>or</b> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>1</td></tr> </table>	0	1	0	1	0	0	1	1	0	1	1	1	<b>xor</b> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>0</td><td>1</td><td></td><td></td></tr> <tr><td>0</td><td>0</td><td></td><td></td></tr> <tr><td>0</td><td>1</td><td></td><td></td></tr> </table>	0	1			0	0			0	1		
0	1	0	1																																			
0	0	1	1																																			
0	0	0	1																																			
0	1	0	1																																			
0	0	1	1																																			
0	1	1	1																																			
0	1																																					
0	0																																					
0	1																																					
<i>Illustration 17: Logical Operations</i>																																						

The logical instructions are summarized as follows:



Instruction	Explanation
<b>and    &lt;dest&gt;, &lt;src&gt;</b>	Perform logical AND operation on two operands, (<dest> and <src>) and place the result in <dest> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<pre>and    ax, bx and    rcx, rdx and    eax, dword [dN] and    qword [qNum],</pre>
<b>or    &lt;dest&gt;, &lt;src&gt;</b>	Perform logical OR operation on two operands, (<dest>    <src>) and place the result in <dest> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<pre>or    ax, bx or    rcx, rdx or    eax, dword [dN] or    qword [qNum],</pre>



Instruction	Explanation
<code>xor &lt;dest&gt;, &lt;src&gt;</code>	Perform logical XOR operation on two operands, ( <code>&lt;dest&gt; ^ &lt;src&gt;</code> ) and place the result in <code>&lt;dest&gt;</code> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand cannot be an immediate.
Examples:	<code>xor ax, bx</code> <code>xor rcx, rdx</code> <code>xor eax, dword [dN]</code> <code>xor qword [qNum],</code>
<code>not &lt;op&gt;</code>	Perform a logical not operation (one's complement on the operand 1's→0's and 0's→1's). <i>Note</i> , operand cannot be an immediate.
Examples:	<code>not bx</code> <code>not rdx</code> <code>not dword [dN]</code> <code>not qword [qN]</code>

The `&` refers to the logical AND operation, the `||` refers to the logical OR operation, and the `^` refers to the logical XOR operation as per C/C++ conventions. The `~` refers to the logical NOT operation.

A more complete list of the instructions is located in Appendix E.

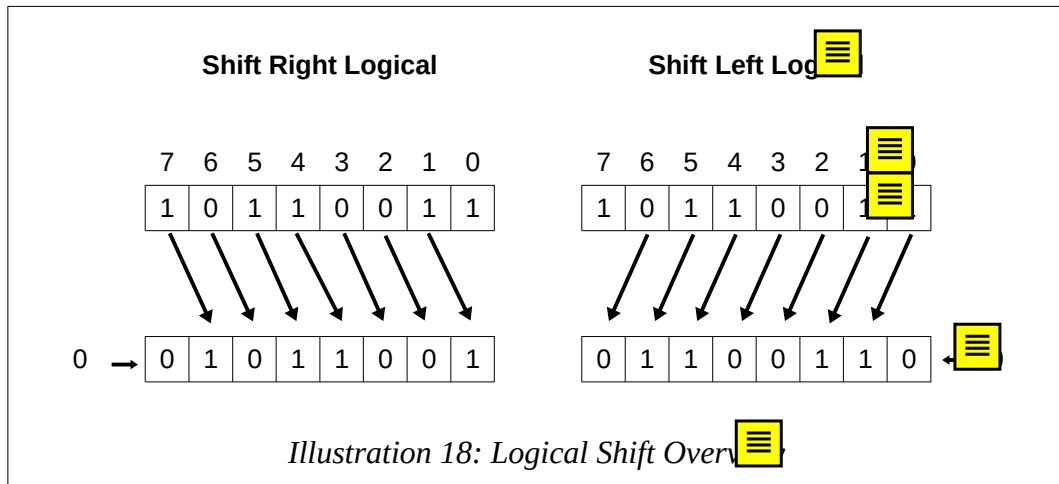
## 7.6.2 Shift Operations

The shift operation shifts bits within an operand, either left or right. Two typical reasons for shifting bits include isolating a subset of the bits within an operand for some specific purpose or possibly for performing multiplication or division by powers of two. All bits are shifted one position. The bit that is shifted outside the operand is lost and a 0-bit added at the other side.

### 7.6.2.1 Logical Shift

The logical shift is a bitwise operation that shifts all the bits of its source register by the specified number of bits and places the result into the destination register. The bits can be shifted left or right as needed. Every bit in the source operand is moved the specified number of bit positions and the newly vacant bit positions are filled in with zeros.

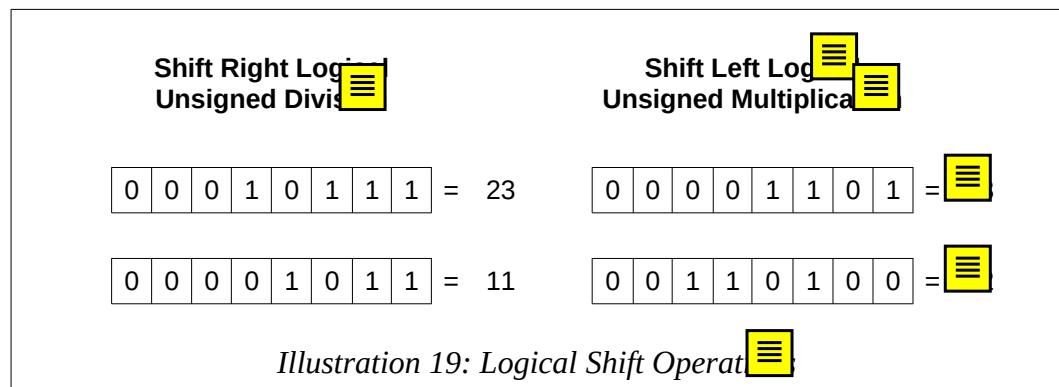
The following diagram shows how the right and left shift operations work for byte sized operands.



The logical shift treats the operand as a sequence of bits rather than as a number.

The shift instructions may be used to perform unsigned integer multiplication and division operations for powers of 2. Powers of two would be 2, 4, 8, etc. up to the limit of the operand size (32-bits for register operands).

In the examples below, 23 is divided by 2 by performing a shift right logical one bit. The resulting 11 is shown in binary. Next, 13 is multiplied by 4 by performing a shift left logical two bits. The resulting 52 is shown in binary.



As can be seen in the examples, a 0 was entered in the newly vacated bit locations on either the right or left (depending on the operation).

The logical shift instructions are summarized as follows:

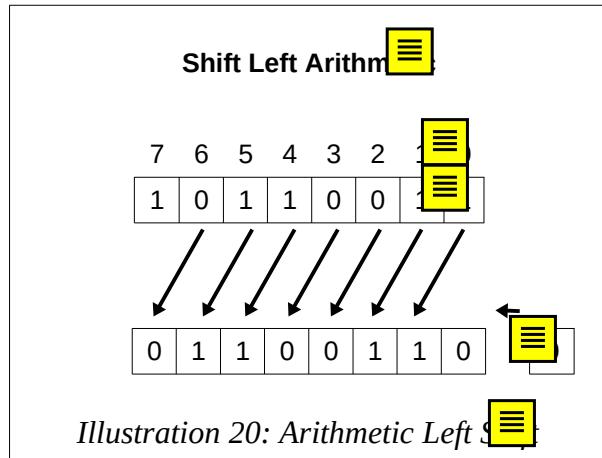
Instruction	Explanation
<code>shl &lt;dest&gt;, &lt;imm&gt;</code> <code>shl &lt;dest&gt;, cl</code>	Perform logical shift left operation on destination operand. Zero fills from right (as needed). The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:  <code>shl ax, 8</code> <code>shl rcx, 32</code> <code>shl eax, cl</code> <code>shl qword [qNum], cl</code>	
<code>shr &lt;dest&gt;, &lt;imm&gt;</code> <code>shr &lt;dest&gt;, cl</code>	Perform logical shift right operation on destination operand. Zero fills from left (as needed). The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:  <code>shr ax, 8</code> <code>shr rcx, 32</code> <code>shr eax, cl</code> <code>shr qword [qNum], cl</code>	

A more complete list of the instructions is located in Appendix E.

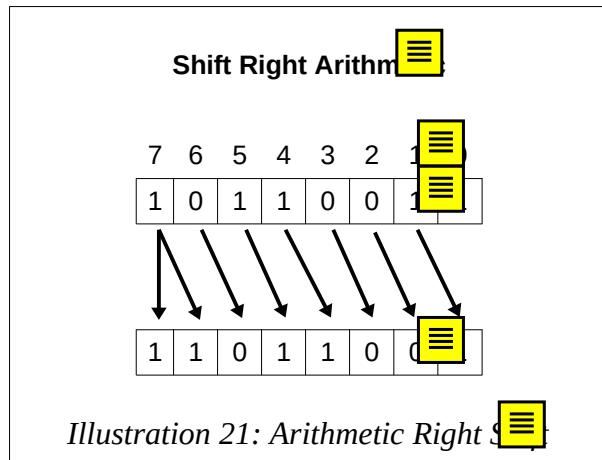
### 7.6.2.2 Arithmetic Shifts

The arithmetic shift right is also a bitwise operation that shifts all the bits of its source register by the specified number of bits and places the result into the destination register. Every bit in the source operand is moved the specified number of bit positions, and the newly vacant bit positions are filled in. For an arithmetic left shift, the original leftmost bit (the sign bit) is replicated to fill in all the vacant positions. This is referred to as sign extension.

The following diagrams show how the shift left and shift right arithmetic operations works for a byte sized operand.



The arithmetic left shift moves bits the number of specified places to the left and zero fills the from the least significant bit position (left). The leading sign bit is not preserved. The arithmetic left shift can be useful to perform an efficient multiplication by a power of two. If the resulting value does not fit an overflow is generated.



The arithmetic right shift moves bits the number of specified places to the right and treats the operand as a signed number which extends the sign (negative in this example).



The arithmetic shift rounds always rounds down (towards negative infinity) and the standard divide instruction truncates (rounds toward 0). As such, the arithmetic shift is not typically used to replace the signed divide instruction.



The arithmetic shift instructions are summarized as follows:



Instruction	Explanation
<code>sal &lt;dest&gt;, &lt;imm&gt;</code> <code>sal &lt;dest&gt;, cl</code>	Perform arithmetic shift left operation on destination operand. Zero fills from right (as needed). The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:	<code>sal ax, 8</code> <code>sal rcx, 32</code> <code>sal eax, cl</code> <code>sal qword [qNum],</code>
<code>sar &lt;dest&gt;, &lt;imm&gt;</code> <code>sar &lt;dest&gt;, cl</code>	Perform arithmetic shift right operation on destination operand. Sign fills from left (as needed). The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:	<code>sar ax, 8</code> <code>sar rcx, 32</code> <code>sar eax, cl</code> <code>sar qword [qNum],</code>

A more complete list of the instructions is located in Appendix E.

### 7.6.3 Rotate Operations

The rotate operation shifts bits within an operand, either left or right, with the bit that is shifted outside the operand is rotated around and placed at the other end.

For example, if a byte operand,  $10010110_2$ , is rotated to the right 1 place, the result would be  $01001011_2$ . If a byte operand,  $10010110_2$ , is rotated to the left 1 place, the result would be  $00101101_2$ .

The logical shift instructions are summarized as follows.

Instruction	Explanation
<code>rol &lt;dest&gt;, &lt;imm&gt;</code> <code>rol &lt;dest&gt;, cl</code>	Perform rotate left operation on destination operand. The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:	<code>rol ax, 8</code> <code>rol rcx, 32</code> <code>rol eax, cl</code> <code>rol qword [qNum],</code>
<code>ror &lt;dest&gt;, &lt;imm&gt;</code> <code>ror &lt;dest&gt;, cl</code>	Perform rotate right operation on destination operand. The <imm> or the value in cl register must be between 1 and 64. Note, destination operand cannot be an immediate.
Examples:	<code>ror ax, 8</code> <code>ror rcx, 32</code> <code>ror eax, cl</code> <code>ror qword [qNum],</code>

A more complete list of the instructions is located in Appendix E.

## 7.7 Control Instructions

Program control refers to basic programming structures such as IF statements and looping.

All of the high-level language control structures must be performed with the limited assembly language control structures. For example, an IF-THEN-ELSE statement does not exist at the assembly language level. Assembly language provides an unconditional branch (or jump) and a conditional branch or an IF statement that will jump to a target label or not jump.

The control instructions refer to unconditional and conditional jumping. Jumping is required for basic conditional statements (i.e., IF statements) and looping.

## 7.7.1 Labels

A program label is the target, or a location to jump to, for control statements. For example, the start of a loop might be marked with a label such as “loopStart”. The code may be re-executed by jumping to the label.

Generally, a label starts with a letter, followed by letters, numbers, or symbols (limited to “\_”), terminated with a colon (“:”). It is possible to start labels with non-letter characters (i.e., digits, “\_”, “\$”, “#”, “@”, “~” or “?”). However, these typically convey special meaning and, in general, should not be used by programmers. Labels in yasm are case sensitive.

For example,

```
loopStart:  
last:
```

are valid labels. Program labels may be defined only once.

The following sections describe how labels are used.

## 7.7.2 Unconditional Control Instructions

The unconditional instruction provides an unconditional jump to a specific location in the program denoted with a program label. The target label must be defined exactly once and accessible and within scope from the originating jump instruction.

The unconditional jump instruction is summarized as follows:

Instruction	Explanation
<code>jmp &lt;label&gt;</code>	Jump to specified label. <i>Note, label must be defined exactly once.</i>
Examples:	<code>jmp startLoop</code> <code>jmp ifDone</code> <code>jmp last</code>

A more complete list of the instructions is located in Appendix C.

## 7.7.3 Conditional Control Instructions

The conditional control instructions provide a conditional jump based on a comparison. This provides the functionality of a basic IF statement.

Two steps are required for a comparison; the compare instruction and the conditional jump instruction. The conditional jump instruction will jump or not jump to a label.

provided label based on the result of the previous comparison operation. The compare instruction will compare two operands and store the results of the comparison in the **rFlag** registers. The conditional jump instruction will act (jump or not jump) based on the contents of the **rFlag** register. This requires that the compare instruction is immediately followed by the conditional jump instruction. If other instructions are placed between the compare and conditional jump, the **rFlag** register will be altered and the conditional jump may not reflect the correct condition.

The general form of the compare instruction:

```
cmp    <op1>, <op2>
```

Where **<op1>** and **<op2>** are not changed and must be of the same size. Either, but not both, may be a memory operand. The **<op1>** operand cannot be an immediate, but the **<op2>** operand may be an immediate value.

The conditional control instructions include the jump equal (**je**) and jump not equal (**jne**) which work the same for both signed and unsigned data.

The signed conditional control instructions include the basic set of comparison operations; jump less than (**jl**), jump less than or equal (**jle**), jump greater than (**jg**), and jump greater than or equal (**jge**).

The unsigned conditional control instructions include the basic set of comparison operations; jump below than (**jb**), jump below or equal (**jbe**), jump above than (**ja**), and jump above or equal (**jae**).

The general form of the signed conditional instructions along with an explanatory comment are as follows:

<b>je</b>	<b>&lt;label&gt;</b>	<i>; if &lt;op1&gt; == &lt;op2&gt;</i>
<b>jne</b>	<b>&lt;label&gt;</b>	<i>; if &lt;op1&gt; != &lt;op2&gt;</i>
<b>jl</b>	<b>&lt;label&gt;</b>	<i>; signed, if &lt;op1&gt; &lt; &lt;op2&gt;</i>
<b>jle</b>	<b>&lt;label&gt;</b>	<i>; signed, if &lt;op1&gt; &lt;= &lt;op2&gt;</i>
<b>jg</b>	<b>&lt;label&gt;</b>	<i>; signed, if &lt;op1&gt; &gt; &lt;op2&gt;</i>
<b>jge</b>	<b>&lt;label&gt;</b>	<i>; signed; if &lt;op1&gt; &gt;= &lt;op2&gt;</i>
<b>jb</b>	<b>&lt;label&gt;</b>	<i>; unsigned, if &lt;op1&gt; &lt; &lt;op2&gt;</i>
<b>jbe</b>	<b>&lt;label&gt;</b>	<i>; unsigned, if &lt;op1&gt; &lt;= &lt;op2&gt;</i>
<b>ja</b>	<b>&lt;label&gt;</b>	<i>; unsigned, if &lt;op1&gt; &gt; &lt;op2&gt;</i>
<b>jae</b>	<b>&lt;label&gt;</b>	<i>; unsigned, if &lt;op1&gt; &gt;= &lt;op2&gt;</i>

For example, given the following pseudo-code for signed double-words:

```
if (currNum > myMax)
    myMax = currN
```

And, assuming the following data declarations:

currNum	dq	
myMax	dq	

Assuming that the values are updating appropriately within the program (not shown), the following instructions could be used:

```
mov    rax, qword [currN]
cmp    rax, qword [myMax]
jle    notNewMax
mov    qword [myMax], rax
notNewM
```

; if currNum <= myMax  
; skip set new

Note that the logic for the IF statement has been reversed. The compare and conditional jump provide functionality for jump or not jump. As such, if the condition from the original IF statement is false, the code must not be executed. Thus, when false, in order to skip the execution, the conditional jump will jump to the target label immediately following the code to be skipped (not executed). While there is only one line in this example, there can be many lines of code.

A more complex example might be as follows:

```
if (x != 0)
    ans = x / y;
    errFlg = FALSE
} else
    ans = 0;
    errFlg = TRUE
```

This basic compare and conditional jump do not provide a typical IF-ELSE structure. It must be created. Assuming the *x* and *y* variables are signed double-words that will be set during the program execution, and the following declarations:

TRUE	equ	
FALSE	equ	
x	dd	
y	dd	
ans	dd	
errFlg	db	

The following code could be used to implement the above IF-ELSE statement:

```

    cmp      dword [x], 0          ; if statement
    je       doE[...]
    mov      eax, dword [...]
    cdq
    idiv    dword [...]
    mov      dword [ans], [...]
    mov      byte [errFlg], FA[...]
    jmp     skpE[...]
doEl   [...]
    mov      dword [ans], [...]
    mov      byte [errFlg], T[...]
skpEl [...]

```

In this example, since the data was signed, a signed division (**idiv**) and the appropriate conversion (**cdq** in this case) were required. It should also be noted that the **edx** register was overwritten even though it did not appear explicitly. If a value was previously placed in **edx** (or **rdx**), it has been altered.

### 7.7.3.1 Jump Out of Range

The target label is referred to as a short-jump. Specifically, that means the target label must be within  $\pm 128$  bytes from the conditional jump instruction. While this limit is not typically a problem, for very large loops, the assembler may generate an error referring to “jump out-of-range”. The unconditional jump (**jmp**) is not limited in range. If a “jump out-of-range” is generated, it can be eliminated by reversing the logic and using an unconditional jump for the long jump. For example, the following code:

```

    cmp      rcx [...]
    jne     startOfL [...]

```

might generate a “jump out-of-range” assembler error if the label, **startOfLoop**, is a long distance away. The error can be eliminated with the following code:

```

    cmp      rcx [...]
    je      endOfL [...]
    jmp     startOfL [...]
endOfLo [...]

```

Which accomplishes the same thing using an unconditional jump for the long jump and adding a conditional jump to a very close label.

The conditional jump instructions are summarized as follows:

Instruction	Explanation
<code>cmp &lt;op1&gt;, &lt;op2&gt;</code>	Compare <code>&lt;op1&gt;</code> with <code>&lt;op2&gt;</code> . Results are stored in the <b>rFlag</b> register. <i>Note 1</i> , operands are not changed. <i>Note 2</i> , both operands cannot be memory. <i>Note 3</i> , <code>&lt;op1&gt;</code> operand cannot be an immediate.
Examples:	<code>cmp rax, 5</code> <code>cmp ecx, edx</code> <code>cmp ax, word [wN]</code>
<code>je &lt;label&gt;</code>	Based on preceding comparison instruction, jump to <code>&lt;label&gt;</code> if <code>&lt;op1&gt; == &lt;op2&gt;</code> . Label must be defined exactly once.
Examples:	<code>cmp rax, 5</code> <code>je wasEq</code>
<code>jne &lt;label&gt;</code>	Based on preceding comparison instruction, jump to <code>&lt;label&gt;</code> if <code>&lt;op1&gt; != &lt;op2&gt;</code> . Label must be defined exactly once.
Examples:	<code>cmp rax, 5</code> <code>jne wasNotEq</code>
<code>jl &lt;label&gt;</code>	For signed data, based on preceding comparison instruction, jump to <code>&lt;label&gt;</code> if <code>&lt;op1&gt; &lt; &lt;op2&gt;</code> . Label must be defined exactly once.
Examples:	<code>cmp rax, 5</code> <code>jl wasL</code>
<code>jle &lt;label&gt;</code>	For signed data, based on preceding comparison instruction, jump to <code>&lt;label&gt;</code> if <code>&lt;op1&gt; ≤ &lt;op2&gt;</code> . Label must be defined exactly once.
Examples:	<code>cmp rax, 5</code> <code>jle wasLessOrEq</code>



Instruction	Explanation
<b>jg &lt;label&gt;</b>	For signed data, based on preceding comparison instruction, jump to <label> if <op1> > <op2>. Label must be defined exactly once.
Examples:	<code>cmp rax, 5 jg wasGreater</code>
<b>jge &lt;label&gt;</b>	For signed data, based on preceding comparison instruction, jump to <label> if <op1> ≥ <op2>. Label must be defined exactly once.
Examples:	<code>cmp rax, 5 jge wasGreaterOrEq</code>
<b>jb &lt;label&gt;</b>	For unsigned data, based on preceding comparison instruction, jump to <label> if <op1> < <op2>. Label must be defined exactly once.
Examples:	<code>cmp rax, 5 jb wasLess</code>
<b>jbe &lt;label&gt;</b>	For unsigned data, based on preceding comparison instruction, jump to <label> if <op1> ≤ <op2>. Label must be defined exactly once.
Examples:	<code>cmp rax, 5 jbe wasLessOrEq</code>
<b>ja &lt;label&gt;</b>	For unsigned data, based on preceding comparison instruction, jump to <label> if <op1> > <op2>. Label must be defined exactly once.
Examples:	<code>cmp rax, 5 ja wasGreater</code>



Instruction	Explanation
<code>jae &lt;label&gt;</code>	For unsigned data, based on preceding comparison instruction, jump to <code>&lt;label&gt;</code> if $<\text{op1}> \geq <\text{op2}>$ . Label must be defined exactly once.
Examples:	<code>cmp rax, 5 jae wasGreaterOrEq</code>

A more complete list of the instructions is located in Appendix E.

#### 7.7.4 Iteration

The basic control instructions outlined provide a means to iterate or loop.

A basic loop can be implemented consisting of a counter which is checked at either the bottom or top of a loop with a compare and conditional jump.

For example, assuming the following declarations:

```
lpCnt dq
sum dq
```

The following code would sum the odd integers from 1 to n.

```
mov rcx, qword [lpCnt] ; loop counter
mov rax, 1 ; odd integer counter
sumLoop:
    add qword [sum], rax ; sum current odd integer
    add rax, 2 ; set next odd integer
    dec rcx ; decrement loop counter
    cmp rcx, [lpCnt] ; compare loop counter to lpCnt
    jne sumLoop ; if not equal, loop back to sumLoop
```

This is just one of many different ways to accomplish the odd integer summation task. In this example, `rcx` was used as a loop counter and `rax` was used for the current odd integer (appropriately initialized to 1 and incremented by 2).

The process shown using `rcx` as a counter is useful when looping a predetermined number of times. There is a special instruction, `loop`, provides looping support.

The general format is as follows:

```
loop <label>
```

Which will perform the decrement of the **rcx** register, comparison to 0, and jump to the specified label if **rcx** ≠ 0. The label must be defined exactly once.

As such, the loop instruction provides the same functionality as the three lines of code from the previous example program. The following sets of code are equivalent:

Code Set 1

```
loop  <label>
```

Code Set 2

```
dec   rcx
cmp   rcx
jne  <label>
```

For example, the previous program can be written as follows:

```
mov    rcx, qword [maxN]      ; loop count
mov    rax, 1                  ; odd integer count
sumLo<label>
add    qword [sum], rax        ; sum current odd integer
add    rax, 2                  ; set next odd integer
loop   sumLo
```

Both code examples produce the exact same result in the same manner.

Since the **rcx** register is decremented and then checked, forgetting to set the **rcx** register could result in looping an unknown number of times. This is likely to generate an error during the loop execution, which can be very misleading when debugging.

The **loop** instruction can be useful when coding, but it is limited to the **rcx** register and to counting down. If nesting loops are required, the use of a loop instruction for both the inner and outer loop can cause a conflict unless additional actions are taken (i.e. save/restore **rcx** register as required for inner loop).

While some of the programming examples in this text will use the loop instruction, it is not required.

The loop instruction is summarized as follows:

Instruction	Explanation
<b>loop &lt;label&gt;</b>	Decrement <b>rcx</b> register and jump to <b>&lt;label&gt;</b> if <b>rcx</b> is ≠ 0. Note, label must be defined exactly once.

Instruction	Explanation
Examples:	<pre>loop  startLoop loop  ifDone loop  sumLoop</pre>

A more complete list of the instructions is located in Appendix E.

## 7.8 Example Program, Sum of Squares

The following is a complete example program to find the sum of squares from 1 to  $n$ . For example, the sum of squares for 10 is as follows:

$$\boxed{1}^2 + \boxed{2}^2 + \dots + \boxed{n}^2 = \boxed{\text{sum}}$$

This example main initializes the  $n$  value to 10 to match the above example.

```
; Simple example program to compute the
; sum of squares from 1 to n.
; *****
; Data declarations

section .data
; -----
; Define constants
SUCCESS equ 0 ; Successful operation
SYS_exit equ 60 ; call code for termination

; Define Data
n dd 10
sumOfSquares dq 0
; *****

section .text
global _start
_start:
; -----
; Compute sum of squares from 1 to n (inclusive)
```

```

; Approach:
;   for (i=1; i<=n; i++)
;     sumOfSquares += i██

    mov      rbx, 1██
    mov      ecx, dword██
sumLoop██
    mov      rax, rbx██
    mul      rax██
    add      dword [sumOfSquares], rax██
    inc      rbx██
    loop    sumLoop██

; -----
; Done, terminate program██

la██
    mov      rax, SYS_exit██
    mov      rdi, SUCCESS██
    sysc██ ; call code for exit██
                  ; exit with success██

```

The debugger can be used to examine the results and verify correct execution of the program.

## 7.9 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 7.9.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) Which of the following instructions is legal / illegal? As appropriate, provide an explanation.

1. mov rax,<sup>█</sup><sub>█</sub>
2. mov ax,<sup>█</sup><sub>█</sub>
3. mov al, 3<sup>█</sup><sub>█</sub>
4. mov rax, r<sup>█</sup><sub>█</sub>
5. mov rax, r1<sup>█</sup><sub>█</sub>

```
6. mov    54, e[=]
7. mov    rax, qword [qVar1][=]
8. mov    rax, qword [bVar1][=]
9. mov    rax, [qVar1][=]
10. mov   rax, q[=]
11. mov   eax, dword [bVar1][=]
12. mov   qword [qVar2], qword [qVar1][=]
13. mov   qword [bVar2], qword [qVar1][=]
14. mov   r15, [=]
15. mov   r16, [=]
16. mov   r11b, [=]
```

2) Explain what each of the following instructions do[=]

1. movzx rsi, byte [bVar1][=]
2. movsx rsi, byte [bVar1][=]

3) What instruction is used[=]

1. convert an *unsigned* byte in **al** into a word in [=]
2. convert a *signed* byte in **al** into a word in [=]

4) What instruction is used[=]

1. convert an *unsigned* word in **ax** into a double-word in [=]
2. convert a *signed* word in **ax** into a double-word in [=]

5) What instruction is used[=]

1. convert an *unsigned* word in **ax** into a double-word in **dx** [=]
2. convert a *signed* word in **ax** into a double-word in **dx** [=]

6) Explain the difference between the **cwd** instruction and the **mov**[=] instruction

7) Explain why the explicit type specification (*dword* in this example) is required on the first instruction and is not required on the second instruction[=]

1. add dword [dVar], [=]
2. add [dVar], e[=]



- 8) Given the following code fragment

```
mov    rax [ ]  
mov    rbx [ ]  
add    rbx, [ ]
```

What would be in the **rax** and **rbx** registers after execution? Show answer in hex, full register size.



- 9) Given the following code fragment

```
mov    rax [ ]  
mov    rbx [ ]  
sub    rax, [ ]
```

What would be in the **rax** and **rbx** registers after execution? Show answer in hex, full register size.



- 10) Given the following code fragment

```
mov    rax [ ]  
mov    rbx [ ]  
sub    rbx, [ ]
```

What would be in the **rax** and **rbx** registers after execution? Show answer in hex, full register size.



- 11) Given the following code fragment

```
mov    rax [ ]  
mov    rbx [ ]  
imul   [ ]
```

What would be in the **rax** and **rdx** registers after execution? Show answer in hex, full register size.



- 12) Given the following code fragment

```
mov    rax [ ]  
cqo  
mov    rdx [ ]  
idiv   [ ]
```

What would be in the **rax** and **rdx** registers after execution? Show answer in hex, full register size.



13) Given the following code fragm

```
    mov    rax, 11  
    cqo  
    mov    rbx, 4  
    idiv   rbx
```



What would be in the **rax** and **rdx** registers after execution? Show answer in hex, full register size.



14) Explain why each of the following statements will not w

1. mov 42, ea
2. div :
3. mov dword [num1], dword [num1]
4. mov dword [ax], 8



15) Explain why the following code fragment will not work corre

```
    mov  eax,  
    mov  ebx,  
    idiv ebx
```



16) Explain why the following code fragment will not work corre

```
    mov  eax, -  
    cdq  
    mov  ebx,  
    div  ebx
```



17) Explain why the following code fragment will not work corre

```
    mov  ax, -  
    cwd  
    mov  bx,  
    idiv bx  
    mov  dword [ans],
```



18) Under what circumstances can the three-operand multiple be us





## 7.9.2 Suggested Projects



Below are some suggested projects based on this chapter.



- 1) Create a program to compute the following expressions using unsigned **b** sized variables and unsigned operations. Note, the first letter of the variable name denotes the size (**b** → byte and **w** → word).



1. **bAns1** = **bNum1** + **bNum2**
2. **bAns2** = **bNum1** + **bNum3**
3. **bAns3** = **bNum3** + **bNum1**
4. **bAns6** = **bNum1** - **bNum2**
5. **bAns7** = **bNum1** - **bNum3**
6. **bAns8** = **bNum2** - **bNum1**
7. **wAns11** = **bNum1** \* **bNum2**
8. **wAns12** = **bNum2** \* **bNum1**
9. **wAns13** = **bNum2** \* **bNum3**
10. **bAns16** = **bNum1** / **bNum2**
11. **bAns17** = **bNum3** / **bNum1**
12. **bAns18** = **wNum1** / **bNum2**
13. **bRem18** = **wNum1** % **bNum2**



Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.



- 2) Repeat the previous program using signed values and signed operations. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.
- 3) Create a program to complete the following expressions using unsigned **w** sized variables. Note, the first letter of the variable name denotes the size (**w** → word and **d** → double-word).



1. **wAns1** = **wNum1** + **wNum2**
2. **wAns2** = **wNum1** + **wNum3**
3. **wAns3** = **wNum3** + **wNum1**



```
4. wAns6 = wNum1 - wNuANSWER  
5. wAns7 = wNum1 - wNuANSWER  
6. wAns8 = wNum2 - wNuANSWER  
7. dAns11 = wNum1 * wNuANSWER  
8. dAns12 = wNum2 * wNuANSWER  
9. dAns13 = wNum2 * wNuANSWER  
10. wAns16 = wNum1 / wNuANSWER  
11. wAns17 = wNum3 / wNuANSWER  
12. wAns18 = dNum1 / wNuANSWER  
13. wRem18 = dNum1 % wNuANSWER
```

Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.

- 4) Repeat the previous program using signed values and signed operations. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.
- 5) Create a program to complete the following expressions using unsigned double-word sized variables. *Note*, the first letter of the variable name denotes the size (**d** → double-word and **q** → quadword).

```
1. dAns1 = dNum1 + dNumANSWER  
2. dAns2 = dNum1 + dNumANSWER  
3. dAns3 = dNum3 + dNuANSWER  
4. dAns6 = dNum1 - dNuANSWER  
5. dAns7 = dNum1 - dNuANSWER  
6. dAns8 = dNum2 - dNuANSWER  
7. qAns11 = dNum1 * dNuANSWER  
8. qAns12 = dNum2 * dNuANSWER  
9. qAns13 = dNum2 * dNuANSWER  
10. dAns16 = dNum1 / dNuANSWER  
11. dAns17 = dNum3 / dNuANSWER
```



12. `dAns18 = qNum1 / dNum1`
13. `dRem18 = qNum1 % dNum1`



Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.



- 6) Repeat the previous program using signed values and signed operations. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.
- 7) Implement the example program to compute the sum of squares from 1 to  $n$ . Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.
- 8) Create a program to compute the square of the sum from 1 to  $n$ . Specifically, compute the sum of integers from 1 to  $n$  and then square the value. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results in both decimal and hexadecimal.
- 9) Create a program to iteratively find the  $n$ th Fibonacci number<sup>37</sup>. The value of  $n$  should be set as a parameter (e.g., a programmer defined constant). The formula for computing Fibonacci is as follows:

$$\text{fibonacci}(n) = \begin{cases} n & \text{if } n=0 \text{ or } n=1 \\ \text{fibonacci}(n-1) + \text{fibonacci}(n-2) & \text{if } n > 1 \end{cases}$$

Use the debugger to execute the program and display the final results. Test the program for various values of  $n$ . Create a debugger input file to show the results in both decimal and hexadecimal.

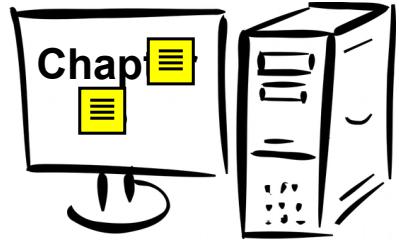



---

<sup>37</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Fibonacci\\_number](http://en.wikipedia.org/wiki/Fibonacci_number)



*Why did the programmer quit his job?*  
Because he didn't get arrays.



## 8.0 Addressing Modes

This chapter provides some basic information regarding addressing modes and the associated address manipulations on the x86-64 architecture.

The addressing modes are the supported methods for accessing a value in memory using the address of a data item being accessed (read or written). This might include the name of a variable or the location in an array.

The basic addressing modes

- Register
- Immediate
- Memory

Each of these modes is described with examples in the following sections. Additionally a simple example for accessing an array is presented.

### 8.1 Addresses and Values

On a 64-bit architecture, addresses require 64-bits.

As noted in the previous chapter, the only way to access memory is with the brackets ([]'s). Omitting the brackets will not access memory and instead obtain the address of the item. For example:

```
mov    rax, qword [var1]      ; value of var1 in memory  
mov    rax, var1             ; address of var1 in memory
```

Since omitting the brackets is not an error, the assembler will not generate error messages or warnings.

## Chapter 8.0 ◀ Addressing Modes

When accessing memory, in many cases the operand size is clear. For example, the instruction

```
mov eax, [r] [list icon]
```

moves a double-word from memory. However, for some instructions the size can be ambiguous. For example,

```
inc [rbx] ; er [list icon]
```

is ambiguous since it is not clear if the memory being accessed is a byte, word, or double-word. In such a case, operand size must be specified with either the *byte*, *word*, or *dword*, *qword* size qualifier. For example,

```
inc byte [rbx]  
inc word [rbx]  
inc dword [rbx] [list icon]
```

each instruction requires the size specification in order to be clear and legal.

### 8.1.1 Register Mode Addressing

Register mode addressing means that the operand is a CPU register (**eax**, **ebx**, etc.). For example:

```
mov eax, [list icon]
```

Both **eax** and **ebx** are in register mode addressing.

### 8.1.2 Immediate Mode Addressing

Immediate mode addressing means that the operand is an immediate value. For example:

```
mov eax, [list icon]
```

The destination operand, **eax**, is register mode addressing. The **123** is immediate mode addressing. It should be clear that the destination operand in this example cannot be immediate mode.

### 8.1.3 Memory Mode Addressing

Memory mode addressing means that the operand is a location in memory (accessed

an address). This is referred to as *indirection* or *dereferencing*.

The most basic form of memory mode addressing has been used extensively in the previous chapter. Specifically, the instruction:

```
mov rax, qword [qNum]
```

Will access the memory location of the variable **qNum** and retrieve the value stored there. This requires that the CPU wait until the value is retrieved before completing the operation and thus might take slightly longer to complete than a similar operation using an immediate value.

When accessing arrays, a more generalized method is required. Specifically, an address can be placed in a register and indirection performed using the register (instead of the variable name).

For example, assuming the following declaration:

```
lst dd 101, 103, 105,
```

The decimal value of 101 is 0x00000065 in hex. The memory picture would be as follows:

Value	Address	Offset	Label
00	0x6000ef	lst + 0	
00	0x6000ee	lst + 4	
00	0x6000ed	lst + 8	
6b	0x6000ec	lst + 12	
00	0x6000eb	lst + 16	
00	0x6000ea	lst + 20	
00	0x6000e9	lst + 24	
69	0x6000e8	lst + 28	
00	0x6000e7	lst + 32	
00	0x6000e6	lst + 36	
00	0x6000e5	lst + 40	
67	0x6000e4	lst + 44	
00	0x6000e3	lst + 48	
00	0x6000e2	lst + 52	
00	0x6000e1	lst + 56	
65	0x6000e0	lst + 60	

The first element of the array could be accessed as follo

```
mov    eax, dword [1]
```

Another way to access the first element is as follo

```
mov    rbx, 1
mov    eax, dword [rbx]
```

In this example, the starting address, or base address, of the list is placed in **rbx** (first line) and then the value at that address is accessed and placed in the **rax** register (second line). This allows us to easily access other elements in the array.

Recall that memory is “byte addressable”, which means that each address is one byte of information. A double-word variable is 32-bits or 4 bytes so each array element uses 4 bytes of memory. As such, the next element (103) is the starting address (lst) plus 4 and the next element (105) is the starting address (lst) 8.

Increasing the offset by 4 for each successive element. A list of bytes would increase by 1, a list of words would increase by 2, a list of double-words would increase by 4, and a list of quadwords would increase by 8.

The offset is the amount added to the base address. The index is the array element number as used in a high-level language.

There are several ways to access the array elements. One is to use a base address and add a displacement. For example, given the initializations:

```
mov    rbx
mov    rsi
```

Each of the following instructions access the third element (105 in the above)

```
mov    eax, dword [lst]
mov    eax, dword [rbx]
mov    eax, dword [rbx+r]
```

In each case, the starting address plus 8 was accessed and the value of 105 placed in the **eax** register. The displacement is added and the memory location accessed while none of the source operand registers (**rbx**, **rsi**) are altered. The specific method used is up to the programmer.

In addition, the displacement may be computed in more complex w

The general format of memory addressing is as follows:

[ **baseAddr** + (**indexReg** \* **scaleValue** ) + **displacement** ]

Where **baseAddr** is a register or a variable name. The **indexReg** must be a register. The **scaleValue** is an immediate value of 1, 2, 4, 8 (1 is legal, but not useful). The **displacement** must be an immediate value. The total represents a 64-bit address.

Elements may be used in any combination, but must be legal and result in a valid address.

Some examples of memory addressing for the source operand are as follows:

```
mov    eax, dword [val]      ; base
mov    rax, qword [rbx+r]    ; index
mov    ax, word [lst]        ; scale 1
mov    bx, word [lst+rdx]    ; displacement
mov    rcx, qword [lst+(rsi*4)] ; base + index * scale
mov    al, byte [buff-1+r]   ; displacement
mov    eax, dword [rbx+(rsi*4)+d] ; base + index * scale + displacement
```

For example, to access the 3<sup>rd</sup> element of the previously defined double-word array (which is index 2 since index's start at 0):

```
mov    rsi, 2                ; index
mov    eax, dword [lst+rsi*4] ; get lst[rsi]
```

Since addresses are always *qword* (on a 64-bit architecture), a 64-bit register is used for the memory mode addressing (even when accessing double-word values). This allows a register to be used more like an array index (from a high-level language).

For example, the memory operand, **[lst+rsi\*4]**, is analogous to **lst[rsi]** from a high-level language. The **rsi** register is multiplied by the data size (4 in this example since each element is 4 bytes).

## 8.2 Example Program, List Summarization

The following example program will sum the numbers in a list.

```
; Simple example to the sum and average for
; a list of numbers.
```

```
; **** Data declarations ****
```

```
section    .data
```

```

; -----
; Define constants
EXIT_SUCCESS    equ 0          ; successful operation
SYS_exit        equ 60         ; call code for termination

; -----
; Define Data
section .data
lst    dd 1002, 1004, 1006, 1008, 1010
len    dd 5
sum    dd 0

; *****
section .text
global _start
_start:

; -----
; Summation loop
    mov    ecx, dword [len]      ; get length value
    mov    rsi, 0                ; index register

sumLoop:
    mov    eax, dword [lst+(rsi*4)] ; get lst[rsi]
    add    dword [sum], eax       ; update sum
    inc    rsi                   ; next index
    loop   sumLoop               ; loop back

; -----
; Done, terminate program
label:
    mov    rax, SYS_exit         ; call code for exit
    mov    rdi, EXIT_SUCCESS     ; exit with success
    syscall

```

The ()'s within the [ ]'s are not required and added only for clarity. As such, the `[lst+(rsi*4)]`, is exactly the same as `[lst+rsi*4]`.

## 8.3 Example Program, Pyramid Areas and Volumen

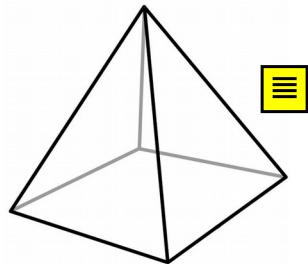
This example is a simple assembly language program to calculate some geometric information for each square pyramid in a series of square pyramids. Specifically, the program will find the lateral total surface area (including the base) and volume of each square pyramid in a set of square pyramids.

Once the values are computed, the program finds the minimum, maximum, sum, and average for the total surface areas and volumes.

All data are unsigned values (i.e., uses **mul** and **div**, not **imul** or **idiv**).

This basic approach used in this example is the loop to calculate the surface areas and volumes arrays. A second loop is used to find the sum, minimum, and maximum for each array. To find the minimum and maximum values, the minimum and maximum variables are each initialized to the first value in the list. Then, every element in the list is compared to the current minimum and maximum. If the current value from the list is less than the current minimum, the minimum is set to the current value (over-writing the previous value). When all values have been checked, the minimum will represent the true minimum from the list. If the current value from the list is more than the current maximum, the maximum is set to the current value (over-writing the previous value). When all values have been checked, the maximum will represent the true maximum from the list.

```
; Example assembly language program to calculate the  
; geometric information for each square pyramid in  
; a series of square pyramids.  
  
; The program calculates the total surface area  
; and volume of each square pyramid.  
  
; Once the values are computed, the program finds  
; the minimum, maximum, sum, and average for the  
; total surface areas and volumes.  
  
; -----  
; Formulas:  
; totalSurfaceAreas(n) = aSides(n)  
; (2*aSides(n)*sSides(n))  
; volumes(n) = (aSides(n)^2 * heights(n))
```



```
; ****
section      .data
; -----
; Define constants
EXIT_SUCCESS    equ     0          ; successful operation
SYS_exit        equ     60         ; call code for termination

; -----
; Provided Data
aSides      db      10,      14,      13,      37,
            db      31,      13,      20,      61,
            db      14,      53,      44,      19,
            db      27,      41,      53,      62,
            db      19,      18,      14,      10,
            db      15,      11,      22,      33,
            db      15,      23,      15,      63,
            db      24,      33,      10,      61,
            db      14,      34,      13,      71,
            db      38,      13,      29,      17,
sSides      dw      1233,    1114,    1773,    1131,    1123,
            dw      1164,    1973,    1974,    1123,    1142,
            dw      1344,    1752,    1973,    1142,    1175,
            dw      1165,    1754,    1273,    1175,    1567,
            dw      1153,    1673,    1453,    1567,    1567,
            dw      1144,    1579,    1764,    1567,    1567,
            dw      1456,    1563,    1564,    1753,    1753,
            dw      1646,    1862,    1457,    1167,    1167,
            dw      1867,    1864,    1757,    1755,    1756,
heights     dd      14145,   11134,   15123,   15123,   14145,
            dd      18454,   15454,   12156,   12164,   12164,
            dd      18453,   18453,   11184,   15142,   12156,
            dd      14564,   14134,   12156,   12344,   13134,
            dd      11153,   18543,   17156,   12352,   15156,
            dd      18455,   14134,   12123,   15324,   13134,
            dd      11134,   14134,   15156,   15234,   17175
```

```

dd 19567, 14134, 12134, 17546, 16
dd 11134, 14134, 14576, 15457, 17
dd 13153, 11153, 12184, 14142, 17

length      dd
taMin        dd
taMax        dd
taSum        dd
taAve        dd

volMin       dd
volMax       dd
volSum       dd
volAve       dd

; -----
; Additional variables
ddTwo        dd
ddThree      dd

; -----
; Uninitialized data
section      .bss
totalAreas   resd
volumes      resd

; *****
section      .text
global _start
_start:

; Calculate volume, lateral and total surface area
    mov     ecx, dword [length]           ; length counter
    mov     rsi, 0                         ; index
calculationLoop:
    ; totalAreas(n) = aSides(n) * (2*aSides(n)*sSides(n))
    movzx   r8d, byte [aSides+rsi]        ; aSides

```

## Chapter 8.0 ◀ Addressing Modes

```
movzx  r9d, word [sSides+rsi*2]           ; sSides
mov    eax, [rsi]
mul    ddword [ddT]
mul    ddword [ddT]
mul    ddword [ddT]
mov    dword [totalAreas+rsi*4], [rsi]

; volumes(n) = (aSides(n)^2 * heights(n))
; -----
; Find min, max, sum, and average for the total
; areas and volumes.

mov    eax, dword [totalArea]
mov    dword [taMin], [rsi]
mov    dword [taMax], [rsi]

mov    eax, dword [volume]
mov    dword [volMin], [rsi]
mov    dword [volMax], [rsi]

mov    dword [taSum], [rsi]
mov    dword [volSum], [rsi]

mov    ecx, dword [length]
mov    rsi, [rsi]

statsLoop:
    mov    eax, dword [totalAreas+rsi]
    add    dword [taSum], [rsi]

    cmp    eax, dword [taMin]
    jae    notNewTotal
    mov    dword [taMin], [rsi]
```

```

notNewTaM:
    cmp    eax, dword [taM]
    jbe    notNewTaM
    mov    dword [taMax], 

notNewTaM:
    mov    eax, dword [volume+rsi]
    add    dword [volSum], 
    cmp    eax, dword [volM]
    jae    notNewVolM
    mov    dword [volMin], 

notNewVolM:
    cmp    eax, dword [volM]
    jbe    notNewVolM
    mov    dword [volMax], 

notNewVolM:
    inc    rsi
    loop   statsL

; -----
; Calculate average

    mov    eax, dword [taS]
    mov    edx, 
    div    dword [leng]
    mov    dword [taAve], 

    mov    eax, dword [volS]
    mov    edx, 
    div    dword [leng]
    mov    dword [volAve], 

; -----
; Done, terminate program

    la
    mov    rax, SYS_exit
    mov    rdi, EXIT_SUCCESS
    sysc

; call code for exit
; exit with success

```

This is one example. There are multiple other valid approaches to solving this problem.

**8.4 Exercises**

Below are some quiz questions and suggested projects based on this chapter.

**8.4.1 Quiz Questions**

Below are some quiz questions based on this chapter.

- 1) Explain the difference between the following two instructions:

1. `mov rdx, qword [qVar]`
2. `mov rdx, qVar`

- 2) What is the address mode of the source operand for each of the instructions listed below. Respond with *Register*, *Immediate*, *Memory*, or *Illegal Instruction*.

Note, `mov <dest>, <sour`

```

mov    ebx,
mov    ecx, dword [rdx]
mov    byte [rbx+4],
mov    10,
mov    dl,
mov    ax, word [rsi]
mov    cx, word [rbx+r]
mov    ax, byte [r]

```

- 3) Given the following variable declarations and code fragments:

```

ans1   dd
mov    rax
mov    rbx, a
add    eax, dword [r]

```

What would be in the `eax` register after execution? Show answer in hex, full register size.

- 4) Given the following variable declarations and code fragm

list1 dd 2, 3, 4, 5, 6

```
mov rbx, list1  
add rbx, 4  
mov eax, dword [rbx]  
mov edx, dword [lis
```

What would be in the **eax** and **edx** registers after execution? Show answer in hex, full register size.

- 5) Given the following variable declarations and code fragm

lst dd 2, 3, 5, 7

```
lp:  
    mov     rsi  
    mov     eax  
    mov     rcx  
    add     eax, dword [lst+r  
    add     rsi  
    loop  
    mov     ebx, dword [l
```

What would be in the **eax**, **ebx**, **rcx**, and **rsi** registers after execution? Show answer in hex, full register size. Note, pay close attention to the register sizes (32-bit vs. 64-bit).

- 6) Given the following variable declarations and code fragm

list dd 8, 6, 4, 2, 1

```
lp:  
    mov     rbx  
    mov     rsi  
    mov     rcx  
    mov     edx, dword [r  
    mov     eax, dword [list+rsi  
    inc  
    loop  
    imul    dword [li
```

What would be in the **eax**, **edx**, **rcx**, and **rsi** registers after execution? Show answer in hex, full register size. Note, pay close attention to the register sizes (32-bit vs. 64-bit).



- 7) Given the following variable declarations and code fragm



```
list      dd    8, 7, 6, 5, 4, 3, 2, 1
          mov    rbx, 1
          mov    rsi, 2
          mov    rcx, 3
          mov    edx, dword [rbx]
lp:      add    eax, dword [list+rsi]
          inc    rsi
          loop   1
          cdq
          idiv   dword [list]
          
```

What would be in the **eax**, **edx**, **rcx**, and **rsi** registers after execution? Show answer in hex, full register size. Note, pay close attention to the register sizes (32-bit vs. 64-bit).



- 8) Given the following variable declarations and code fragm



```
list      dd    2, 7, 4, 5, 6
          mov    rbx, 1
          mov    rsi, 2
          mov    rcx, 3
          mov    eax, 4
          mov    edx, dword [rbx]
lp:      add    eax, dword [rbx+rsi]
          add    rsi, 1
          loop   1
          imul   dword [rsi]
          
```

What would be in the **eax**, **edx**, **rcx**, and **rsi** registers after execution? Show answer in hex, full register size. Note, pay close attention to the register sizes (32-bit vs. 64-bit).



### 8.4.2 Suggested Proj



Below are some suggested projects based on this chap



- 1) Implement the example program to sum a list of numbers. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.



- 2) Update the example program from the previous question to find the maximum, minimum, and average for the list of numbers. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 3) Implement the example program to compute the lateral total surface area (including the base) and volume of each square pyramid in a set of square pyramids. Once the values are computed, the program finds the minimum, maximum, sum, and average for the total surface areas and volumes. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 4) Write an assembly language program to find the minimum, middle value, maximum, sum, and integer average of a list of numbers. Additionally, the program should also find the sum, count, and integer average for the negative numbers. The program should also find the sum, count, and integer average for the numbers that are evenly divisible by 3. Unlike the median, the 'middle value' does not require the numbers to be sorted. *Note*, for an odd number of items, the middle value is defined as the median value. For an even number of values, it is the integer average of the two median values. Assume all data is unsigned. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 5) Repeat the previous program using signed values and signed operations. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

- 6) Create a program to sort a list of numbers. Use the following bubble sort algorithm:

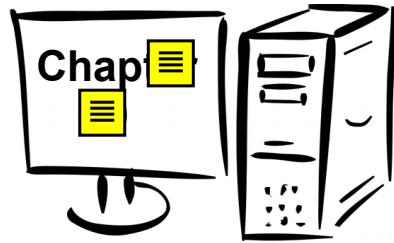
```
for ( i = (len-1) to 0 ) {  
    swapped = false  
    for ( j = 0 to i-1 ) {  
        if ( lst(j) > lst(j+1) ) {  
            tmp = lst(j)  
            lst(j) = lst(j+1)  
            lst(j+1) = tmp  
            swapped = true  
        }  
        if ( swapped = false ) exit  
    }  
}
```

Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

---

38 For more information, refer to: [http://en.wikipedia.org/wiki/Bubble\\_sort](http://en.wikipedia.org/wiki/Bubble_sort)

*A programmer is heading out to the grocery store, and is asked to "get a gallon of milk, and if they have eggs, get a dozen." He returns with 12 gallons of milk.*



## 9.0 Process Sta



In a computer, a stack is a type of data structure where items are added and then removed from the stack in reverse order. That is, the most recently added item is the very first one that is removed. This is often referred to as Last-In, First-Out (LIFO).



A stack is heavily used in programming for the storage of information during procedure or function calls. The following chapter provides information and examples regarding the stack.



Adding an item to a stack is referred to as a **push** or push operation. Removing an item from a stack is referred to as a **pop** or pop operation.



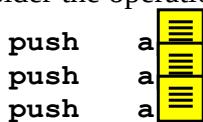
It is expected that the reader will be familiar with the general concept of a st



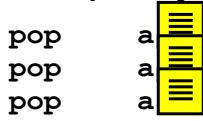
### 9.1 Stack Exam



To demonstrate the general usage of the stack, given an array, `a = {7, 19, 37}`, consider the operations:



Followed by the operati

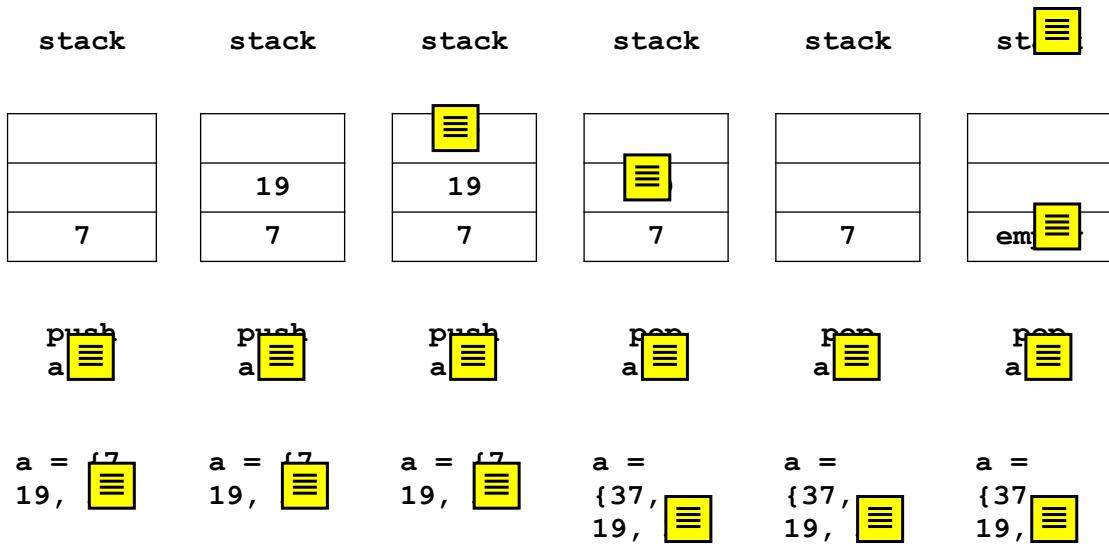


The initial push will push the 7, followed by the 19, and finally the 37. Since the stack is last-in, first-out, the first item popped off the stack will be the last item pushed, or 37 in this example. The 37 is placed in the first element of the array (over-writing the 7). As this continues, the order of the array elements is reversed.



## Chapter 9.0 ◀ Process S

The following diagram shows the progress and the results of stack operations.



The following sections provide more detail regarding the stack implementation and applicable stack operations and instructions.

## 9.2 Stack Instructions

A push operation puts things onto the stack, and a pop operation takes things off the stack. The format for these commands is:

**push <operand**  
**pop <operand**

The operand can be a register or memory, but an immediate is not allowed. In general, push and pop operations will push the architecture size. Since the architecture is 64-bit, we will push and pop quadwords.

The stack is implemented in reverse in memory. Refer to the following sections for a detailed explanation of why.

The stack instructions are summarized as follows:

Instruction	Explanation
<code>push &lt;op64&gt;</code>	Push the 64-bit operand on the stack. First, adjusts <b>rsp</b> accordingly ( <b>rsp</b> -8) and then copy the operand to <b>[rsp]</b> . The operand may not be an immediate value. Operand is not changed.
Examples:	<code>push rax</code> <code>push qword [qVal] ; value</code> <code>push qVal ; address</code>
<code>pop &lt;op64&gt;</code>	Pop the 64-bit operand from the stack. Adjusts <b>rsp</b> accordingly ( <b>rsp</b> +8). The operand may not be an immediate value. Operand is overwritten.
Examples:	<code>pop rax</code> <code>pop qword [qVal]</code> <code>pop rsi</code>

If more than 64-bits must be pushed, multiple push operations would be required. While it is possible to push and pop operands less than 64-bits, it is not recommended.

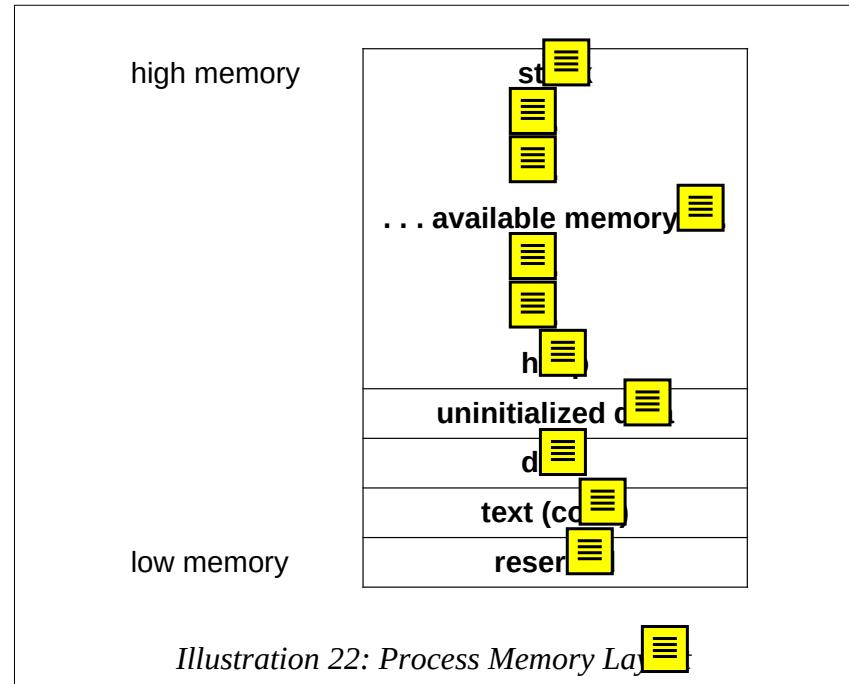
A more complete list of the instructions is located in Appendix E.

## 9.3 Stack Implementation

The **rsp** register is used to point to the current top of stack in memory. In this architecture, as with most, the stack is implemented growing downward in memory.

### 9.3.1 Stack Layout

As noted in Chapter 2, Architecture, the general memory layout for a program is as follows:

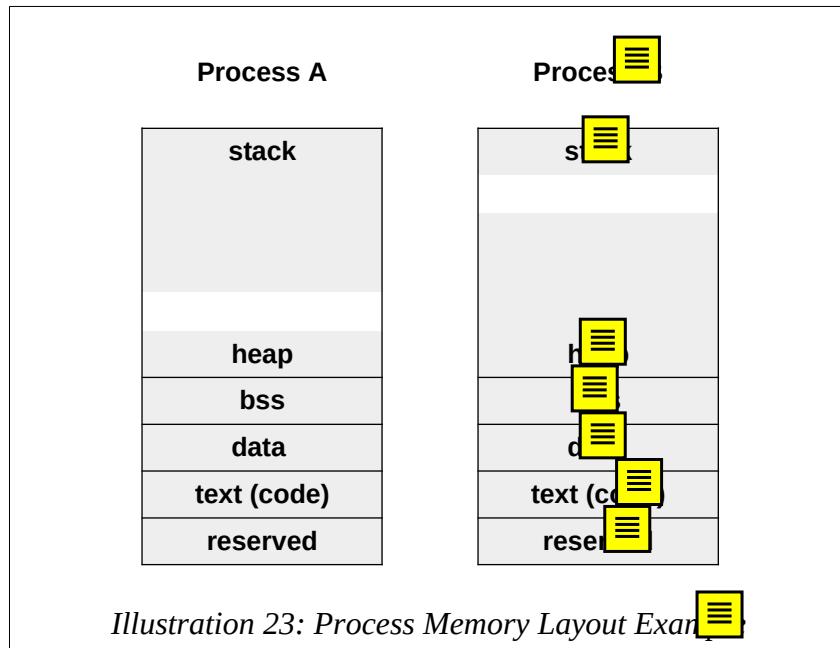


The heap is where dynamically allocated data will be placed (if requested). For example, items allocated with the C++ `new` operator or the C `malloc()` system call. As dynamically allocated data is created (at run-time), the heap typically grows upward. However, the stack starts in high memory and grows downward. The stack is used to temporarily store information such as call frames for function calls. A large program or a recursive function may use a significant amount of stack space.

As the heap and stack expand, they grow toward each other. This is done to ensure the most effective overall use of memory.

A program (Process A) that uses a significant amount of stack space and a minimal amount of heap space will function. A program (Process B) that uses a minimal amount of stack space and a very large amount of heap space will also function.

For example:



Of course, if the stack and heap meet, the program will crash. If that occurs, there is no memory available.

### 9.3.2 Stack Operations

The basic stack operations of push and pop adjust the stack pointer register, **rsp**, during their operation.

For a push operation:

1. The **rsp** register is decreased by 8 (1 quadword).
2. The operand is copied to the stack at [**rsp**].

The operand is not altered. The order of these operations is important.

For a pop operation:

1. The current top of the stack, at [**rsp**], is copied into the operand.
2. The **rsp** register is increased by 8 (1 quadword).

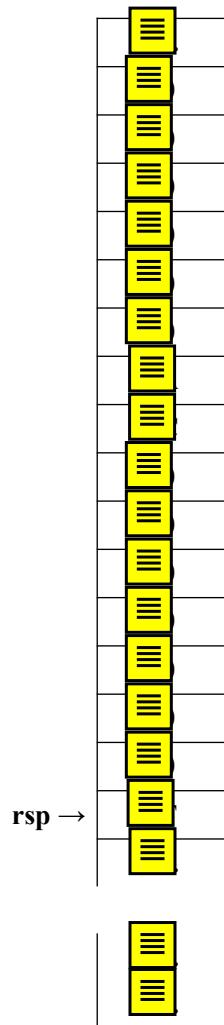
The order of these operations is the exact reverse of the push. The item popped is not actually deleted. However, the programmer cannot count on the item remaining on the stack after the pop operation. Previously pushed, but not popped, items can be accessed.

## Chapter 9.0 ◀ Process S

For example:

```
mov    [rsp], 6700      ; 670010 = 00001A
push   [rsp]
mov    [rsp], 31        ; 3110 = 000000
push   [rsp]
```

Would produce the following stack configuration (where each box is a byte):



The layout shows the architecture is little-endian in that the least significant byte is placed into the lowest memory location.

## 9.4 Stack Examples

The following is an example program to use the stack to reverse a list of quadwords in place. Specifically, each value in a quadword array is placed on the stack in the first loop. In the second loop, each element is removed from the stack and placed back into the array (over-writing) the previous value.

```
; Simple example demonstrating basic stack operations

; Reverse a list of numbers - in place.
; Method: Put each number on stack, then pop each number
;           back off, and then put back into memory.

; ****
; Data declarations

section      .data

; -----
; Define constants

EXIT_SUCCESS    equ     0          ; successful operation
SYS_exit        equ     60         ; call code for termination

; -----
; Define Data

numbers          dq      121, 122, 123, 124,
len              dq

; ****
section      .text
global _start
_start:

; Loop to put numbers on stack

    mov     rcx, qword [l]
    mov     rbx, numbe
    mov     r12
    mov     rax

pushLoop:
    push   qword [rbx+r12]

; ****
```

```

inc      [ ]      [ ]
loop    pushL [ ]      [ ]
; -----
; All the numbers are on stack (in reverse order).
; Loop to get them back off. Put them back into
; the original list...
popLo [ ]      [ ]
        mov rcx, qword [l]      [ ]
        mov rbx, numb [ ]      [ ]
        mov r12 [ ]      [ ]
        pop [ ]      [ ]
        mov qword [rbx+r12*8], [ ]      [ ]
        inc [ ]      [ ]
        loop popL [ ]      [ ]
; -----
; Done, terminate program [ ]
la [ ]      [ ]
        mov rax, SYS_exit      ; call code for [ ]
        mov rdi, EXIT_SUCCESS  ; exit with success [ ]
        sysc [ ]      [ ]

```

There are other ways to accomplish this function (reversing a list), however this is meant to demonstrate the stack operations.

## 9.5 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 9.5.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) Which register refers to the top of the stack?
- 2) What happens as a result of a `push rax` instruction (two thin lines)?
- 3) How many **bytes** of data does the `pop rax` instruction remove from the stack?

- 4) Given the following code fragment

```
    mov    r10
    mov    r11
    mov    r12
    push
    push
    push
    pop
    pop
    pop
```

What would be in the **r10**, **r11**, and **r12** registers after execution? Show answer in hex, full register size.

- 5) Given the following variable declarations and code fragment

```
lst    dq   1, 3, 5, 7
      mov    rsi
      mov    rcx
lp1: push   word [lst+rsi]
      inc
      loop
      mov    rsi
      mov    rcx
lp2: pop    word [lst+rsi]
      inc
      loop
      mov    rbx, qword [l
```

Explain what would be the **result** of the code (after execution)

- 6) Provide one advantage to the stack growing downward in memory

### 9.5.2 Suggested Projects

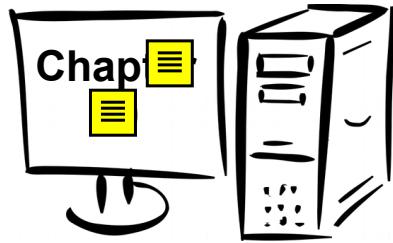
Below are some suggested projects based on this chapter

- 1) Implement the example program to reverse a list of numbers. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

- 2) Create a program to determine if a NULL terminated string representing a word is a palindrome<sup>39</sup>. A palindrome is a word that reads the same forward or backwards. For example, “anna”, “civic”, “hannah”, “kayak”, and “madam” are palindromes. This can be accomplished by pushing the characters on the stack one at a time and then comparing the stack items to the string starting from the beginning. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 3) Update the previous program to test if a phrase is a palindrome. The general approach using the stack is the same, however, spaces and punctuation must be skipped. For example, “A man, a plan, a canal – Panama!” is a palindrome. The program must ignore the comma, dash, and exclamation point. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

---

39 For more information, refer to: <http://en.wikipedia.org/wiki/Palindrome>



## 10.0 Program Development

Writing or developing programs is easier when following a clear methodology. The main steps in the methodology are:

- Understand the Problem
- Create the Algorithm
- Implement the Program
- Test/Debug the Program

To help demonstrate this process in detail, these steps will be applied to a simple example problem in the following sections.

### 10.1 Understand the Problem

Before attempting to create a solution, it is important to fully understand the problem. Ensuring a complete understanding of the problem can help reduce errors and save time and effort. The first step is to understand what is required, especially the applicable input information and expected results or output.

Consider the problem of converting a single integer number into a string or series of characters representing that integer. To be clear, an integer can be used for numeric calculations, but cannot be displayed to the console (as it is). A string can be displayed to the console but not used in numeric calculations.

For this example, only unsigned (positive only) values will be considered. The small extra effort to address signed values is left to the reader as an exercise.

As an unsigned double-word integer, the numeric value  $1498_{10}$  would be represented as  $0x000005DA$  in hex (double-word sized). The integer number  $1498_{10}$  ( $0x000005DA$ ) would be represented by the string “1”, “4”, “9”, “8” with a NULL termination. This would require a total of 5 bytes since there is no sign or leading spaces required for



specific example. As such, the string “1498” would be represented as follo



Character	“1”	“4”	“9”	“8”	NULL
ASCII Value (decimal)	49	52	57	56	
ASCII Value (hex)	0x31	0x34	0x39	0x38	

The goal is to convert the single integer number into the appropriate series of characters to form a NULL terminated string.



## 10.2 Create the Algorithm



The algorithm is the name for the unambiguous, ordered sequence of steps involved in solving the problem. Once the program is understood, a series of steps can be developed to solve that problem. There can be, and usually are, multiple correct solutions to a given problem.



The process for creating an algorithm can be different for different people. In general, some time should be devoted to thinking about possible solutions. This may involve working on some possible solutions using a scratch piece of paper. Once an approach is selected, that solution can be developed into an algorithm. The algorithm should be written down, reviewed, and refined. The algorithm is then used as the outline of the program.



For example, we will consider the integer to ASCII conversion problem outlined in the previous section. To convert a single digit integer (0-9) into a character,  $48_{10}$  (or “0” or  $0x30$ ) can be added to the integer. For example,  $0x01 + 0x30$  is  $0x31$  which is the ASCII value of “1”. It should be obvious that this trick will only work for single digit numbers (0-9).



In order to convert a larger integer ( 10) into a string, the integer must be broken into its component digits. For example,  $123_{10}$  ( $0x7B$ ) would be 1, 2, and 3. This can be accomplished by repeatedly performing integer division by 10 until a 0 result is obtained.



For example:

$$\frac{123}{10} = 12 \quad remainder$$

$$\frac{12}{10} = 1 \quad remainder$$

$$\frac{1}{10} = 0 \quad remainder$$

As can be seen, the remainder represents the individual digits. However, they are obtained in reverse order. To address this, the program can push the remainder and, when done dividing, pop the remainders and convert to ASCII and store in a string (which is an array of bytes).

This process forms the basis for the algorithm. It should be noted, that there are many ways to develop this algorithm. One such approach is shown as follows:

```
; Part A - Successive division
; digitCount = 0
; get integer
; divideLoop
;   divide number by
;   push remainder
;   increment digitCount
;   if (result > 0) goto divideLoop

; Part B - Convert remainders and store
; get starting address of string (array of bytes)
; index
; popLoop
;   pop intDigit
;   charDigit = intDigit + "0" (0x0)
;   string[idx] = charDigit
;   increment index
;   decrement digitCount
;   if (digitCount > 0) goto popLoop
;   string[idx] = NUL
```

The algorithm steps are shown as program comments for convenience. The algorithm is typically started on paper and then more formally written in pseudo-code as shown above. In the unlikely event the program does not work the first time, the comments are the primary debugging checklist.

Some programmers skip the comments and will end up spending much more time debugging. The commenting represents the algorithm and the code is the implementation of that algorithm.

### 10.3 Implement the Progr

Based on the algorithm, a program can be developed and implemented. The algorithm is expanded and the code added based on the steps outlined in the algorithm. This allows the programmer to focus on the specific issues for the current section being coded including the data types and data sizes. This example addresses only unsigned data so the unsigned divide (DIV, not IDIV) is used. Since the integer is a double-word, it must be converted into a quadword for the division. However, the result and the remainder after division will also be a double-words. Since the stack is quadwords, the entire quadword register will be pushed. The upper-order portion of the register will not be accessed, so its contents are not relevant.

One possible implementation of the algorithm is as follo

```
; Simple example program to convert an integer into an ASCII string.  
; *****  
; Data declarations  
  
section      .d  
  
; -----  
; Define consta  
  
NULL          equ             ; successful operation  
EXIT_SUCCESS  equ   0           ; code for termination  
SYS_exit      equ   60          ;  
  
; -----  
; Define Da  
  
intNum        dd    1
```

```

section .data
strNum    resb 100h
; *****
section .text
global _start
_start:
; Convert an integer to an ASCII string
; -----
; Part A - Successive division
    mov     eax, dword [intNum]           ; get integer
    mov     rcx, 0                      ; digitCount = 0
    mov     ebx, 10                     ; set for dividing by 10
divideLoop:
    mov     edx, eax                   ; divide number by 10
    div     ebx
    push   rdx                       ; push remainder
    inc    rcx                        ; increment digitCount
    cmp    eax, 0                     ; if (result > 0)
    jne    divideLoop                ;   goto divideLoop
; -----
; Part B - Convert remainders and store
    mov     rbx, strNum               ; get addr of str
    mov     rdi, 0                     ; idx = 0
popLoop:
    pop    rax                       ; pop intDiv
    add    al, "0"                   ; char = int + '0'
    mov    byte [rbx+rdi], al        ; string[idx] = char
    inc    rdi                      ; increment idx
    loop   popLoop                  ; if (digitCount > 0)
    mov    byte [rbx+rdi], NULL      ; string[idx] = NULL

```

```
; -----  
; Done, terminate program  
la mov rax, SYS_exit  
mov rdi, EXIT_SUCCESS ; call code for ; exit with succ  
sysc
```

There are many different valid implementations for this algorithm. The program should be assembled to address any typos or syntax errors.

## 10.4 Test/Debug the Program

Once the program is written, testing should be performed to ensure that the program works. The testing will be based on the specific parameters of the program.

In this case, the program can be executed using the debugger and stopped near the end of the program (e.g., at the label “last” in this example). After starting the debugger with **ddd**, the command **b last** and **run** can be entered which will run the program up to, but not executing the line referenced by the label “last”. The resulting string, **strNum** can be viewed in the debugger with **x/s &strNum** will display the string address and the contents which should be “1498”. For example;

```
(gdb) x/s &strNum  
0x600104: "14
```

If the string is not displayed properly, it might be worth checking each character of the five (5) byte array with the **x/5cb &strNum** debugger command. The output will show the address of the string followed by both the decimal and ASCII representation.

For example

```
(gdb) x/5cb &strNum  
0x600104: 49 '1' 52 '4' 57 '9' 56 '8' 0 '\000'
```

The format of this output can be confusing initially.

If the correct output is not provided, the programmer will need to debug the code. For this example, there are two main steps; successive division and conversion/storing the remainders. The second step requires the first step to work, so the first step should be verified. This can be done by using the debugger to focus only on the first section. In this example, the first step should iterate exactly 4 times, so **rcx** should be 4. Additionally, 8, 9, 4, and 1 should be pushed on the stack in that order. This is ea

verified in the debugger by looking at the register contents of **rdx** when it is pushed or by viewing the top 4 entries in the stack.

If that section works, the second section can be verified. Here, the values 1, 4, 9, and 8 should be coming off the stack (in that order). If so, the integer is converted into a character by adding “0” (0x30) and that stored in the string, one character at a time. The string can be viewed character by character to see if they are being entered into the string correctly.

In this manner, the problem can be narrowed down fairly quickly. Efficient debugging is a critical skill and must be honed by practice.

Refer to Chapter 6, DDD Debugger for additional information on specific debugger commands.

## 10.5 Error Terminology

In case the program does not work, it helps to understand some basic terminology about where or what the error might be. Using the correct terminology ensures that you can communicate effectively about the problem with others.

### 10.5.1 Assembler Errors

Assembler errors are generated when the program is assembled. This means that the assembler does not understand one or more of the instructions. The assembler will provide a list of errors and the line number of each error. It is recommended to address the errors from the top down. Resolving an error at the top can clear multiple errors further down.

Typical assembler errors include misspelling an instruction and/or omitting a variable declaration.

### 10.5.2 Run-time Errors

A run-time error is something that causes the program to crash.

### 10.5.3 Logic Errors

A logic error is when the program executes, but does not produce the correct result. For example, coding a provided formula incorrectly or attempting to compute the average of a series of numbers before calculating the sum.

If the program has a logic error, one way to find the error is to display intermediate values. Further information will be provided in later chapters regarding advice on finding logic errors.

## 10.6 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 10.6.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) What is an algorithm? 
- 2) What are the four main steps in algorithm development? 
- 3) Are the four main steps in algorithm development applicable only to assembly language programming? 
- 4) What type of error, if any, occurs if the one operand multiply instruction uses an immediate value operand? Respond with assemble-time or run-time. 
- 5) If an assembly language instruction is spelled incorrectly (e.g., “mv” instead of “mov”), when will the error be found? Respond with assemble-time or run-time. 
- 6) If a label is referenced, but not defined, when will the error be found? Respond with assemble-time or run-time. 
- 7) If a program performing a series of divides on values in an array divides by zero, when will the error be found? Respond with assemble-time or run-time. 

### 10.6.2 Suggested Projects

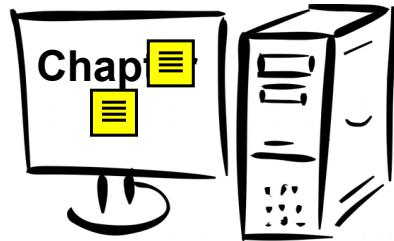
Below are some suggested projects based on this chapter.

- 1) Implement the example program to convert an integer into a string. Change the original integer to a different value. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 
- 2) Update the example program to address signed integers. This will require handling a preceding sign, “+” or “-” in the string. For example,  $-123_{10}$  ( $0xFFFFFFFF85$ ) would be “-123” with a NULL termination (total of 5 bytes). Additionally, the signed divide (IDIV, not DIV) and signed conversions (e.g., CDQ) must be used. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 

- 3) Create a program to convert a string representing a numeric value into integer. For example, given the NULL terminated string “41275” (a total of 6 bytes), convert the string into a double-word sized integer (0x0000A13B). You may assume the string and resulting integer is unsigned. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 4) Update the previous program to address strings with a preceding sign (“+” or “-”). This will require including a sign, “+” or “-” in the string. You must ensure the final string is NULL terminated. You may assume the input strings are valid. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 5) Update the previous program to convert strings into integers to include checking on the input string. Specifically, the sign must be valid and be the first character in the string, each digit must be between “0” and “9”, and the string NULL terminated. For example, the string “-321” is valid while “1+32” and “+1R3” are both invalid. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.



*Why did C++ decide not to go out with C?  
Because C has no class.*



## 11.0 Macro

An assembly language macro is a predefined set of instructions that can easily be inserted wherever needed. Once defined, the macro can be used as many times as necessary. It is useful when the same set of code must be utilized numerous times. A macro can be useful to reduce the amount of coding, streamline programs, and reduce errors from repetitive coding.

The assembler contains a powerful macro processor, which supports conditional assembly, multi-level file inclusion, and two forms of macros (single-line and multi-line), and a 'context stack' mechanism for extra macro power.

Before using a macro, it must be defined. Macro definitions should be placed in the source file **before** the data and code sections. The macro is used in the text (code) section. The following sections will present a detailed example with the definition and use.

### 11.1 Single-Line Macro

There are two key types of macros; single-line macros and multi-line macros. Each of these is described in the following sections.

Single-line macros are defined using the **%define** directive. The definitions work in a similar way to C/C++; so you can do things like:

```
%define mulby4 (x)    shl x
```

And, then use the macro by entering

```
mulby4 (r
```

in the source, which will multiply the contents to the **rax** register by 4 (via shifting **two** bits).

## 11.2 Multi-Line Mac

Multi-line macros can include a varying number of lines (including one). The multi-line macros are more useful and the following sections will focus primarily on multi-line macros.

### 11.2.1 Macro Definition

Before using a multi-line macro, it must first be defined. The general format is as follows:

```
%macro <name> <number of arguments>
    ; [body of macro]
%endmacro
```

The arguments can be referenced within the macro by %<number>, with %1 being the first argument, and %2 the second argument, and so forth.

In order to use labels, the labels within the macro must be prefixing the label name with a %%.

This will ensure that calling the same macro multiple times will use a different label each time. For example, a macro definition for the absolute value function would be as follows:

```
%macro abs 1
    cmp %1, 0
    jge %%done
    neg %1
%%done:
%endmacro
```

Refer to the sample macro program for a complete example.

### 11.2.2 Using a Macro

In order to use or “invoke” a macro, it must be placed in the code segment and referred to by name with the appropriate number of arguments.

Given a data declaration as follows:

```
qVar dq
```

Then, to invoke the “abs” macro (twice):

```
mov qVar, abs  
abs qword [qV]
```

The list file will display the code as follows (for the first invocation):

```
27 00000000 B8FDFFFFFF      mov eax, -3  
28                                     abs eax  
29 00000005 3D00000000 <1> cmp %1, 0  
30 0000000A 7D02    <1> jge %%done  
31 0000000C F7D8    <1> neg %1  
32                                     <1> %%done:
```

The macro will be copied from the definition into the code, with the appropriate arguments replaced in the body of the macro, *each* time it is used. The <1> indicates code copied from a macro definition. In both cases, the %1 argument was replaced with the given argument; **eax** in this example.

Macros use more memory, but do not require overhead for transfer of control (like functions).

### 11.3 Macro Examples

The following example program demonstrates the definition and use of a simple macro:

```
; Example Program to demonstrate a simple macro  
*****  
; Define the macro  
; called with three arguments:  
;     aver <lst>, <len>, <avg>  
  
%macro aver  
    mov eax, [%1]  
    mov ecx, dw [ %2 ] ; len  
    mov r12, [%3]  
    lea rbx, [%1]
```

```
%%sumLo
    add    [rbx+r12*4], dword [rbx+r12*4]      ; get list
    inc
    loop   %%sumL

    cdq
    idiv   dword [rbx+r12*4]
    mov    dword [%3], rax

%endma

; *****
; Data declarations

section .d
; -----
; Define consta
EXIT_SUCCESS equ 0
; success code
SYS_exit     equ 60
; code for termination

; Define Da
section .d
list1        dd -4, 5, 2, -3
len1         dd 4
avel          dd 0
; *****

list2        dd 2, 6, 3, -2, 1, 8
len2         dd 6
ave2          dd 0
; *****

section .t
global _start
_start:
; -----
; Use the macro in the program
aver    list1, len1, avel
; 1st, data se
```

```

aver    list2, len2, ave2          ; 2nd, data se[3]
;
; -----
; Done, terminate program[3]
la[3]
    mov rax, SYS_exit
    mov rdi, EXIT_SUCCESS
    sysc[3]           ; exit[3]
; succ[3]

```

In this example, the macro is invoked twice. Each time the macro is used, it is copied from the definition into the text section. As such, macros typically use more memory[3]

## 11.4 Debugging Mac[3]

The code for a macro will not be displayed in the debugger source window. When a macro is working correctly, this is very convenient. However, when debugging macros[3] the code must be viewable.

In order to see the macro code, display the machine code window (**View → Machine Code Window**). In the window, the machine code for the instructions are displayed. The step and next instructions will execute the entire macro. In order to execute the macro instructions, the **stepi** and **nexti** commands must be used.

The code, when viewed, will be the expanded code (as opposed to the original macro's definition).

## 11.5 Exercise[3]

Below are some quiz questions and suggested projects based on this chapter[3]

### 11.5.1 Quiz Questions[3]

Below are some quiz questions based on this chapter[3]

- 1) Where is the macro definition placed in the assembly language source file[3]
- 2) When a macro is invoked, how many times is the code placed in the current segment[3]
- 3) Explain why, in a macro, labels are typically preceded by a **%%** (double percent sign)[3]
- 4) Explain what might happen if the **%%** is not included on a label[3]
- 5) Is it legal to jump to a label that does not include the **%%**? If not legal, explain why. If legal, explain under what circumstances that might be used[3]

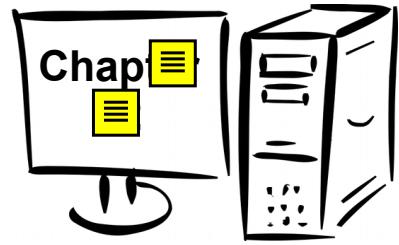
- 6) When does the macro argument substitution occur? 

### 11.5.2 Suggested Projects

Below are some suggested projects based on this chapter. 

- 1) Implement the example program for a list average macro. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 
- 2) Update the program from the previous question to include the minimum and maximum values. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 
- 3) Create a macro to update an existing list by multiplying every element by a constant. Invoke the macro at least three times of three different data sets. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 
- 4) Create a macro from the integer to ASCII conversion example from the previous chapter. Invoke the macro at least three times of three different data sets. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results. 

*Why do programmers mix up Halloween  
and Christmas?  
Because 31 Oct = 25 Dec.*



## 12.0 Function

Functions and procedures (i.e., void functions) help break-up a program into smaller parts making it easier to code, debug, and maintain. Function calls involve two main actions:

- Linking
  - Since the function can be called from multiple different places in the code, the function must be able to return to the correct place in which it was originally called.
- Argument Transmission
  - The function must be able to access parameters to operate on or to return results (i.e., access call-by-reference parameters).

The specifics of how each of these actions are accomplished is explained in the following sections.

### 12.1 Updated Linking Instructions

When writing and debugging functions, it is easier for the C compiler (either GCC or G++) to link the program as the C compiler is aware of the appropriate locations for the various C/C++ libraries.

For example, assuming that the source file is named *example.asm*, the commands to compile, assemble, link, and execute as follows:

```
yasm -g dwarf2 -f elf64 example.asm -l example.lst  
gcc -g -o example example.o
```

Note, Ubuntu 18 will require the **no-pie** option on the gcc command as shown:

```
gcc -g -no-pie -o example example.o
```

This will use the GCC compiler to call the linker, reading the *example.o* object file and creating the *example* executable file. The “-g” option includes the debugging information in the executable file in the usual manner. The file names can be changed as desired.

## 12.2 Debugger Command

When using the debugger to debug programs with functions, a review of the **step** and **next** debugger commands may be helpful.

### 12.2.1 Debugger Command, **n**

With respect to a function call, the debugger **next** command will execute the entire function and go to the next line. When debugging functions, this is useful to quickly execute the entire function and then just verify the results. It will not display any of the function code.

### 12.2.2 Debugger Command, **s**

With respect to a function call, the debugger **step** command will step into the function and go to the first line of the function code. It will display the function code. When debugging functions, this is useful to debug the function code.

## 12.3 Stack Dynamic Local Variables

In a high-level language, non-static local variables declared in a function are stack dynamic local variables by default. Some C++ texts refer to such variables as *automatics*. This means that the local variables are created by allocating space on the stack and assigning these stack locations to the variables. When the function completes, the space is recovered and reused for other purposes. This requires a small amount of additional run-time overhead, but makes a more efficient overall use of memory. If a function with a large number of local variables is never called, the memory for the local variables is never allocated. This helps reduce the overall memory footprint of the program which generally helps the overall performance of the program.

In contrast, statically declared variables are assigned memory locations for the entire execution of the program. This uses memory even if the associated function is not being executed. However, no additional run-time overhead is required to allocate the space since the space allocation has already been performed (when the program was initially loaded into memory).

## 12.4 Function Declaration

A function must be written before it can be used. Functions are located in the `code` segment. The general format is:

```
global <procName>
<procName>:
    ; function b
```

A function may be defined only once. There is no specific order required for how functions are defined. However, functions cannot be nested. A function definition should be started and ended before the next function's definition can be started.

Refer to the sample functions for examples of function declarations and usage.

## 12.5 Standard Calling Convention

To write assembly programs, a standard process for passing parameters, returning values, and allocating registers between functions is needed. If each function did these operations differently, things would quickly get very confusing and require programmers to attempt to remember for each function how to handle parameters and which registers were used. To address this, a standard process is defined and used which is typically referred to as a *standard calling convention*<sup>40</sup>. There are actually a number of different standard calling conventions. The 64-bit C calling convention, called **System V AMD64 ABI**<sup>41</sup><sup>42</sup>, is described in the remainder of this document.

This calling convention is also used for C/C++ programs by default. This means that interfacing assembly language code and C/C++ code is easily accomplished since the same calling convention is used.

It must be noted that the standard calling convention presented here applies to Linux-based operating systems. The standard calling convention for Microsoft Windows is slightly different and not presented in this text.

<sup>40</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Calling\\_convention](http://en.wikipedia.org/wiki/Calling_convention)

<sup>41</sup> For more information, refer to:

[https://en.wikipedia.org/wiki/X86\\_calling\\_conventions#System\\_V\\_AMD64](https://en.wikipedia.org/wiki/X86_calling_conventions#System_V_AMD64)

<sup>42</sup> For complete details, refer to: <https://software.intel.com/sites/default/files/article/402129/intel-64-and-ia-32-architectures-software-developer-manual-volume-2.pdf>

## 12.6 Link

The linkage is about getting to and returning from a function call correctly. There are two instructions that handle the linkage, **call <funcName>** and **ret** instructions.

The **call** transfers control to the named function, and **ret** returns control back to the calling routine.

- The **call** works by saving the address of where to return to when the function completes (referred to as the *return address*). This is accomplished by placing contents of the **rip** register on the stack. Recall that the **rip** register points to the next instruction to be executed (which is the instruction immediately after the call).
- The **ret** instruction is used in a procedure to return. The **ret** instruction pops the current top of the stack (**rsp**) into the **rip** register. Thus, the appropriate return address is restored.

Since the stack is used to support the linkage, it is important that within the function the stack must not be corrupted. Specifically, any items pushed must be popped. Pushing a value and not popping would result in that value being popped off the stack and placed in the **rip** register. This would cause the processor to attempt to execute code at that location. Most likely the invalid location will cause the process to crash.

The function calling or linkage instruction is summarized as follows:

Instruction	Explanation
<b>call &lt;funcName&gt;</b>	Calls a function. Push the 64-bit <b>rip</b> register and jump to the < <i>funcName</i> >.
Examples:	<code>call printStr</code>
<b>ret</b>	Return from a function. Pop the stack into the <b>rip</b> register, effecting a jump to the line after the call.
Examples:	

A more complete list of the instructions is located in Appendix

## 12.7 Argument Transmission

Argument transmission refers to sending information (variables, etc.) to a function and obtaining a result as appropriate for the specific function.

The standard terminology for transmitting values to a function is referred to as *call-by-value*. The standard terminology for transmitting addresses to a function is referred to as *call-by-reference*. This should be a familiar topic from a high-level language.

There are various ways to pass arguments to and/or from a function:

- Placing values in registers:
  - Easiest, but has limitations (i.e., the number of registers)
  - Used for first six integer arguments
  - Used for system calls
- Globally defined variables:
  - Generally poor practice, potentially confusing, and will not work in many cases
  - Occasionally useful in limited circumstances
- Putting values and/or addresses on stack:
  - No specific limit to count of arguments that can be passed
  - Incurs higher run-time overhead

In general, the calling routine is referred to as the *caller* and the routine being called is referred to as the *callee*.

## 12.8 Calling Convention

The function *prologue* is the code at the beginning of a function and the function *epilogue* is the code at the end of a function. The operations performed by the prologue and epilogue are generally specified by the standard calling convention and deal with stack, registers, passed arguments (if any), and stack dynamic local variables (if any).

The general idea is that the program state (i.e., contents of specific registers and the stack) are saved, the function executed, and then the state is restored. Of course, the function will often require extensive use of the registers and the stack. The prologue code helps save the state and the epilogue code restores the state.

### 12.8.1 Parameter Pass

As noted, a combination of registers and the stack is used to pass parameters to and/or from a function.

The first six integer arguments are passed in registers as follows:

Argument Num	Argument			
	64-bits	32-bits	16-bits	8-bits
1	<b>rdi</b>	<b>edi</b>	<b>di</b>	
2	<b>rsi</b>	<b>esi</b>	<b>si</b>	
3	<b>rdx</b>	<b>edx</b>	<b>dx</b>	
4	<b>rcx</b>	<b>ecx</b>	<b>cx</b>	
5	<b>r8</b>	<b>r8d</b>	<b>r8w</b>	
6	<b>r9</b>	<b>r9d</b>	<b>r9w</b>	

The seventh and any additional arguments are passed on the stack. The standard calling convention requires that, when passing arguments (values or addresses) on the stack, the arguments should be pushed in reverse order. That is “**someFunc (one, two, three, four, five, six, seven, eight, nine)**” would imply a push order of: *nine, eight, and then seven*.

For floating-point arguments, the floating-point registers **xmm0** to **xmm7** are used in that order for the first eight float arguments.

Additionally, when the function is completed, the calling routine is responsible for clearing the arguments from the stack. Instead of doing a series of pop instructions, the stack pointer, **rsp**, is adjusted as necessary to clear the arguments off the stack. Since each argument is 8 bytes, the adjustment would be adding [(number of arguments) \* 8] to the **rsp**.

For value returning functions, the result is placed in the **A** register based on the size of the value being returned.

Specifically, the values are returned as follows:

Return Value	Local Variables
byte	
word	
double-word	
quadword	
floating-point	

The **rax** register may be used in the function as needed as long as the return value is set appropriately before returning.

## 12.8.2 Register Usage

The standard calling convention specifies the usage of registers when making function calls. Specifically, some registers are expected to be preserved across a function call. That means that if a value is placed in a *preserved register* or *saved register*, and the function must use that register, the original value must be preserved by placing it on the stack, altered as needed, and then restored to its original value before returning to the calling routine. This register preservation is typically performed in the prologue and the restoration is typically performed in the epilogue.

The following table summarizes the register usage:

Register	Usage
<b>rax</b>	Return Value
<b>rbx</b>	Callee Saved
<b>rcx</b>	4 <sup>th</sup> Argument
<b>rdx</b>	3 <sup>rd</sup> Argument
<b>rsi</b>	2 <sup>nd</sup> Argument
<b>rdi</b>	1 <sup>st</sup> Argument
<b>rbp</b>	Callee Saved
<b>rsp</b>	Stack Pointer
<b>r8</b>	5 <sup>th</sup> Argument
<b>r9</b>	6 <sup>th</sup> Argument
<b>r10</b>	Temporary

<b>r11</b>	Temporary
<b>r12</b>	Callee Saved
<b>r13</b>	Callee Saved
<b>r14</b>	Callee Saved
<b>r15</b>	Callee Saved

The temporary registers (**r10** and **r11**) and the argument registers (**rdi**, **rsi**, **rdx**, **rcx**, **r8**, and **r9**) are not preserved across a function call. This means that any of these registers may be used in the function without the need to preserve the original value.

Additionally, none of the floating-point registers are preserved across a function call. Refer to Chapter 18 for more information regarding floating-point operations.

### 12.8.3 Call Frame

The items on the stack as part of a function call are referred to as a *call frame* (also referred to as an *activation record* or *stack frame*). Based on the standard calling convention, the items on the stack, if any, will be in a specific general format.

The possible items in the call frame include:

- Return address (required)
- Preserved registers (if any)
- Passed arguments (if any)
- Stack dynamic local variables (if any)

Other items may be placed in the call frame such as static links for dynamically scoped languages. Such topics are outside the scope of this text and will not be discussed here.

For some functions, a full call frame may not be required. For example, if the function

- Is a leaf function (i.e., does not call another function)
- Passes its arguments only in registers (i.e., does not use the stack)
- Does not alter any of the saved registers
- Does not require stack-based local variables

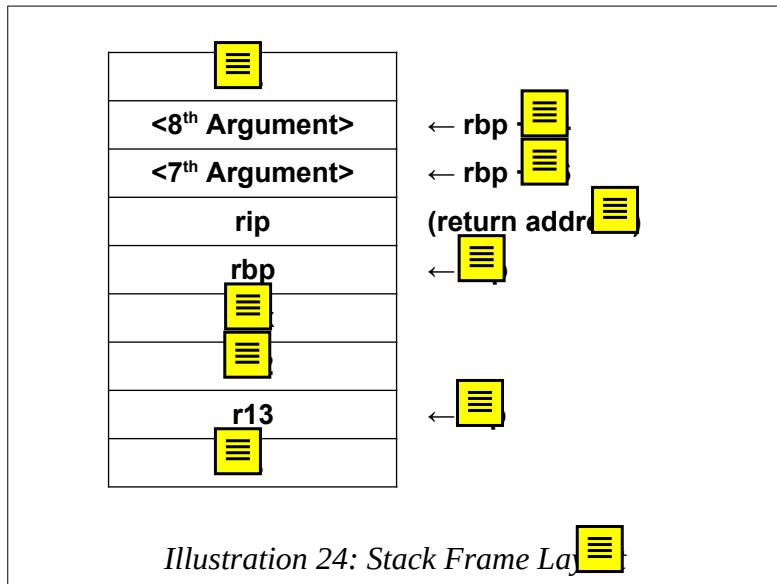
This can occur for simpler, smaller leaf functions. However, if any of these conditions is not true, a full call frame is required.

For more non-leaf or more complex functions, a more complete call frame is required.

The standard calling convention does not explicitly require use of the frame pointer register, **rbp**. Compilers are allowed to optimize the call frame and not use the frame pointer. To simplify and clarify accessing stack-based arguments (if any) and stack dynamic local variables, this text will utilize the frame pointer register. This is similar to how many other architectures use a frame pointer register.

As such, if there are any stack-based arguments or any local variables needed within a function, the frame pointer register, **rbp**, should be pushed and then set pointing to itself. As additional pushes and pops are performed (thus changing **rsp**), the **rbp** register will remain unchanged. This allows the **rbp** register to be used as a reference to access arguments passed on the stack (if any) or stack dynamic local variables (if any).

For example, assuming a function call has eight (8) arguments and assuming the function uses **rbx**, **r12**, and **r13** registers (and thus must be pushed), the call frame would be as follows:

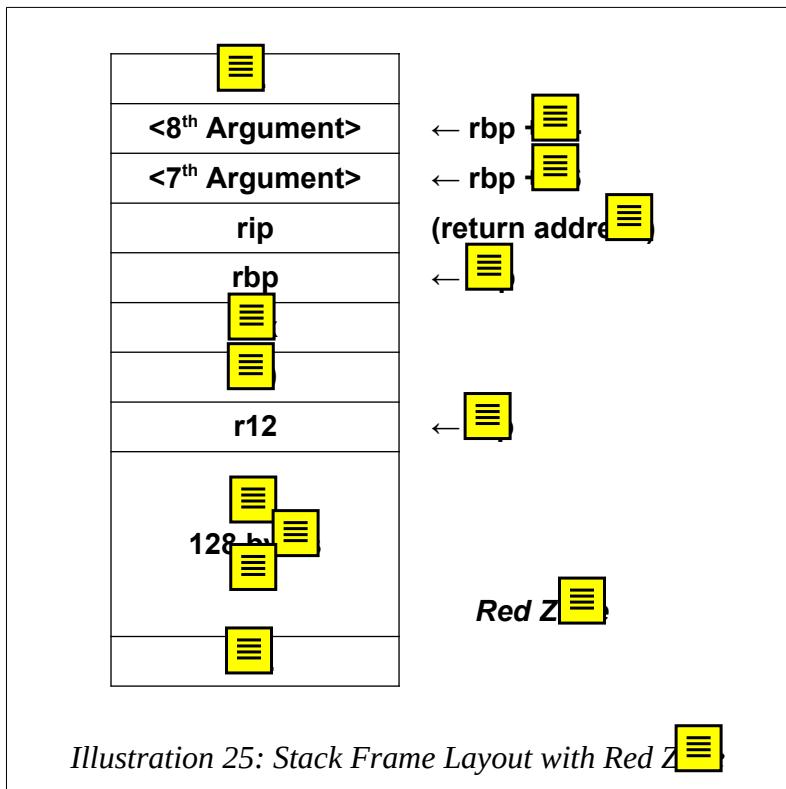


The stack-based arguments are accessed relative to the **rbp**. Each item push is a quadword which uses 8 bytes. For example, **[rbp+16]** is the location of the first passed argument (7<sup>th</sup> integer argument) and **[rbp+24]** is the location of the second passed argument (8<sup>th</sup> integer argument).

In addition, the call frame would contain the assigned locations of local variables (if any). The section on local variables details the specifics regarding allocating and using local variables.

**12.8.3.1 Red Z**

In the Linux standard calling convention, the first 128-bytes after the stack pointer, `rsp`, are reserved. For example, extending the previous example, the call frame would be as follows:



This red zone may be used by the function without any adjustment to the stack pointer. The purpose is to allow compiler optimizations for the allocation of local variables. This does not directly impact programs written directly in assembly language.

**12.9 Example, Statistical Function 1 (l)**

This simple example will demonstrate calling a simple void function to find the sum and average of an array of numbers. The High-Level Language (HLL) call for C/C++ is as follows:

```
stats1(arr, len, sum, av)
```

As per the C/C++ convention, the array, ***arr***, is call-by-reference and the length, ***len***, is call-by-value. The arguments for ***sum*** and ***ave*** are both call-by-reference (since there are no values as yet). For this example, the array ***arr***, ***sum***, and ***ave*** variables are all signed double-word integers. Of course, in context, the ***len*** must be unsigned.

### 12.9.1 Call

In this case, there are 4 arguments, and all arguments are passed in registers in accordance with the standard calling convention. The assembly language code in the calling routine for the call to the stats function would be as follows:

```
; stats1(arr, len, sum, ave);
mov    rcx, ave           ; 4th arg, addr of [list]
mov    rdx, sum            ; 3rd arg, addr of [list]
mov    esi, dword [len]    ; 2nd arg, value of [list]
mov    rdi, arr             ; 1st arg, addr of [list]
call   sta [list]
```

There is no specific required order for setting the argument registers. This example sets them in reverse order in preparation for the next, extended example.

Note, the setting of the **esi** register also sets the upper-order double-word to zero, thus ensuring the **rsi** register is set appropriately for this specific usage since length is unsigned.

No return value is provided by this void routine. If the function was a value returning function, the value returned would be in the **A** register (of appropriate size).

### 12.9.2 Call

The function being called, the callee, must perform the prologue and epilogue operations (as specified by the standard calling convention) before and after the code to perform the function goal. For this example, the function must perform the summation of values in the array, compute the integer average, return the sum and average value.

The following code implements the **stats1** example.

```
; Simple example function to find and return
; the sum and average of an array.

; HLL call:
; stats1(arr, len, sum, ave)
; -----
; Arguments:
; arr, address - [list]
; len, dword value - [list]
```

```

;    sum, address - [ ]  

;    ave, address - [ ]  
  

global stats1  

stats1:  

    push    r12           ; prolo [ ]  

    mov     r12, 0         ; counter/initial value [ ]  

    mov     rax, 0         ; running total [ ]  

sumLoop:  

    add     rax, dword [rdi+r12*4] ; sum += arr[ ]  

    inc     r12            ; sumLoop [ ]  

    cmp     r12, [ ]        ; sumLoop [ ]  

    jl      sumLoop        ; sumLoop [ ]  
  

    mov     dword [rdx], eax ; return [ ]  
  

    cdq  

    idiv    esi            ; compute average [ ]  

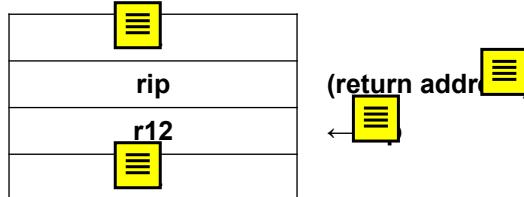
    mov     dword [rcx], eax ; return [ ]  
  

    pop    r12             ; epilo [ ]  


```

The choice of the **r12** register is arbitrary, however a 'saved register' was selected.

The call frame for this function would be as follows:



The minimal use of the stack helps reduce the function call run-time overhead.

## 12.10 Example, Statistical Function2 (non-local)

This extended example will demonstrate calling a simple void function to find the minimum, median, maximum, sum and average of an array of numbers.

The High-Level Language (HLL) call for C/C++ is as follows:

```
stats2(arr, len, min, med1, med2, max, sum, ave)
```

For this example, it is assumed that the array is sorted in ascending order. Additionally, for this example, the median will be the middle value. For an even length list, there are two middle values, **med1** and **med2**, both of which are returned. For an odd length list, the single middle value is returned in both **med1** and **med2**.

As per the C/C++ convention, the array, **arr**, is call-by-reference and the length, **len**, is call-by-value. The arguments for **min**, **med1**, **med2**, **max**, **sum**, and **ave** are all call-by-reference (since there are no values as yet). For this example, the array **arr**, **min**, **med1**, **med2**, **max**, **sum**, and **ave** variables are all signed double-word integers. Of course, in context, the **len** must be unsigned.

### 12.10.1 Calling

In this case, there are 8 arguments and only the first six can be passed in registers. The last two arguments are passed on the stack. The assembly language code in the calling routine for the call to the stats function would be as follows:

```
; stats2(arr, len, min, med1, med2, max, sum, ave);
push    ave                      ; 8th arg, add of [esi]
push    sum                      ; 7th arg, add of [esi+4]
mov     r9, max                  ; 6th arg, add of [esi+8]
mov     r8, med2                ; 5th arg, add of [esi+12]
mov     rcx, med1                ; 4th arg, add of [esi+16]
mov     rdx, min                  ; 3rd arg, addr of [esi+20]
mov     esi, dword [len]          ; 2nd arg, value of [esi+24]
mov     rdi, arr                  ; 1st arg, addr of [esi+28]
call    stats
add    rsp, 16                   ; clear passed arguments
```

The 7<sup>th</sup> and 8<sup>th</sup> arguments are passed on the stack and pushed in reverse order in accordance with the standard calling convention. After the function is completed, the arguments are cleared from the stack by adjusting the stack point register (**rsp**). Since two arguments, 8 bytes each, were passed on the stack, 16 is added to the stack point.

Note, the setting of the **esi** register also sets the upper-order double-word to zero, thus ensuring the **rsi** register is set appropriately for this specific usage since length is unsigned.

No return value is provided by this void routine. If the function was a value returning function, the value returned would be in the **A** register.

## 12.10.2 Call

The function being called, the callee, must perform the prologue and epilogue operations (as specified by the standard calling convention). Of course, the function must perform the summation of values in the array, find the minimum, medians, and maximum, compute the average, return all the values.

When call-by-reference arguments are passed on the stack, two steps are required to return the value.

- Get the address from the stack
- Use that address to return the value

A common error is to attempt to return a value to a stack-based location in a single step, which will not change the referenced variable. For example, assuming the double-word value to be returned is in the **eax** register and the 7<sup>th</sup> argument is call-by-reference and where the **eax** value is to be returned, the appropriate code would be as follows:

```
mov     r12, qword [rbp+16] ; Step 1
mov     dword [r12], eax    ; Step 2
```

These steps cannot be combined into a single step. The following code

```
mov     dword [rbp+16], eax
```

Would overwrite the address passed on the stack and not change the reference variable.

The following code implements the **stats2** example.

```
; Simple example function to find and return the minimum,
; maximum, sum, medians, and average of an array.
; -----
; HLL call:
; stats2(arr, len, min, med1, med2, max, sum, ave);

; Arguments:
; arr, address -
; len, dword value
; min, address -
; med1, address -
; med2, address -
; max, address -
; sum, address - stack (rbp+16)
```

```

; ave, address - stack (rbp+)

global stats2
stats2:
    push    rbp
    mov     rbp, 
    push    
    ; prolo   

;  

;  

; Get min and m   

    mov     eax, dword [rdi]
    mov     dword [rdx], eax  

    ; get   

    ; return   

    mov     r12, rsi
    dec     r12
    mov     eax, dword [rdi+r12*4]
    mov     dword [r9], eax  

    ; get   

    ; set   

    ; get   

    ; return   

;  

;  

; Get medi   

    mov     rax, 
    mov     rdx, 
    mov     r12, 
    div     r12  

    ; rax = length   

    cmp     rdx, 0
    je      evenLen   

    ; even/odd leng   

    mov     r12d, dword [rdi+rax*4]
    mov     dword [rcx], r12d
    mov     dword [r8], r12d
    jmp     medDo   

evenLen 
    mov     r12d, dword [rdi+rax*4]
    mov     dword [r8], r12d
    dec     r12d
    mov     r12d, dword [rdi+rax*4]
    mov     dword [rcx], r12d  

medDo   

;  

;  

; Find 
    ; get arr[] 
    ; return m   

    ; get arr[] 
    ; return m 
```

```

    mov      r12, 0          ; counter/iteration
    mov      rax, 0          ; running total

sumLoop:
    add      eax, dword [rdi+r12*4]   ; sum += arr[i]
    inc      r12, 1
    cmp      r12, 10           ; i <= 10?
    jl       sumLoop

    mov      r12, qword [rbp+16]     ; get sum
    mov      dword [r12], eax        ; return sum

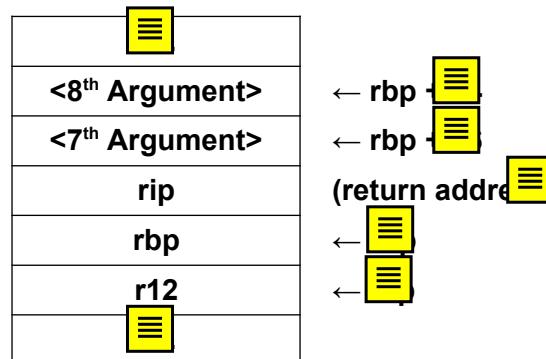
; -----
; Calculate average

    cdq
    idiv    rsi
    mov      r12, qword [rbp+24]     ; average = sum / n
    mov      dword [r12], eax        ; get average
    ; return average

    pop      r12
    ; epilogue

```

The choice of the registers is arbitrary with the bounds of the calling convention  
The call frame for this function would be as follows:



In this example, the preserved registers, **rbp** and then **r12**, are pushed. When popped, they must be popped in the exact reverse order **r12** and then **rbp** in order to correctly restore their original values.



## 12.11 Stack-Based Local Variables

If local variables are required, they are allocated on the stack. By adjusting the **rsp** register, additional memory is allocated on the stack for locals. As such, when the function is completed, the memory used for the stack-based local variables is released (and no longer uses memory).



Further expanding the previous example, if we assume all array values are between 0 and 99, and we wish to find the mode (number that occurs the most often), a single double-word variable **count** and a one hundred (100) element local double-word array **tmpArr[100]** might be used.



As before, the frame register, **rbp**, is pushed on the stack and set pointing to itself. The frame register plus an appropriate offset will allow accessing any arguments passed on the stack. For example, **rbp+16** is the location of the first stack-based argument integer argument.



After the frame register is pushed, an adjustment to the stack pointer register, **rsp**, is made to allocate space for the local variables, a 100-element array in this example. Since the count variable is a one double-word, 4-bytes is needed. The temporary array is 100 double-word elements, 400 bytes is required. Thus, a total of 404 bytes is required. Since the stack is implemented growing downward in memory, the 404 bytes is subtracted from the stack pointer register.



Then any saved registers, **rbx** and **r12** in this example, are pushed on the stack.



When leaving the function, the saved registers and then the locals must be cleared from the stack. The preferred method of doing this is to pop the saved registers and then top copy the **rbp** register into the **rsp** register, thus ensuring the **rsp** register points to the correct place on the stack.



```
mov    rsp, [rbp]
```

This is generally better than adding the offset back to the stack since allocated space may be altered as needed without also requiring adjustments to the epilogue code.

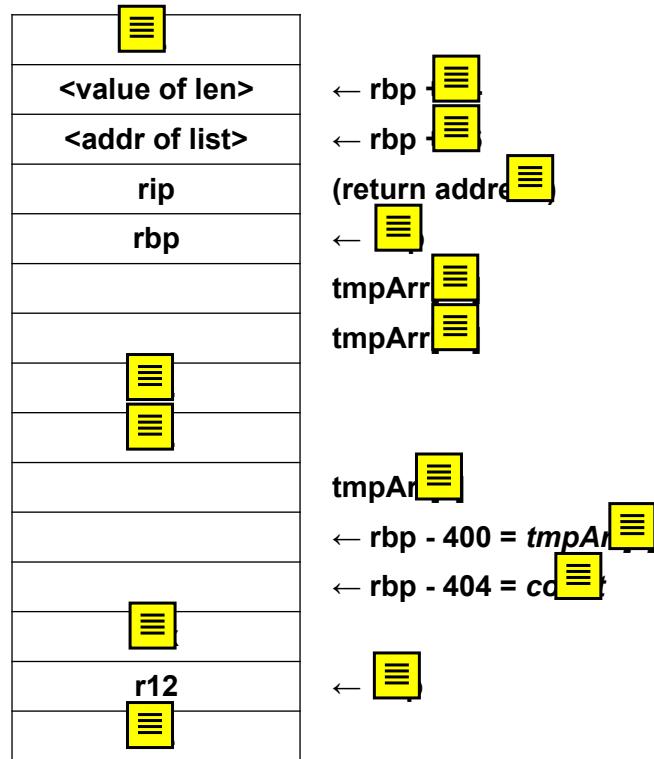


It should be clear that variables allocated in this manner are uninitialized. Should the function require the variables to be initialized, possibly to 0, such initializations must be explicitly performed.



For this example, the call frame would be formatted as follows:





The layout and order of the local variables within the allocated 404 bytes is arbitrary.

For example, the updated prologue code for this expanded example would

```
push    rbp ; prologue  
mov     rbp, [rbp] ; allocate loc  
sub    rbp, 404  
push    [rbp] ;  
push    [rbp] ;
```

The local variables can be accessed relative to the frame pointer register, **rbp**. For example, to initialize the count variable, now allocated to **rbp-404**, the following instruction could be used:

mov dword [rbp-404] 

To access the *tmpArr*, the starting address must be obtained which can be performed with the lea instruction. For example,

lea rbx, dword [rbp-4]

Which will set the appropriate stack address in the **rbx** register where **rbx** was chosen arbitrarily. The **dword** qualifier in this example is not required, and may be misleading, since addresses are always 64-bits (on a 64-bit architecture). Once set as above, the **tmpArr** starting address in **rbx** is used in the usual manner.

For example, a small incomplete function code fragment demonstrating the accessing of stack-based local variables is as follows:

```
; -----  
; Example function  
  
global expFunc  
expFunc:  
    push rbp  
    mov rbp, [rbp - 404] ; prolo  
    sub rbp, rbp, 404 ; allocate loc  
    push rbp  
    push rbp  
  
; -----  
; Initialize count local variable to 0  
    mov dword [rbp - 404] 0  
  
; -----  
; Increment count variable (for example)  
    inc dword [rbp - 404] ; coun  
  
; -----  
; Loop to initialize tmpArr to all 0  
zeroLoop:  
    lea rbx, dword [rbp - 400] ; tmpArr a  
    mov r12, 0 ; in  
    mov dword [rbx + r12 * 4], 0 ; tmpArr[index]  
    inc r12  
    cmp r12, zeroL  
    jl zeroL  
  
; -----  
; Done, restore all and return to calling routine  
    pop r12 ; epilo
```

```

pop    [ ]      ; clear loc [ ]
mov    [ ] , rbp
[ ]      [ ]      ; clear loc [ ]

```

Note, this example function focuses only on how stack-based local variables are accessed and does not perform anything useful.

## 12.12 Summary

This section presents a brief summary of the standard calling convention requirements which are as follows:

### Caller Operations

- The first six integer arguments are passed in registers:
  - **rdi, rsi, rdx, rcx, r8**
- The 7<sup>th</sup> and on arguments are passed on the stack-based frame:
  - Pushes the arguments on the stack in reverse order (right to left, so that the first stack argument specified in the function call is pushed last)
  - Pushed arguments are passed as quadwords
- The caller executes a **call** instruction to pass control to the function (call)
- Stack-based arguments are cleared from the stack:
  - **add rsp, <argCount**

### Callee Operations

- Function Prologue:
  - If arguments are passed on the stack, the callee must save **rbp** to the slot above it and move the value of **rsp** into **rbp**. This allows the callee to use **rbp** as a frame pointer to access arguments on the stack in a uniform manner.
    - The callee may then access its parameters relative to **rbp**. The quadword at **[rbp]** holds the previous value of **rbp** as it was pushed; the next quadword, at **[rbp+8]**, holds the return address, pushed by the **call**. The parameters start after that, at **[rbp+16]**.

- If local variables are needed, the callee decreases **rsp** further to allow space on the stack for the local variables. The local variables are accessible at negative offsets from **rbp**.  
[ANSWER]
- The callee, if it wishes to return a value to the caller, should leave the value in **al**, **ax**, **eax**, **rax**, depending on the size of the value being returned.  
  - A floating-point result is returned in **xmm0**
- If altered, registers **rbx**, **r12**, **r13**, **r14**, **r15** and **rbp** must be saved on the stack.  
[ANSWER]
- Function Execution  
[ANSWER]
  - The function code is executed.  
[ANSWER]
- Function Epilogue  
[ANSWER]
  - Restores any pushed registers.  
[ANSWER]
  - If local variables were used, the callee restores **rsp** from **rbp** to clear stack-based local variables.  
[ANSWER]
  - The callee restores (i.e., pops) the previous value of **rbp**.  
[ANSWER]
  - The call returns via **ret** instruction (return).  
[ANSWER]

Refer to the sample functions to see specific examples of the calling convention.  
[ANSWER]

## 12.13 Exercises [ANSWER]

Below are some quiz questions and suggested projects based on this chapter.  
[ANSWER]

### 12.13.1 Quiz Questions [ANSWER]

Below are some quiz questions based on this chapter.  
[ANSWER]

- 1) What are the two main actions of a function call?  
[ANSWER]
- 2) What are the two instructions that implement *linkage*?  
[ANSWER]
- 3) When arguments are passed using *values*, it is referred to as  
[ANSWER]
- 4) When arguments are passed using *addresses*, it is referred to as  
[ANSWER]
- 5) If a function is called fifteen (15) times, how many times is the code placed in memory by the assembler?  
[ANSWER]
- 6) What happens during the execution of a **call** instruction (two things)?  
[ANSWER]
- 7) According to the standard calling convention, as discussed in class, what is the purpose of the initial pushes and final pops within most procedures?  
[ANSWER]

- 8) If there are six (6) 64-bit integer arguments passed to a function, where specifically should each of the arguments be passed?
- 9) If there are six (6) 32-bit integer arguments passed to a function, where specifically should each of the arguments be passed?
- 10) What does it mean when a register is designated as temporary?
- 11) Name two temporary registers.
- 12) What is the name for the set of items placed on the stack as part of a function?
- 13) What does it mean when a function is referred to as a *leaf function*?
- 14) What is the purpose of the `add rsp, <immediate>` after the `sub` statement?
- 15) If **three** arguments are passed on the stack, what is the value for `rsp` after `<immediate>`?
- 16) If there are seven (7) arguments passed to a function, and the function it calls pushes the **rbp**, **rbx**, and **r12** registers (in that order), what is the correct offset of the stack-based argument when using the standard calling convention?
- 17) What, if any, is the limiting factor for how many times a function can be called?
- 18) If a function must return a result for the variable **sum**, how should the variable be passed (call-by-reference or call-by-value)?
- 19) If there are eight (8) arguments passed to a function, and the function it calls pushes the **rbp**, **rbx**, and **r12** registers (in that order), what are the correct offsets for each of the two stack-based arguments (7<sup>th</sup> and 8<sup>th</sup>) when using the standard calling convention?
- 20) What is the advantage of using stack dynamic local variables (as opposed to using all global variables)?

### 12.13.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Create a main and implement the **stats1** example function. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

- 2) Create a main and implement the **stats2** example function. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 3) Create a main program and a function that will sort a list of numbers in ascending order. Use the following selection<sup>43</sup> sort algorithm.

```

be[=]
  for i = 0 to len[=]
    small = arr(i)
    index = i
    for j = i to len[=]
      if ( arr(j) < small ) t[=]
        small = arr(i)
        index = j
      end[=]
    end_for[=]
    arr(index) = arr(i)
    arr(i) = small
  end[=]
end_be[=]

```

The main should call the function on at least three different data sets. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

- 4) Update the program from the previous question to add a stats function that finds the minimum, median, maximum, sum, and average for the sorted list. The stats function should be called after the sort function to make the minimum and maximum easier to find. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 5) Update the program from the previous question to add an integer square root function and a standard deviation function. To estimate the square root of a number, use the following algorithm:

$$iSqrt_{est} = iNum$$

$$iSqrt_{est} := \left( \frac{iNum}{iSqrt_{est}} \right) + iSqrt_{est}$$

iterate 50 times

---

<sup>43</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Selection\\_sort](http://en.wikipedia.org/wiki/Selection_sort)

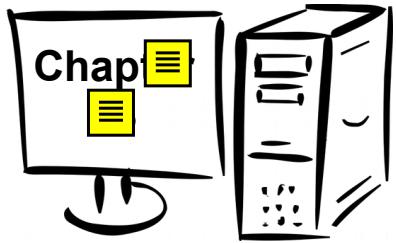
The formula for standard deviation is as follows:

$$iStandardDeviation = \sqrt{\frac{\sum_{i=1}^n (list[i] - average)^2}{n}}$$

Note, perform the summation and division using integer values. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

- 6) Convert the integer to ASCII macro from the previous chapter into a function. The function should convert a signed integer into a right-justified string of a given length. This will require including any leading blanks, a sign ("+" or "-"), the digits, and the NULL. The function should accept the value for the integer and the address of where to place the NULL terminated string, and the value of the maximum string length - in that order. Develop a main program to call the function on a series of different integers. The main should include the appropriate data declarations. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.
- 7) Create a function to convert an ASCII string representing a number into an integer. The function should read the string and perform appropriate error checking. If there is an error, the function should return FALSE (a defined constant set to 0). If the string is valid, the function should convert the string into an integer. If the conversion is successful, the function should return TRUE (a defined constant set to 1). Develop a main program to call the function on a series of different integers. The main should include the appropriate data declarations and applicable the constants. Use the debugger to execute the program and display the final results. Create a debugger input file to show the results.

*Linux is basically a simple operating system, but you have to be a genius to understand the simplicity.*



## 13.0 System Services

There are many operations that an application program must use the operating system to perform. Such operations include console output, keyboard input, file services (open, read, write, close, etc.), obtaining the time or date, requesting memory allocation, and many others.

Accessing system services is how the application requests that the operating system perform some specific operation (on behalf of the process). More specifically, the *system call* is the interface between an executing process and the operating system.

This section provides an explanation of how to use some basic system service calls. More information on additional system service calls is located in Appendix C, System Service Calls.

### 13.1 Calling System Services

A system service call is logically similar to calling a function, where the function code is located within the operating system. The function may require privileges to operate which is why control must be transferred to the operating system.

When calling system services, arguments are placed in the standard argument registers. System services do not typically use stack-based arguments. This limits the arguments of a system services to six (6), which does not present a significant limitation.

To call a system service, the first step is to determine which system service is desired. There are many system services (see Appendix C). The general process is that the system service call code is placed in the `rax` register. The call code is a number that has been assigned for the specific system service being requested. These are assigned as part of the operating system and cannot be changed by application programs. To simplify the process, this text will define a very small subset of system service `call` codes to a set of constants. For this text, and the associated examples, the subse

system call code constants are defined and shown in the source file to help provide complete clarity for new assembly language programmers. For more experienced programmers, typically developing larger or more complex programs, a complete list of constants is in a file and included into the source file.

If any are needed, the arguments for system services are placed in the **rdi**, **rsi**, **rdx**, **r10**, **r8**, and **r9** registers (in that order). The following table shows the argument locations which are consistent with the standard calling convention.

Register	Usage
<b>rax</b>	Call code (see table)
<b>rdi</b>	1st argument (if needed)
<b>rsi</b>	2nd argument (if needed)
<b>rdx</b>	3rd argument (if needed)
<b>r10</b>	4th argument (if needed)
<b>r8</b>	5th argument (if needed)
<b>r9</b>	6th argument (if needed)

This is very similar to the standard calling convention for function calls, however the argument, if needed, uses the **r10** register.

Each system call will use a different number of arguments (from none up to 6). However, the system service call code is always required.

After the call code and any arguments are set, the **syscall** instruction is executed. The **syscall** instruction will pause the current process and transfer control to the operating system which will attempt to perform the service specified in the **rax** register. When the system service returns, the process will be resumed.

## 13.2 Newline Character

As a refresher, in the context of output, a newline means move the cursor to the start of the next line. In many languages, including C, it is often noted as “\n” as part of a string. C++ uses **endl** in the context of a **cout** statement. For example, “Hello World 1” and “Hello\nWorld 2” would be displayed as follows:

```
Hello World 1
Hello
World 2
```

Nothing is displayed for the newline, but the cursor is moved to the start of the next line as shown.

In Unix/Linux systems, the linefeed, abbreviated LF with an ASCII value of 10 (or 0x0A), is used as the newline character. In Windows systems, the newline is carriage return, abbreviated as CR with an ASCII value 13 (or 0x0D) followed by the LF. The LF is used in the code examples in the text.

The reader may have seen instances where a text file is downloaded from a web page and displayed using older versions Windows Notepad (pre-Windows 10) where all the formatting is lost and it looks like the text is one very long line. This is typically due to a Unix/Linux formatted file, which uses only LF's, being displayed with a Windows utility that expects CR/LF pairs and does not display correctly when only LF's are found. Other Windows software, like Notepad++ (open source text editor) will recognize and handle the different newline formats and display correctly.

### 13.3 Console Out

The system service to output characters to the console is the system write (SYS\_write). Like a high-level language characters are written to standard out (STDOUT) which is the console. The STDOUT is the default file descriptor for the console. The file descriptor is already opened and available for use in programs (assembly and high-level languages).

The arguments for the write system service are as follows:

Register	SYS_write
rax	Call code = SYS_write
rdi	Output location, STDOUT
rsi	Address of characters to output
rdx	Number of characters to output

Assuming the following declarations:

```
STDOUT      equ      1          ; standard out
SYS_write   equ      1          ; call code for write
msg         db       "Hello Wor
msgLen     dq        12
```

For example, to output “Hello World” (it’s traditional) to the console, the system write (SYS\_write) would be used. The code would be as follows:

```

mov    rax, SYS...  

mov    rdi, STD...  

mov    rsi, msg  

        ; msg addr  

mov    rdx, qword [msgLen]    ; length va...  

sysc

```

Refer to the next section for a complete program to display the above message. It should be noted that the operating system does not check if the string is valid.

### 13.3.1 Example, Console Out

This example is a complete program to output some strings to the console. In this example, one string includes new line and the other does not.

```

; Example program to demonstrate console output.  

; This example will send some messages to the scre...  

; *****  

section .d...  

; -----  

; Define standard constan...  

LF          equ   10      ; line f...  

NULL        equ   0       ; end of str...  

TRUE        equ   1       ;  

FALSE       equ   0       ;  

EXIT_SUCCESS equ   0       ; success c...  

STDIN       equ   0       ; standard in...  

STDOUT      equ   1       ; standard out...  

STDERR      equ   2       ; standard er...  

SYS_read    equ   0       ; r...  

SYS_write   equ   1       ; wr...  

SYS_open    equ   2       ; file o...  

SYS_close   equ   3       ; file cl...  

SYS_fork    equ   57      ; f...  

SYS_exit    equ   60      ; termin...  

SYS_creat   equ   85      ; file open/cre...

```

```

SYS_time      equ     201          ; get t[...]
; -----
; Define some strings[...]
message1      db      "Hello World.", LF, N[...]
message2      db      "Enter Answer: ", N[...]
newLine        db      LF, N[...]
; -----
section       .t[...]
global _start
_start:
; -----
; Display first message[...]
    mov     rdi, message1[...]
    call   printStr[...]
; -----
; Display second message and then newLine[...]
    mov     rdi, message2[...]
    call   printStr[...]
    mov     rdi, newLine[...]
    call   printStr[...]
; -----
; Example program do[...]
exampleDo:
    mov     rax, SYS_exit[...]
    mov     rdi, EXIT_SUCC[...]
    sysc[...]
; *****
; Generic function to display a string to the screen.
; String must be NULL terminated.
; Algorithm:
; Count characters in string (excluding NULL)
; Use syscall to output character[...]

```

## Chapter 13.0 ◀ System Serv

```
; Arguments:  
; 1) address, str  
; Returns:  
; nothing
```

```
global printString  
printString:  
    push    [rbx]  
  
; -----  
; Count characters in str  
strCountLoop:  
    mov     rbx, [rdx]  
    mov     rdx, [rbx]  
    cmp     byte [rbx], N  
    je      strCountDone  
    inc     rbx  
    inc     rdx  
    jmp     strCountLoop  
strCountDone:  
    cmp     rdx, [rbx]  
    je      prtDone  
  
; -----  
; Call OS to output str  
    mov     rax, SYS_write          ; system code for write  
    mov     rsi, rdi               ; address of chars to write  
    mov     rdi, STDOUT            ; standard out  
    syscall                      ; RDX=count to write, set ab  
                                ; system call
```

```
; -----  
; String printed, return to calling routine  
prtDone:  
    pop    [rbx]
```

The output would be as follo

Hello World  
Enter Answe

The newline (LF) was provided as part of the first string (**message1**) thus placing the cursor on the start of the next line. The second message would leave the cursor on the same line which would be appropriate for reading input from the user (which is not part of this example). A final newline is printed since no actual input is obtained in this example.

The additional, unused constants are included for refere

## 13.4 Console In

The system service to read characters from the console is the system read (SYS\_read). Like a high-level language, for the console, characters are read from standard input (STDIN). The STDIN is the default file descriptor for reading characters from the keyboard. The file descriptor is already opened and available for use in program (assembly and high-level languages).

Reading characters interactively from the keyboard presents an additional complication. When using the system service to read from the keyboard, much like the write system service, the number of characters to read is required. Of course, we will need to declare an appropriate amount of space to store the characters being read. If we request 10 characters to read and the user types more than 10, the additional characters will be lost, which is not a significant problem. If the user types less than 10 characters, for example 5 characters, all five characters will be read plus the newline (LF) for a total of six characters.

A problem arises if input is redirected from a file. If we request 10 characters, and there are 5 characters on the first line and more on the second line, we will get the six characters from the first line (5 characters plus the newline) and the first four characters from the next line for the total of 10. This is undesirable.

To address this, for interactively reading input, we will read one character at a time until a LF (the Enter key) is read. Each character will be read and then stored, one at a time, in an appropriately sized array.

The arguments for the read system service are as follo

Register	SYS_
<b>rax</b>	Call code = SYS_read
<b>rdi</b>	Input location, STDIN

<b>rsi</b>	Address of where to store characters	
<b>rdx</b>	Number of characters to	

Assuming the following declaration:

```
STDIN      equ      0          ; standard input
SYS_read   equ      0          ; call code for read
inChar     db      ?
```

For example, to read a single character from the keyboard, the system read (SYS\_read) would be used. The code would be as follows:

```
mov      rax, SYS_read
mov      rdi, STDIN
mov      rsi, inChar
mov      rdx, 1
sysc
```

Refer to the next section for a complete program to read characters from the keyboard.

### 13.4.1 Example, Console In

The example is a complete program to read a line of 50 characters from the keyboard. Since space for the newline (LF) along with a final NULL termination is included, an input array allowing 52 bytes would be required.

This example will read up to 50 characters from the user and then echo the input back to the console to verify that the input was read correctly.

```
; Example program to demonstrate console output.
; This example will send some messages to the screen
; ****
section    .data
;
; -----
; Define standard constants
;
LF          equ      10          ; line feed
NULL        equ      0           ; end of string
```

```

TRUE      equ   
FALSE     equ   
EXIT_SUCCESS equ  0          ; success code 
STDIN      equ   0          ; standard input 
STDOUT     equ   1          ; standard output 
STDERR     equ   2          ; standard error 
SYS_read    equ   0          ; read system call 
SYS_write   equ   1          ; write system call 
SYS_open    equ   2          ; file open system call 
SYS_close   equ   3          ; file close system call 
SYS_fork    equ   57         ; fork system call 
SYS_exit    equ   60         ; terminate system call 
SYS_creat   equ   85         ; file open/create system call 
SYS_time   equ   201        ; get time system call 
; -----
; Define some strings 
STRLEN     equ   
pmpt       db    "Enter Text: ", NUL 
newLine    db    LF, NUL 
section .bss
chr        resb 
inLine     resb STRLEN+2      ; total of 
; -----
section .text
global _start
_start:

; -----
; Display prompt 
    mov     rdi, pmpt 
    call   printStr 
; -----
; Read characters from user (one at a time) 

```

## Chapter 13.0 ◀ System Serv

```
        mov      rbx, inLine  
        mov      r12, 0  
readCharacter:  
        mov      rax, SYS_read  
        mov      rdi, STDIN  
        lea      rsi, byte [chr]  
        mov      rdx, 1  
        syscall  
  
        mov      al, byte [chr]  
        cmp      al, LF  
        je      readDo  
  
        inc      r12  
        cmp      r12, STRLEN  
        jae      readCharacters  
  
        mov      byte [rbx], al  
        inc      rbx  
  
        jmp      readCharacter  
readDo:  
        mov      byte [rbx], NULL  
; add NULL terminator  
  
; -----  
; Output the line to verify successful read  
        mov      rdi, inLine  
        call     printStr  
  
; -----  
; Example do  
exampleDo:  
        mov      rax, SYS_execve  
        mov      rdi, EXIT_SUCCESS  
        syscall  
  
; ****  
; Generic procedure to display a string to the screen.  
; String must be NULL terminated.  
; Algorithm:  
;     Count characters in string (excluding NULL)
```

```

;      Use syscall to output character
;
; Arguments:
;   1) address, str
;
; Returns:
;   nothing

global printString
printString:
    push    ; [rbx]
    push    ; [rdx]

; -----
; Count characters in string
    mov     rbx, [rdx] ; rbx = address of string
    mov     rdx, [rbx] ; rdx = count of characters

strCountLoop:
    cmp     byte [rbx], NUL ; if character is NUL, exit loop
    je      strCountDone
    inc     rbx
    inc     rdx
    jmp     strCountLoop

strCountDone:
    cmp     rdx, [rbx] ; if count = address, exit loop
    je      prtDone

; -----
; Call OS to output string
    mov     rax, SYS_write      ; system code for write
    mov     rsi, rdi             ; address of char's to write
    mov     rdi, STDOUT          ; standard out
                                ; RDX=count to write, set above
    syscall                     ; system call

; -----
; String printed, return to calling routine
prtDone:
    pop    ; [rbx]
    pop    ; [rdx]

```



If we were to completely stop reading at 50 (STRLEN) characters and the user enters more characters, the characters might cause input errors for successive read operations. To address any extra characters the user might enter, the extra characters are read from the keyboard but not placed in the input buffer (*inLine* above). This ensures that the extra input is removed from the input stream and but does not overrun the array.



The additional, unused constants are included for reference.



## 13.5 File Open Operati



In order to perform file operations such as read and write, the file must first be opened. There are two file open operations, open and open/create. Each of the two `open` operations are explained in the following sections.



After the file is opened, in order to perform file read or write operations the operating system needs detailed information about the file, including the complete status and current read/write location. This is necessary to ensure that read or write operations pick up where they left off (from last time).



If the file open operation fails, an error code will be returned. If the file open operation succeeds, a file descriptor is returned. This applies to both high-level languages and assembly code.



The file descriptor is used by the operating system to access the complete information about the file. The complete set of information about an open file is stored in an operating system data structure named File Control Block (FCB). In essence, the file descriptor is used by the operating system to reference the correct FCB. It is the programmer's responsibility to ensure that the file descriptor is stored and used correctly.



### 13.5.1 File O



The file open requires that the file exists in order to be opened. If the file does not exist it is an error.



The file open operation also requires the parameter flag to specify the access mode. The access mode must include one of the following:



- Read-Only Access → `O_RDONLY`
- Write-Only Access → `O_WRONLY`
- Read/Write Access → `O_RDWR`



One of these access modes must be used. Additional access modes may be used by OR'ing with one of these. This might include modes such as append mode (which is not addressed in this text). Refer to Appendix C, System Services for additional information regarding the file access modes.

The arguments for the file open system service are as follows:

Register	SYS_open
rax	Call code = SYS_open
rdi	Address of NULL terminated file name string
rsi	File access mode

Assuming the following declarations:

```
SYS_open    equ     2          ; file open
O_RDONLY    equ     000000q    ; read only
O_WRONLY    equ     000001q    ; write only
O_RDWR      equ     000002q    ; read and write
```

After the system call, the **rax** register will contain the return value. If the file open operation fails, **rax** will contain a negative value (i.e., < 0). The specific negative value provides an indication of the type of error encountered. Refer to Appendix C, System Services for additional information on error codes. Typical errors might include invalid file descriptor, file not found, or file permissions error.

If the file open operation succeeds, **rax** contains the file descriptor. The file descriptor will be required for further file operations and should be saved.

Refer to the section on Example File Read for a complete example that opens a file.

### 13.5.2 File Open/Create

A file open/create operation will create a file. If the file does not exist, a new file will be created. If the file already exists, it will be erased and a new file created. Thus, the previous contents of the file will be lost.

A file access mode must be specified. Since the file is being created, the access mode must include the file permissions that will be set when the file is created. This would include specifying read, write, and/or execute permissions for the *user*, *group*, or *world* as is typical for Linux file permissions. The only permissions addressed in this example are for the user or owner of the file. As such, other users (i.e., using other accounts) will not be able to access the file our program creates. Refer to Appendix C, System Services for additional information on file permissions.

Services for additional information regarding the file access mode  
The arguments for the file open/create system service are as follows:

Register	SYS_create
rax	Call code = SYS_create
rdi	Address of NULL terminated file name string
rsi	File access mode

Assuming the following declarations:

```

SYS_create    equ     85      ; file open/create
O_CREAT       equ     0x00000001
O_TRUNC       equ     0x00000002
O_APPEND      equ     0x00000004

S_IRUSR       equ     00400q    ; owner, read permission
S_IWUSR       equ     00200q    ; owner, write permission
S_IXUSR       equ     00100q    ; owner, execute permission

```

The file status flags “S\_IRUSR | S\_IWUSR” would allow simultaneous read and write which is typical. The “|” is a logical OR operation, thus combining the selections.

If the file open/create operation does not succeed, a negative value is returned in the rax register. If file open/create operation succeeds, a file descriptor is returned. The file descriptor is used for all subsequent file operations.

Refer to the section on Example File Write for a complete example file open/create.

## 13.6 File Read

A file must be opened with the appropriate file access flags before it can be read.

The arguments for the file read system service are as follows:

Register	SYS_read
rax	Call code = SYS_read
rdi	File descriptor (of open)
rsi	Address of where to place characters

**rdx**

Count of characters to



Assuming the following declaration:

**SYS\_read**

**equ**

**0**

**; file r**



If the file read operation does not succeed, a negative value is returned in the **rax** register. If the file read operation succeeds, the number of characters actually read is returned.

Refer to the next section on example file read for a complete file read example.



## 13.7 File W



The arguments for the file write system service are as follows:

Register	SYS_w
<b>rax</b>	Call code = SYS_write
<b>rdi</b>	File descriptor (of open)
<b>rsi</b>	Address of characters to write
<b>rdx</b>	Count of characters to write

Assuming the following declaration:

**SYS\_write**

**equ**

**1 ; file wr**



If the file write operation does not succeed, a negative value is returned in the **rax** register. If the file write operation does succeed, the number of characters actually written is returned.



Refer to the section on Example File Read for a complete file write example.



## 13.8 File Operations Examples



This section contains some simple example programs to demonstrate very basic file I/O operations. The more complex issues regarding file I/O buffering are addressed in a subsequent chapter.



### 13.8.1 Example, File W



This example program writes a short message to a file. The file created contains a simple message, a URL in this example. The file name and message to be written to the file are hard-coded. This helps simplify the example, but is not realistic.





Since the open/create service is used, the file will be created (even if an old version must be overwritten).



```
; Example program to demonstrate file I/O. This example  
; will open/create a file, write some information to the  
; file, and close the file. Note, the file name and  
; write message are hard-coded for the example.
```



```
section .d
```



```
; -----  
; Define standard constan
```



LF	equ	10	; line f
NULL	equ	0	; end of str
TRUE	equ		
FALSE	equ		
EXIT_SUCCESS	equ	0	; success c
STDIN	equ	0	; standard in
STDOUT	equ	1	; standard out
STDERR	equ	2	; standard er
SYS_read	equ	0	; r
SYS_write	equ	1	; wr
SYS_open	equ	2	; file o
SYS_close	equ	3	; file cl
SYS_fork	equ	57	; f
SYS_exit	equ	60	; termin
SYS_creat	equ	85	; file open/cre
SYS_time	equ	201	; get t
O_CREAT	equ	0	
O_TRUNC	equ	0x	
O_APPEND	equ	0x	
O_RDONLY	equ	000000q	; read o
O_WRONLY	equ	000001q	; write o
O_RDWR	equ	000002q	; read and wr
S_IRUSR	equ	004	
S_IWUSR	equ	002	
S_IXUSR	equ	001	



```

; -----
; Variables for main program
newLine      db      LF, NUL
header       db      LF, "File Write Example", NUL
              db      LF, LF, NUL
fileName     db      "url.txt", NUL
url          db      "http://www.google.com", NUL
              db      LF, NUL
len          dq      $-url

writeDone    db      "Write Completed.", LF, NUL
fileDesc     dq
errMsgOpen   db      "Error opening file.", LF, NUL
errMsgWrite  db      "Error writing to file.", LF, NUL

;-----

section      .text
global _start
_start:

; -----
; Display header line
        mov     rdi, header
        call    printStr

; -----
; Attempt to open file.
; Use system service for file open/create

; System Service - Open/Create
;     rax = SYS_creat (file open/create)
;     rdi = address of file name string
;     rsi = attributes (i.e., read only, etc)

; Returns:
;     if error -> eax
;     if success -> eax = file descriptor number

; The file descriptor points to the File Control Block (FCB).
; The FCB is maintained by the operating system.
; The file descriptor is used for all subsequent
; file operations (read, write, close, etc)

```

```
openInputFile
    mov     rax, SYS_creat
    mov     rdi, fileName
    mov     rsi, S_IRUSR | S_IWUSR
    syscall

    cmp     rax, 0
    jl      errorOnOpen

    mov qword [fileDesc], rax
; -----;
; Write to file.
; In this example, the characters to write are in a predefined string containing a Unicode character.
; System Service - write
; rax = SYS_write
; rdi = file descriptor
; rsi = address of characters to write
; rdx = count of characters to write
; Returns:
; if error -> rax = error code
; if success -> rax = count of characters actually written
    mov     rax, SYS_write
    mov     rdi, qword [fileDesc]
    mov     rsi, writeStr
    mov     rdx, qword [writeStrLen]
    syscall

    cmp     rax, writeStrLen
    jl      errorOnWrite

    mov     rdi, writeStr
    call    printString

; -----
; Close the file.
; System Service - close
; rax = SYS_close
; rdi = file descriptor
    mov     rax, SYS_close
    mov     rdi, fileDesc
```

```

    mov rdi, qword [fileDe≡]
    sysc≡

    jmp exampleD≡

; -----
; Error on open.
; note, rax contains an error code which is not u≡
; for this examp≡

errorOnOp≡
    mov rdi, errMsgO≡
    call printStr≡

    jmp exampleD≡

; -----
; Error on write.
; note, rax contains an error code which is not u≡
; for this examp≡

errorOnWri≡
    mov rdi, errMsgWr≡
    call printStr≡

    jmp exampleD≡

; -----
; Example program do≡

exampleDo≡
    mov rax, SYS_e≡
    mov rdi, EXIT_SUCC≡
    sysc≡

; ****
; Generic function to display a string to the screen.
; String must be NULL terminated.
; Algorithm:
;     Count characters in string (excluding NULL)
;     Use syscall to output character≡

; Arguments:
;     1) address str≡
; Returns: nothing≡

```

```
global printString
printString:
    push    rbp
    mov     rbp, rsp
    push    rbp
    mov     rbp, rsp
    ; Count characters in string
    mov     rbx, rbp
    mov     rdx, rbp
strCountLoop:
    cmp     byte [rbx], NUL
    je      strCountDone
    inc     rbx
    inc     rdx
    jmp     strCountLoop
strCountDone:
    cmp     rdx, rbp
    je      prtDone
    ; Call OS to output string
    mov     rax, SYS_write
    mov     rsi, rbp
    mov     rdi, STDOUT
    ; code for write
    ; addr of character
    ; file descriptor
    ; count set above
    ; system call
    syscall
    ; String printed, return to calling routine
prtDone:
    pop    rbp
    pop    rbp
; ****
```

This example creates the file which is read by the next example

### 13.8.2 Example, File R

This example will read a file. The file to be read contains a simple message, the URL from the previous example. The file name is hard-coded which helps simplify the example, but is not realistic. The file name used matches the previous file write example. If this example program is executed prior to the write example program being executed, it will generate an error since the file will not be found. After the file write example program is executed, this file read example program will read the file and display the contents.

```
; Example program to demonstrate file I/O  
  
; This example will open/create a file, write some  
; information to the file, and close the file.  
  
; Note, the file name is hard-coded for this example  
  
; This example program will open a file, read the  
; contents, and write the contents to the screen.  
; This routine also provides some very simple examples  
; regarding handling various errors on system services.  
  
; -----  
  
section .data  
  
; -----  
; Define standard constants  
  
LF equ 10 ; line feed  
NULL equ 0 ; end of string  
  
TRUE equ 1  
FALSE equ 0  
EXIT_SUCCESS equ 0 ; success code  
STDIN equ 0 ; standard input  
STDOUT equ 1 ; standard output  
STDERR equ 2 ; standard error  
  
SYS_read equ 0 ; read system call  
SYS_write equ 1 ; write system call  
SYS_open equ 2 ; file open system call  
SYS_close equ 3 ; file close system call  
SYS_fork equ 57 ; fork system call  
SYS_exit equ 60 ; terminate system call
```

## Chapter 13.0 ◀ System Serv

```
SYS_creat      equ     85          ; file open/cre
SYS_time       equ     201         ; get t
O_CREAT        equ     0x00000001
O_TRUNC        equ     0x00000002
O_APPEND       equ     0x00000004
O_RDONLY       equ     0000000q    ; read o
O_WRONLY       equ     0000001q    ; write o
O_RDWR         equ     0000002q    ; read and wr
S_IRUSR        equ     00400000
S_IWUSR        equ     00200000
S_IXUSR        equ     00100000
;
; -----
; Variables/constants for main
BUFF_SIZE      equ     2
newLine         db      LF, NU
header          db      LF, "File Read Example", NU
                db      LF, LF, NU
fileName         db      "myrl.txt", NU
fileDesc         dq
errMsgOpen       db      "Error opening the file.", LF, NU
errMsgRead       db      "Error reading from the file.", LF, NU
;
; -----
section .data
readBuffer resb  BUFF_SIZE
;
; -----
section .text
global _start
_start:
;
; -----
; Display header line.
```

```

mov      rdi, head[...]           ; rdi = address of head of linked list
call    printString[...]
; -----
; Attempt to open file - Use system service for file op[...]
;
; System Service - Open[...]
;     rax = SYS_open[...]
;     rdi = address of file name str[...]
;     rsi = attributes (i.e., read only, etc)[...]
; Returns:
;     if error -> eax <[...]
;     if success -> eax = file descriptor num[...]
;
; The file descriptor points to the File Control Block (FCB). The FCB is maintained by the OS[...]
; The file descriptor is used for all subsequent file operations (read, write, close)[...]
;
openInputFile[...]
    mov      rax, SYS_open          ; file op[...]
    mov      rdi, fileName          ; file name str[...]
    mov      rsi, O_RDONLY          ; read only acc[...]
    syscall                         ; call the kernel[...]
    cmp      rax, 0                ; check for success[...]
    jl       errorOnOpen[...]
    mov      qword [fileDesc], rax  ; save descrip[...]
;
; -----
; Read from file.
; For this example, we know that the file has only 1 line[...]
;
; System Service - Read[...]
;     rax = SYS_read[...]
;     rdi = file descrip[...]
;     rsi = address of where to place data[...]
;     rdx = count of characters to read[...]
; Returns:
;     if error -> rax <[...]
;     if success -> rax = count of characters actually read[...]
    mov      rax, SYS_read[...]

```

## Chapter 13.0 ◀ System Serv

```
        mov     rdi, qword [fileDe█]
        mov     rsi, readBuf█
        mov     rdx, BUFF_S█
        sysc█

        cmp     rax█
        jl      errorOnR█

; -----
; Print the buffer.
; add the NULL for the print str█

        mov     rsi, readBuf█
        mov     byte [rsi+rax], N█

        mov     rdi, readBuf█
        call   printStr█

        printNewL█

; -----
; Close the file.
; System Service - close
;     rax = SYS_close█
;     rdi = file descrip█

        mov     rax, SYS_cl█
        mov     rdi, qword [fileDe█]
        sysc█

        jmp     exampleD█

; -----
; Error on open.
; note, eax contains an error code which is not u█
; for this examp█

errorOnOp█
        mov     rdi, errMsgO█
        call   printStr█

        jmp     exampleD█
```

```

; -----
; Error on read.
; note, eax contains an error code which is not u[...]
; for this example[...]

errorOnRe[...]
    mov     rdi, errMsgR[...]
    call    printStr[...]
    jmp    exampleD[...]

; -----
; Example program do[...]

exampleDo[...]
    mov     rax, SYS_e[...]
    mov     rdi, EXIT_SUCC[...]
    sysc[...]

; *****
; Generic procedure to display a string to the screen.
; String must be NULL terminated.
; Algorithm:
; Count characters in string (excluding NULL)
; Use syscall to output character[...]

; Arguments:
; 1) address, str[...]
; Returns:
; nothing[...]

global printString
printString:
    push   rbp[...]
    mov    rbp, [rbp+8][...]
    push   rbp[...]
    mov    rbp, [rbp+8][...]
    push   rbp[...]
    mov    rbp, [rbp+8][...]

; -----
; Count characters in string[...]

    mov     rbx, strCountLo[...]
    mov     rdx, strCountLo[...]
strCountLo:
    cmp    byte [rbx], N[...]
    je    strCountD[...]

```

```
inc          [ ]  
inc          [ ]  
jmp  strCountL [ ]  
strCountDo [ ]  
  
cmp      rdx [ ]  
je       prtD [ ]  
  
; -----  
; Call OS to output string [ ]  
  
mov      rax, SYS_write  
mov      rsi, rdi  
mov      rdi, STDOUT  
  
        ; code for write  
        ; addr of character  
        ; file descriptor  
        ; count set above  
        ; system call [ ]  
  
syscall  
  
; -----  
; String printed, return to calling routine [ ]  
  
prtDo [ ]  
pop  
ret  
[ ]
```

The `printString()` function is the exact same in both examples and is only repeated to allow each program to be assembled and executed independently.

## 13.9 Exerci

Below are some quiz questions and suggested projects based on this chapter.

### 13.9.1 Quiz Questi

Below are some quiz questions based on this chapter.

- When using a system service, where is the *call code* placed? [ ]
  - Where is the code located when the **syscall** instruction is executed? [ ]
  - What is the *call* code and required arguments for a system service call to perform console output? [ ]
  - Why was only one character read for interactive keyboard input? [ ]

- 5) What is returned for a successful file open system service call? 
- 6) What is returned for an unsuccessful file open system service call? 
- 7) If a system service call requires six (6) arguments, where specifically should they be passed? 

### 13.9.2 Suggested Projects

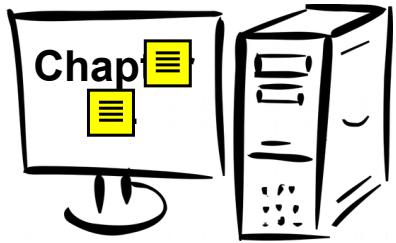
Below are some suggested projects based on this chapter.

- 1) Implement the *printString()* example void function and a simple main to test a series of strings. Use the debugger to execute the program and display the final results. Execute the program without the debugger and verify the appropriate output is displayed to the console. 
- 2) Convert the example program to read input from keyboard into a *readString()* function. The function should accept arguments for the string address and maximum string length (in that order). The maximum length should include space for the NULL (an extra byte), which means the function must not allow more than the maximum minus one characters to be stored in the string. If additional characters are entered by the user, they should be cleared from the input stream, but not stored. The function should not include the newline in the returned string. The function should return the number of characters in the string not including the NULL. The *printString()* function from the previous problem should be used unchanged. When done, create an appropriate main to test the function. Use the debugger as necessary to debug the program. When working correctly, execute the program from the command line which will display the final results to the console. 
- 3) Based on the file write example, create a value returning *fileWrite()* function to write a password to a file. The function should accept arguments for the address of the file name and the address of the NULL terminated password string. The file should be created, opened, the password string written to the file, and the file closed. The function should return SUCCESS if the operations worked correctly or NOSUCCESS if there is a problem. Problems might include not being able to create the file or not being able to write to the file. Create an appropriate main to test the function. Use the debugger as necessary to debug the program. When working correctly, execute the program from the command line which will display the final results to the console. 
- 4) Based on the file read example, create a value returning *fileRead()* function to read a password from a file. The function should accept arguments for the address of file name, the address of where to store the password string, 

maximum length of the password string, and the address of where to store the password length. The function should open the file, read a string representing a password, close the file, and return the number of characters in the password. The maximum length should include space for the NULL, which means the function read must not store more than the maximum minus one characters in the string. The function should return SUCCESS if the operations worked correctly or NOSUCCESS if there is a problem. Problems might include the file not existing or other read errors. Create an appropriate main to test the function. Use the debugger as necessary to debug the program. When working correctly, execute the program from the command line which will display the final results to the console.



*My new computer is so fast, it executes an infinite loop in 6 seconds.*



## 14.0 Multiple Source File

As a program grows in size and complexity, it is common that different parts of the program are stored in different source files. This allows a programmer to ensure the source files do not become unreasonably large and would also allow multiple programmers to more easily work on different parts of the program.

The function calls, even those written by different programmers, will work together correctly due to the standard calling convention as outlined in Chapter 12, Functions. This is even true when interfacing with a high-level language.

This chapter will present a simple assembly language example to demonstrate how to create and use source code in multiple files. In addition, an example of how to interface with a C/C++ source file is provided.

### 14.1 Extern Statement

If a function is called from a source file and the function code is not located in the current source file, the assembler will generate an error. The same applies to variables accessed that are not located in the current file. In order to inform the assembler that the function code or variables are in another file, the **extern** statement is used. The syntax is as follows:

```
extern <symbolName>
```

The symbol name would be the name of the function or a variable that is located in a different source file. In general, using global variables accessed across multiple files is considered poor programming practice and should be used sparingly (if at all). Data is typically passed between functions as arguments for the function call.

The examples in the text focus on using external functions only with no globally declared variables.



## 14.2 Example, Sum and Aver



The following is a simple example of a main that calls an assembly language function, *stats()*, to compute the integer sum and integer average for a list of signed integers. The main and the function are in different source files and are presented as an example of how multiple source files are used. The example itself is really too small to actually require multiple source files.



### 14.2.1 Assembly M



The main is as follo



```
; Simple example to call an external functi
```

---

```
; -----
```

```
; Data section
```

```
section    .d
```

```
; -----
```

```
; Define standard consta
```

```
LF          equ      10           ; line f
```

```
NULL        equ      0            ; end of str
```

```
TRUE        equ      1
```

```
FALSE       equ      0
```

```
EXIT_SUCCESS equ      0           ; success
```

```
SYS_exit    equ      60          ; termin
```

```
; -----
```

```
; Declare the d
```

```
lst1         dd      1, -2,  2, -4,
```

```
len1        dd      7,  9,
```

```
lst2         dd      2, -3,  4, -5,
```

```
len2        dd      7, 10, 12, 14,
```



```

section .data
sum1      resd
ave1      resd
          resd
sum2      resd
ave2      resd

; -----[=]

extern stdio

section .text
global _start
_start:

; -----
; Call the function
;     HLL Call: stats(lst, len, &sum, &ave)
        mov    rdi, lst1           ; data section
        mov    esi, dword [len]
        mov    rdx, sum
        mov    rcx, ave
        call   stdio

        mov    rdi, lst2           ; data section
        mov    esi, dword [len]
        mov    rdx, sum
        mov    rcx, ave
        call   stdio

; -----
; Example program done
exampleDone:
        mov    rax, SYS_exit
        mov    rdi, EXIT_SUCCESS
        syscall

```

The above main can be assembled with the same assemble command as described in Chapter 5, Tool chain. The **extern** statement will ensure that the assembler does not generate errors.

## 14.2.2 Function Source

The function, in a different source file is as follows:

```
; Simple example void function
; **** Data declarations ****
; Note, none needed for this example.
; If needed, they would be declared here as always.

section .data
; ****

section .text
; -----
; Function to find integer sum and integer average
; for a passed list of signed integers.

; Call:
; stats(lst, len, &sum, &avg)

; Arguments Passed:
; 1) rdi - address of array
; 2) rsi - length of passed array
; 3) rdx - address of variable for sum
; 4) rcx - address of variable for average

; Returns:
; sum of integers (via reference)
; average of integers (via reference)

global stats
stats:
    push rbp
    mov rbp, rsp

    ; -----
    ; Find and return sum

    mov r11, 0
    mov r12d, 0
    ; sum = 0
    ; sum = sum + value
```

```

sumLo
    mov    eax, dword [rdi+r11*4]    ; get 1st
    add    r12d, eax                ; update
    inc    r11
    cmp    r11, sumL
    jb     sumL

    mov    dword [rdx], r12d        ; return

; -----
; Find and return avera
    mov    eax, r
    cdq
    idiv
    mov    dword [rcx], eax        ; return aver

; -----
; Done, return to calling functi
    ret

```

The above source file can be assembled with the same assemble command as described in Chapter 5, Tool chain. No **extern** statement is required since no external functions are called.

### 14.2.3 Assemble and L

Assuming the main source file is named *main.asm* and the functions source file is named *stats.asm*, the following command will perform the assemble and link.

```

yasm -g dwarf2 -f elf64 main.asm -l main.lst
yasm -g dwarf2 -f elf64 stats.asm -l stats.lst
ld -g -o main main.o stats.o

```

The files names can be changed as desi

As usual, the debugger is started with the **ddd <executable>** command. It would be appropriate to note that using the debugger **step** command will step into the function, including showing the function source code (even if the source is in a different file). The debugger **next** command will execute the entire function and not show the source. If the function is working, the **next** command would be most useful. In order to de

the function, the **step** command would be most useful.

If the “-g” option is omitted, the debugger will not be able to display the source code.

## 14.3 Interfacing with a High-Level Language

This section provides information on how a high-level language can call an assembly language function and how an assembly language function can call a high-level language function. This chapter presents examples for both.

In brief, the answer of how this is accomplished is through the standard calling convention. As such, no additional or special code is needed when interfacing with a high-level language. The compiler or assembler will need to be informed about the external routines in order to avoid error messages about not being able to find the source code for the non-local routines.

The general process for linking for multiple files was described in Chapter 5, Tool Chain. The process of using multiple source files was described in Chapter 14, Multiple Source Files. It does not matter if the object files are from a high-level language or from an assembly language source.

### 14.3.1 Example, C++ Main / Assembly Function

When calling any functions that are in a separate source file, the compiler must be informed that the function or functions source code are external to the current source file. This is performed with an **extern** statement in C or C++. Other languages will have a similar syntax. For a high-level language, the **extern** statement will include the function prototype which will allow the compiler to verify the function parameters and associated types.

The following is a simple example using a C++ main with an assembly language function.

```
#include <iostream>
using namespace std;
extern "C" void stats(int[], int, int *, int)
```

```
int main()
{
    int lst[] = {1, -2, 3, -4, 5, 7, 9, 11};
    int len = 8;
    int sum, ave;
```

```
stats(lst, len, &sum, &ave)  
  
cout << "Stats:" << endl;  
cout << " Sum = " << sum << endl;  
cout << " Ave = " << ave << endl  
  
return
```

At this point, the compiler does not know the external function is written in assembly (nor does it matter).

The C compiler is pre-installed on Ubuntu. However, the C++ compiler is not installed by default.

A C version of the same program is also presented for completeness.

```
#include<stdio.h>  
extern void stats(int[], int, int *, int)  
  
int main()  
{  
    int lst[] = {1, -2, 3, -4, 5, 7, 9, 11};  
    int len = 8;  
    int sum, ave;  
  
    stats(lst, len, &sum, &ave)  
  
    printf ("Stats:\n");  
    printf (" Sum = %d \n", sum);  
    printf (" Ave = %d \n", ave)  
  
    return
```

The *stats()* function referenced here should be used unchanged from the previous example.

### 14.3.2 Compile, Assemble, and Link

As noted in the Functions Chapter, the C++ compiler should be used. For example, assuming that the C++ main is named *main.cpp*, and the assembly source file is named *stats.asm*, the commands to compile, assemble, link, and execute as follows:

```
g++ -g -Wall -c main.cpp
yasm -g dwarf2 -f elf64 stats.asm -l stats.list
g++ -g -o main main.o stats.o
```

Note, Ubuntu 18 will require the **no-pie** option on the g++ command as shown

```
g++ -g -no-pie -o main main.o stats.o
```

The file names can be changed as desired. Upon execution, the output would be as follows:

```
./main
Stats:
Sum = 20
Ave =
```

If a C main is used, and assuming that the C main is named *main.c*, and the assembly source file is named *stats.asm*, the commands to compile, assemble, link, and execute are as follows:

```
gcc -g -Wall -c main.c
yasm -g dwarf2 -f elf64 stats.asm -l stats.list
gcc -g -o main main.o stats.o
```

Note, Ubuntu 18 will require the **no-pie** option on the gcc command as shown

```
gcc -g -no-pie -o main main.o stats.o
```

The C compiler, **gcc**, or the C++ compiler, **g++**, is used to perform the linking as it is aware of the appropriate locations for the various C/C++ libraries.

## 14.4 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 14.4.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) What is the statement to declare the functions, *func1()* and *func2()*, as extern, assuming no arguments are needed for either function?
- 2) What is the statement to declare the functions, *func1()* and *func2()*, as extern, if two integer arguments are used for each function?

- 3) What will happen if an external function is called but is not declared as external? 
- 4) If an externally declared function is called but the programmer did not actually write the function, when would the error be flagged, assemble-time, link-time, or run-time? 
- 5) If an externally declared function is called but the programmer did not actually write the function, what might the error be? 
- 6) If the “-g” option is omitted from the assemble and link commands, would the program be able to execute? 

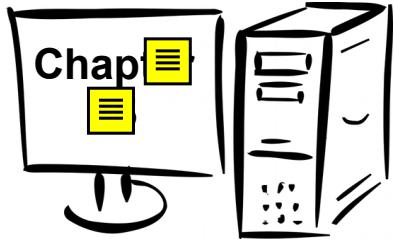
#### 14.4.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Implement the assembly language example program to find the sum and average for a list of signed integers. Ensure that the main and function are in different source files. Use the debugger to execute the program and display the final results. 
- 2) Based on the example function *stats()*, split it into two value returning functions (*lstSum()* and *lstAverage()*). As noted in Chapter 12, Functions, value returning functions return their result in the **A** register. Since these are double-words, the result will be returned in **eax**. Ensure that the main and both functions are in two different source files. The two functions can be in the same source file. Use the debugger to execute the program and display the final results. 
- 3) Extend the previous exercise to display the sum and average to the console. The *printString()* example function (from multiple previous examples) should be placed in a third source file (which can be used on other exercises). This project will require a function to convert an integer to ASCII (as outlined in Chapter 10). Use the debugger as needed to debug the program. When working, execute the program without the debugger and verify that the correct results are displayed to the console. 
- 4) Implement one of the C/C++ example main programs (either one). Additionally, implement the assembly language *stats()* function example. Develop a simple bash script to perform the compile, assemble, and link. The link should be performed with the applicable C/C++ compiler. Use the debugger as needed to debug the program. When working, execute the program without the debugger and verify that the correct results are displayed to the console. 



*My new computer is so fast it requires two HALT instructions to stop it.*



## 15.0 Stack Buffer Overf<sup>44</sup>lo

A stack buffer overflow<sup>44</sup> can occur when a program overflows a stack-based dynamic variable (as described in Chapter 12.9, Stack-Based Local Variables). For example, if a program allocates and uses a stack-based local array holding 50 elements and more than 50 elements are stored in the array, an overflow occurs. Such overflows are generally bad and typically cause program bugs and possibly even crash the program. The stack will contain other important information such as other variables, preserved registers, frame pointer, return address, and/or stack-based parameters. If such data is overwritten, it will likely cause problems which can be very difficult to debug since the symptom will likely be unrelated to where the problem actually occurs.

If a stack buffer overflow is caused deliberately as part of an attack it is referred to as stack smashing. Due to the standard calling convention, the layout of the stack-based call frame or activation record is fairly predictable. Such a stack buffer overflow can be by a malicious individual to inject executable code into the currently running program to perform some inappropriate actions. Under the right circumstances, such code injection could allow a black-hat<sup>45</sup> hacker to perform unwanted actions potentially taking over the system.

The process of how a stack buffer overflow occurs and how it can be exploited are provided in this chapter. This is presented in order to allow developers to clearly understand the problem and thus learn how to protect themselves against such vulnerabilities. The reader must be familiar with the details of the standard calling convention as outlined in Chapter 12, Functions.

It should be noted that the stack buffer overflow problem exists in high-level languages. Working in assembly languages makes it easier to more clearly see and understand the details.

<sup>44</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Stack\\_buffer\\_overflow](http://en.wikipedia.org/wiki/Stack_buffer_overflow)

<sup>45</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Black\\_hat](http://en.wikipedia.org/wiki/Black_hat)

## 15.1 Understanding a Stack Buffer Overf☰

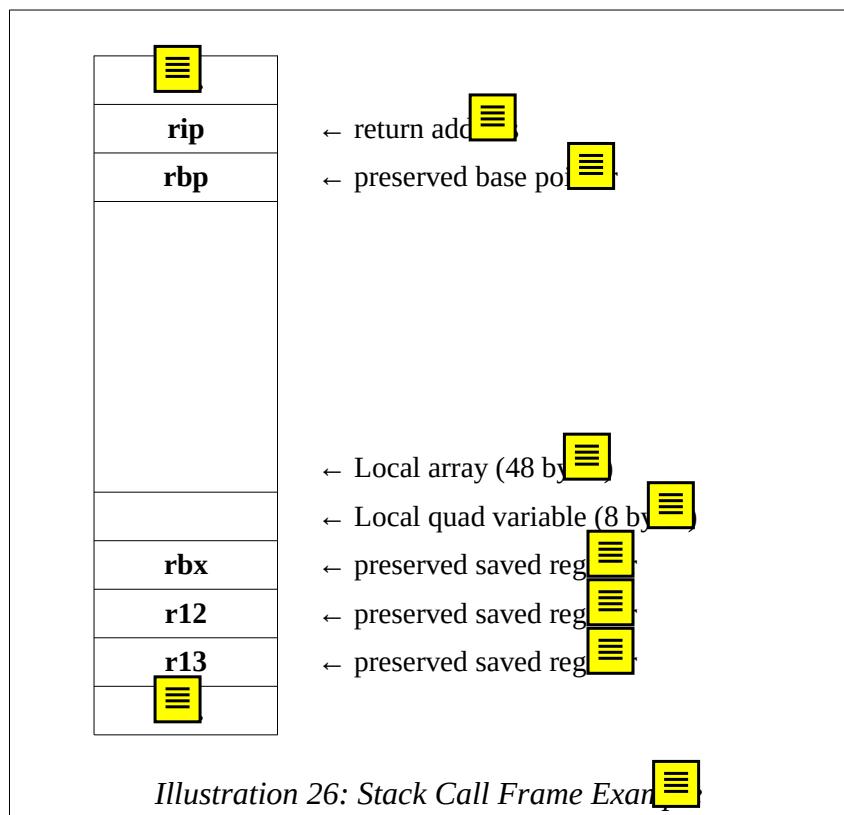
When a program calls a function, the standard calling convention provides the guidance for how the parameters are passed, how the return address is saved, how and which registers must be preserved, and how stack-based local variables are to be allocated.

For example, consider the function

```
expFunc(arg1, arg2, arg3)
```

In addition, we will assume that the function, *expFunc()*, reads in a line of text from the user and stores it in a locally declared array. The local variables include the array of 48 bytes and one quadword local variable (8 bytes). This can be accomplished easily in a high-level language or as is done in Chapter 13, System Services, Console Input.

The resulting call frame would be as folld



When the function completes, all elements of the call frame are removed. The preserved registers are restored to their original contents, the local variables are removed, and the return address is copied off the stack and placed in the **rip** register which effects a jump back to the original calling routine. As noted in Chapter 12 Functions, this layout supports multiple levels of function calls including recursion.

The example in Chapter 13 that reads characters from the user and entered them into an array explicitly checked the character count to ensure that the count does not exceed the buffer size. It would be easy to forget to perform this important check. Indeed, some C functions such as *strcpy()* do not perform verification of sizes and are thus considered unsafe<sup>46</sup> or described as an unsafe function.

In this example, over-writing the 48 character buffer will destroy the contents of the stack that contains the original values for the **rbp** register and possibly the **rip** register. If the stack content of the original value of the **rip** register is altered or corrupted in any way, the function will not be able to return to the calling routine and would attempt to jump to some random location. If the random location is outside the program scope, which is likely, the program will generate a segment fault (i.e., “seg fault” or program crash).

Debugging this type of error can be a significant challenge. The error appears at the very end of the function (on the return). The problem is actually in the body of the function and in fact may be due to code that is not included (as opposed to wrong code that is there).

Testing if a program has this vulnerability would involve typing significantly more characters than expected when prompted. For example, if name is prompted for and 200 characters are entered (possibly by holding down a key), the program crashing would be a sign that such a vulnerability exists. An error message would indicate such a vulnerability does not exist. While straightforward, this type of testing is often not checked thoroughly or even omitted entirely.

## 15.2 Code to Inject

Before discussing how the stack buffer overflow might be exploited, we will review what code might be injected. The code to be injected could be many things. We will assume the program is being executed in a controlled environment with no console access for the user. The lack of console access would limit what a user could do ideally to only what the program allowed. This might be the case if the program is server-based interacting with a user through a web or application front-end. The server would be protected from direct user access for obvious security reasons.

---

46 For more information, refer

[http://en.wikipedia.org/wiki/C\\_standard\\_library#Buffer\\_overflow\\_vulnerabilities](http://en.wikipedia.org/wiki/C_standard_library#Buffer_overflow_vulnerabilities)

As such, one possible attack vector might be to obtain console access to the back-end server (where such access is normally not allowed). 

There is a system service that can execute another program. Using this system service, exec, we could execute the **sh** program (shell) which is a standard Linux program. This will create and open a shell providing console access. The newly created shell is limited to the privileges of the process that started it. In a good security environment, the process will be granted only the privileges it requires. In a poor security environment, the process may have additional, unnecessary privileges which would make it easier for an unauthorized black-hat hacker to cause problems. 

Given the following data declarati<sup>on</sup>

```
NULL          equ      0
progName      db       "/bin/sh", N
```

An example of the exec system service would be as follo<sup>wing</sup>

; Example, system service call for ex<sup>ec</sup>

```
mov    rax, 59
mov    rdi, progName
syscall
```

A new console will appear. The reader is encouraged to try this code example and see it work. 

The list file for this code fragment would be as follo<sup>wing</sup>

40 00000000 48C7C03B000000	mov rax, 59
41 00000007 48C7C7[00000000]	mov rdi, progName
42 0000000E 0F05	syscall 

Recall that the first column is the line number, the second column is relative address in the code section and the third column is the machine language or the hex representation of the human readable instruction shown in the fourth column. The [00000000] represents the relative address of the string in the data section (*progName* in this example). It is zero since it is the first (and only) variable in the data section for this example. 

If the machine language shown is entered into memory and executed, a shell or console would be opened. Getting the hex version of the code into memory can only be performed via the keyboard (since there is no direct access to file system). This w<sup>ould</sup> 

be done one character at a time. The 0x48 is the ASCII code for “0”, so “0” could be entered for that byte. However, the 0x0f and many of the other characters are more difficult to enter directly as ASCII.

The command shell will allow entry of hex code in hex. By typing control key and the left shift followed by a lower-case u and then four hex digits (must be hex), the hex values can be entered one at a time. For example, to enter the 0x3f, it would be;

CTRL SHIFT u 

This can be done for most of the bytes except the 0x00. The 0x00 is a NULL which is a non-printable ASCII characters (used to mark string terminations). As such, the NULL cannot be entered from the keyboard. Additionally, the [00000000] address would not make sense if the code is injected into another program.

To address these issues, we can re-write the example code and eliminate the NULL's and change the address reference. The NULL's can be eliminated by using different instructions. For example, setting **rax** to 59 can be accomplished by xor'ing **rax** with itself and placing the 59 in **al** (having already ensured the upper 56 bits are 0 via the xor). The string can be placed on the stack and the current **rsp** used as the address of the string. The string, “bin\sh” is 7 bytes and the stack operation will require a push of 8 bytes. Again, the NULL cannot be entered and is not counted. An extra, unnecessary “/” can be added to the string which will not impact the operation providing exactly 8 bytes in the string. Since the architecture is little-endian, in order to ensure that the start of the string is in low memory, it must be in the least significant byte of the push. This will make the string appear backwards.

The revised program fragment would be as follo

<b>xor rax, rax</b>	<b>; clear </b>
<b>push rax</b>	<b>; place NULLs on st</b>
<b>mov rbx, 0x68732f6e69622f2f</b>	<b>; string -&gt; “//bin/”</b>
<b>push rbx</b>	<b>; put string in mem</b>
<b>mov al, 59</b>	<b>; call code in </b>
<b>mov rdi, rsp</b>	<b>; rdi = addr of str</b>
<b>syscall</b>	<b>; system c</b>

The list file for the program fragment would be as follo

52 00000013 4831C0	<b>xor rax, rax</b>
53 00000016 50	<b>push rax</b>
54 00000017 48BB2F2F62696E2F73	<b>mov rbx, 0x68732f6e69622f2f</b>
55 00000017 68	<b>push rbx</b> 
56 00000021 53	

```
57 00000022 B03B           mov al, 59
58 00000024 4889E7         mov rdi, r☰
59 00000027 0F05          syscall
```

In this revised code, there are no NULL's and the address reference is obtained from the stack pointer (**rsp**) which points to the correct string.

There is an assembly language instruction **nop** which performs no operation with a machine code of 0x90. In this example, the **nop** instruction is used simply to round out the machine code to an even multiple of 8 bytes.

The series of hex values that would need to be entered is as follows:

```
0x48 0x31 0xC0 0x50 0x48 0xBB 0x2F 0x2F
0x62 0x69 0x6E 0x2F 0x73 0x68 0x53 0xP0
0x3B 0x48 0x89 0xE7 0x0F 0x05 0x90 0x☰
```

While somewhat tedious, these characters can be entered by hand.

### 15.3 Code Inject☰

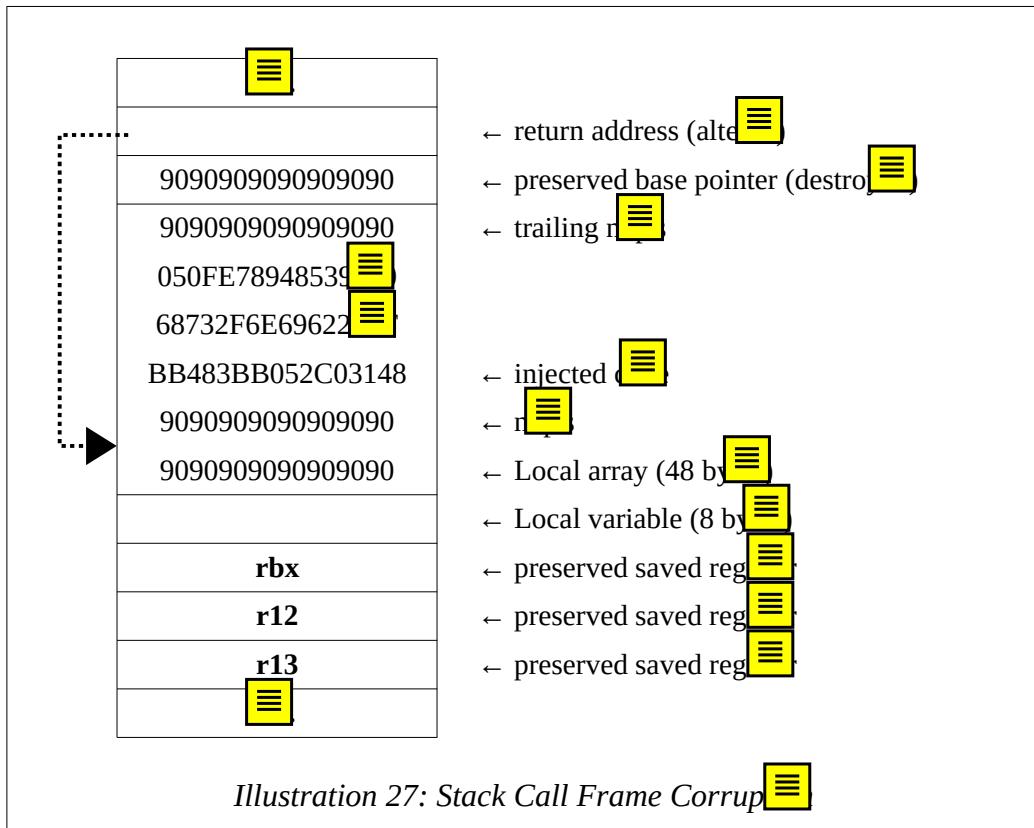
If the code to inject is available and can be entered, the next step would be actually getting the code executed.

Based on the previous example call frame, the code would be entered preceded by a series of **nop**'s (0x90). The exact spot where the **rip** is stored in the stack can be determined through trial and error. When the first byte of the 8 byte address is altered, the program will not be able to return to the calling routine and will likely crash. If the bytes of the **rbp** are corrupted, the program may fail in some way, but it will be different than the immediate crash caused by the corrupted **rip**. The code entered would be extended by 1 byte on each of many successive attempts. Finding this exact location in this manner will take patience.

Once the **rip** location has been determined, the 8 bytes that are entered there will need to be the address of where the injected code is in the stack where the user input was stored. This also would be determined through trial and error. However, the exact address of the start of the injected code is not required. Starting anywhere within the preceding **nop**'s would be sufficient. This is referred to as a NOP slide<sup>47</sup> which will help “slide” the CPU's instruction execution flow to the injected code.

---

47 For more information, refer to: <http://en.wikipedia.org/wiki/NOP>



A larger local array would allow even more room for a longer NOP sequence.

## 15.4 Code Injection Protection

A number of methods have been developed and implemented to protect against the stack buffer overflow. Some of these methods are summarized here. It must be noted that none of these methods are completely perfect.

### 15.4.1 Data Stack Smashing Protector (or Canary)

Data stack smashing protector, also referred to as stack canaries, is used to detect a stack buffer overflow before execution of malicious code can occur. This works by placing an integer, the value of which is randomly chosen at program start, in memory just before the return address (`rip`) in the call frame. In order to overwrite the return address, and thus execute the injected code, the canary value must also be overwritten. This canary value is checked to make sure it has not changed before a routine pops the return address.



For the GNU g++ compiler, this option (**-f-stack-protector**) is enabled by default. It can be turned off with the **-fno-stack-protector** compiler option. Turning it off would be necessary in order to perform testing using a C/C++ program of the injection techniques outlined in this chapter.



### 15.4.2 Data Execution Prevention



Data Execution Prevention<sup>48</sup> (DEP) is a security feature that marks areas of memory as either "executable" or "nonexecutable". Only code in a marked "executable" area is allowed to be executed. Code that is or injected into an area marked "nonexecutable" will not be allowed to execute. This helps prevent stack buffer overflow code injection.



### 15.4.3 Data Address Space Layout Randomization



Address space layout randomization (ASLR) is a technique to prevent an attacker from reliably jumping to the injected code. ASLR randomly arranges the address space positions of key data areas of a process, including the base of the executable and the positions of the stack, heap and libraries.



## 15.5 Exercises



Below are some quiz questions and suggested projects based on this chapter.



### 15.5.1 Quiz Questions



Below are some quiz questions based on this chapter.



- 1) When stack buffer overflow is caused deliberately as part of an attack referred as what?
- 2) What does it mean when a C function is considered unsafe?
- 3) Is a program that reads user input still vulnerable if the input buffer is sufficiently large (e.g., >1024 bytes)?
- 4) How might an attacker determine if an interactive program is vulnerable to buffer overflow attacks?
- 5) What is a “NOP slide”?
- 6) The text example injected code to open a new shell. Provide at least two different ideas for injected code that would cause problems.



---

48 For more information, refer to: [http://en.wikipedia.org/wiki/Data\\_Execution\\_Protection](http://en.wikipedia.org/wiki/Data_Execution_Protection)



- 7) Name three techniques designed to prevent stack buffer overflow attacks.

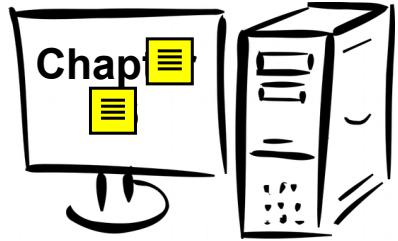
### 15.5.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Implement the second example program fragment to open a new shell. Use the debugger to execute the program and display the final results. Execute the program without the debugger and verify that a new shell is opened.
- 2) Implement the console input program from Chapter 13. Remove the code for the buffer size check. Execute the program without the debugger and ensure the appropriate input is read and that the output is displayed to the console. Verify that entering too many characters will crash the program.
- 3) Using the program from the previous question and the program fragment to open a shell, attempt to inject the code into the running program. In order to save time, print the value of the **rsp** at an appropriate location to allow the guessing of the target address significantly easier.



*The code that is the hardest to debug is the code you know can't possibly be wrong.*



## 16.0 Command Line Argument

This chapter provides a summary of how the operating system handles command [line](#) arguments and how assembly language routines can access the arguments.

The term “command line arguments” is used to refer to the characters, if any, entered on the command line after the program name. The command line arguments are often used to provide information or parameters to the program without the need for the I/O associated with interactive prompting. Of course, if the parameters are not correct, [the](#) entire line must be re-entered.

Command line arguments are used by the assembler and linker to provide information [for](#) input files, output files, and various other options.

### 16.1 Parsing Command Line Argument

The operating system is responsible for parsing, or reading, the command line arguments and delivering the information to the program. It is the responsibility of the program to determine what is considered correct and incorrect. When reading the command line, the operating system will consider an argument a set of non-space characters (i.e., a string). The space or spaces between arguments is removed or ignored by the operating system. In addition, the program name itself is considered the first, and possibly only, argument. If other arguments are entered, at least one space is required [between](#) each argument.

All arguments are delivered to the program as character strings, even number information. If needed, the program must convert the character data into the number value (float or integer).

For example, executing the program **expProg** with the following command [line](#) arguments:

```
./expProg one 42 th
```

will result in four arguments as folld

1. ./expProg
2. one
3. 42
4. three

The command line arguments are delivered to the program as parameters. As such, the program is like a function being called by the operating system. So, the standard calling convention is used to convey the parameters.

## 16.2 High-Level Language Exam

In general, it is assumed that the reader is already familiar with the basic C/C++ command line handling. This section presents a brief summary of how the C/C++ language handles the delivery of command line information to the program.

The count of arguments is passed to the main program as the first integer parameter, typically called *argc*. The second parameter, typically called *argv*, is an array of addresses for each string associated with the corresponding argument.

For example, the following C++ example program will read the command line arguments and display them to the screen.

```
#include <iomanip>
#include <iostream>
using namespace s
```

```
int main(int argc, char* argv[])
{
    string bars;
    bars.append(50, '-')
    cout << bars << endl;
    cout << "Command Line Arguments Exam"
        << endl << en
    cout << "Total arguments provided: "
        << argc << en
    cout << "The name used to start the program"
        << argv[0] << en
    if (argc > 1)
```

```
cout << endl << "The arguments are:" << endl;
for (int n = 1; n < argc; n++)
    cout << setw(2) << n << ": "
        argv[n] << endl;
return 0;
```

Assuming the program is named **argsExp**, executing this program will produce the following output:

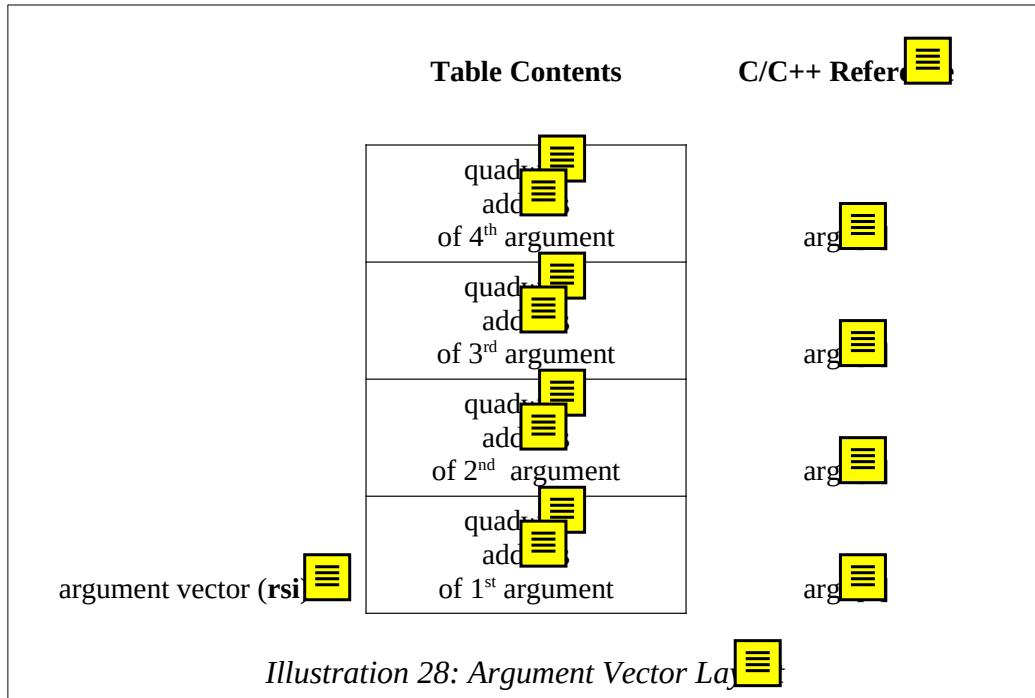
```
./argsExp    one 34      three
-----
Command Line Arguments Example
Total arguments provided: 4
The name used to start the program: ./argsExp
The arguments are:
1: one
2: 34
3: three
-----
```

It should be noted that the parameter '34' is a string. This example simply printed the string to the console. If a parameter is to be used as a numeric value, the program must convert as required. This includes all error checking as necessary.

### 16.3 Argument Count and Argument Vector Table

Since the operating system will call the main program as a function, the standard calling convention applies. The argument count and the argument vector address are passed as parameters in **rdi** and **rsi** in accordance with the standard calling convention. The first argument, in **rdi**, is the integer argument count. The second argument, in **rsi**, is the argument vector table address.

The argument vector table is an array containing the quadword addresses of the string for each argument. Assuming the previous examples with 4 total arguments, the basic argument vector table layout is as follows:



Each string is NULL terminated by the operating system and will not contain a new line character.

## 16.4 Assembly Language Examples

An example assembly language program to read and display the command line arguments is included for reference. This example simply reads and displays the command line arguments.

```
; Command Line Arguments Example
; -----
section .data
; -----
; Define standard constants.
LF equ 10 ; line f☰
NULL equ 0 ; end of str☰
TRUE equ 1
FALSE equ 0
```

```

EXIT_SUCCESS    equ    0          ; success code
STDIN           equ    0          ; standard input
STDOUT          equ    1          ; standard output
STDERR          equ    2          ; standard error

SYS_read         equ    0          ; read system call
SYS_write        equ    1          ; write system call
SYS_open          equ    2          ; file open system call
SYS_close         equ    3          ; file close system call
SYS_fork          equ    57         ; fork system call
SYS_exit          equ    60         ; terminate system call
SYS_creat         equ    85         ; file open/create system call
SYS_time          equ    201        ; get time system call

; -----
; Variables for main
newLine          db     LF, NUL

; -----
section .text
global main
main:

; -----
; Get command line arguments and echo to screen.
; Based on the standard calling convention,
;     rdi = argc (argument count)
;     rsi = argv (starting address of argument vector)

        mov     r12, rdi             ; save for later use
        mov     r13, rsi

; -----
; Simple loop to display each argument to the screen.
; Each argument is a NULL terminated string, so can just
; print directly.

printArgument:
        mov     rdi, newLine
        call   printStr

        mov     rbx

```

## Chapter 16.0 ◀ Command Line Argum

```
printLo[...]
    mov    rdi, qword [r13+rbx][...]
    call   printStr[...]
    mov    rdi, newL[...]
    call   printStr[...]
    inc    rbx[...]
    cmp    rbx, [...]
    jl    printL[...]

; -----
; Example program do[...]
exampleDo[...]
    mov    rax, SYS_e[...]
    mov    rdi, EXIT_SUCC[...]
    sysc[...]
; *****
; Generic procedure to display a string to the screen.
; String must be NULL terminated.
; Algorithm:
; Count characters in string (excluding NULL)
; Use syscall to output character[...]
; Arguments:
; 1) address, str[...]
; Returns:
; nothing[...]

global printString[...]
printString[...]
    push  [...]
    mov   rbp, [...]
    push  [...]
; -----
; Count characters in string[...]
    mov   rbx, [...]
    mov   rdx, [...]
strCountLo[...]
    cmp   byte [rbx], N[...]
    je    strCountD[...]
```

```
inc  
inc  
jmp    strCountL  
strCountDo  
cmp    rdx  
je     prtD  
  
; -----  
; Call OS to output string  
        mov    rax, SYS_write  
        mov    rsi, rdi  
        mov    edi, STDOUT  
        syscall  
        ; code for write  
        ; addr of character  
        ; file descriptor  
        ; count set above  
        ; system call  
  
; -----  
; String printed, return to calling routine  
prtDo  
pop  
*****
```

The `printString()` function is repeated in this example and is unchanged from the previous examples.

It must be noted that in order for this to work, the program should be assembled as usual, but linked with the GNU C compiler, either GCC or G++.

For example, assuming this example program is named cmdLine.asm, the assembly and linking would be as follows:

```
yasm -g dwarf2 -f elf64 cmdLine.asm -l cmdLine.lst  
gcc -g -o cmdLine cmdLine.o
```

*Note*, Ubuntu 18 will require the **no-pie** option on the gcc command as shown:

```
gcc -g -no-pie -o cmdLine cmdLine
```

If the standard linker is used, the arguments will not be passed in the correct map.

## 16.5 Exercise

Below are some quiz questions and suggested projects based on this chapter.

### 16.5.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) What software entity is responsible for parsing or reading the command arguments?
- 2) What software entity is responsible for verifying or checking the command arguments?
- 3) What is the first command line argument?
- 4) Explain what **argc** and **argv** refer to.
- 5) In an assembly language program, where is **argc** passed to the program?
- 6) In an assembly language program, where is **argv** passed to the program?
- 7) If seven spaces are entered between each of the command line arguments, will all the spaces removed when the command line arguments are checked?
- 8) If a number is expected as a command line argument, and the user enters “123456789” (an invalid value), is an error generated by the operating system (i.e., the loader)?

### 16.5.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Implement the example program to read and display the command line arguments. Use the debugger to execute the program and display the final results. Execute the program without the debugger and verify the appropriate output is displayed to the console.
- 2) Convert the command line example into a function that will display each of the command line arguments to the console. Use the debugger as necessary to debug the program. Execute the program without the debugger and verify the appropriate output is displayed to the console.
- 3) Create an assembly language program to accept a file name on the command line and open the file and display the one line message contained in the file. A series of small text files should be created each containing one of the very important messages.

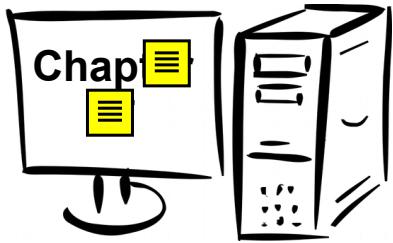
messages at the start of each chapter of this text. The program should perform error checking on the file name, and if valid, open the file. If the file opens successfully, the message should be read from the file displayed to the console. Appropriate error messages should be displayed if the file cannot be opened or read. The program may assume that each message will be < 256 characters. Use the debugger as necessary to debug the program. Execute the program without the debugger and verify the appropriate output is displayed to the console.

- 4) Create an assembly language program that will accept three unsigned integer numbers from the command line, add the three numbers, and display each of the three original numbers and the sum. If too many or too few command line arguments are provided, an error message should be displayed. This program will require that each of the ASCII strings be converted into integer. Appropriate error checking should be included. For example, “123” is correct while “12a3” is incorrect. The main program should call functions as necessary for the ASCII to integer conversion and the output. Use the debugger as necessary to debug the program. Execute the program without the debugger and verify the appropriate output is displayed to the console.



*If I had it all to do over again, I'd spell create with an "e".*

- Kernighan



## 17.0 Input/Output Buffering

This chapter provides information regarding Input/Output (I/O) buffering. I/O buffering is a process that improves the overall throughput and efficiency of I/O operations. The general topic of I/O buffering is quite broad and can encompass many things including hardware, software, and data networking. This text addresses single buffer buffering for reading information from a file. While this is a very specific application, the general concepts regarding how and why buffering is performed apply generally to many different applications.

High-level languages provide library functions that perform the I/O buffering and hide the associated complexity from the programmer. This is very desirable when programming. When working at a low-level, the buffering will need to be implemented explicitly (by us). The goal is to provide a detailed understanding of why and how buffering is performed in order to provide a more in-depth understanding of how a computer operates. This should help dispel notions that such I/O is magic. And, of course, help us appreciate the compiler I/O library functions.

### 17.1 Why Buffering

In order to fully understand the importance of buffering, it is useful to review the memory hierarchy as outlined in Chapter 2, Architecture. As noted, accesses to secondary storage are significantly more expensive in terms of run-time than accesses to main memory. This would strongly encourage us to limit the number of secondary storage accesses by using some temporary storage, referred to as a buffer, in main memory.

Such a buffer would also help reduce the overhead associated with system calls. For example, a file read system call would involve pausing our program and turning over control to the OS. The OS would validate our request (e.g., ensure file descriptor is valid) and then pass the request to the secondary storage control via the system bus.





controller would do whatever is necessary to obtain the requested data and place it directly into main memory location as instructed by the OS, again accessing the system bus to perform the transfer. *Note*, this is referred to as Direct Memory Access<sup>49</sup> (DMA). Once the secondary storage controller has completed the transfer, it notifies the OS. The OS will then notify and resume our program. This process represents system overhead since our program is just waiting for completion of the system service request. It would make sense to obtain more data rather than less data on each system service call. It should be fairly obvious that limiting the number of system service requests will help the overall performance of our program.



Additionally, as described in Chapter 13, System Services, the low-level file read and write operations require the number of characters to read. A typical operation desired by a programmer would be to read a line as is provided by high-level language functions such as the C++ *getline()* function. It is unknown ahead of time how many characters might be on that line. This makes the low-level file read operation more difficult since the exact number of characters to read is required.



We, as humans, place special meaning and significance to the LF character (end of line). In the computer, the LF is just another ASCII character, no different than any other. We see a file as a set of lines, but in the computer the file is just a collection of bytes, one after another.



The process used for reading interactive input involved reading one character at a time. This could be used to address the issue of not knowing ahead of time how many characters are on a line. As was done for interactive input, the program could just read one at a time until an LF is found and stop.



Such a process will be unduly slow for reading large volumes of information from a file. Interactive input expects the user to type characters. The computer is easily able to keep up with a human. Even the fastest typist will not be able to out-type the computer (assuming efficiently written code). Further, interactive input is generally limited to relatively small amounts. Input from a file is not awaiting user action and is generally not limited to small amounts of data. For a sufficiently large file, single character I/O would be quite slow. *Note*, testing and quantizing the performance difference is left as an exercise.



To address this, the input data will be buffered. In this context, we will read a chunk of the file, say 100,000 characters, into a large array which is referred to as a buffer or input buffer. When a “line” of text is required, the line will be obtained from the input buffer. The first time a line is requested, the file is read and the buffer is filled. A



---

49 For more information, refer to: [http://en.wikipedia.org/wiki/Direct\\_memory\\_access](http://en.wikipedia.org/wiki/Direct_memory_access)



the file read completes, the line is returned from the buffer (all characters up to and including the LF). Successive calls to get the next line are satisfied by obtaining the characters from the buffer without needing to read the file again. This can keep happening until all the characters in the buffer have been returned at which point the file will need to be read again. When there are no more characters, the input is completed and the program can be terminated.

This process helps improve performance, but is more complicated. High-level language functions, such as *getline*, hide this complexity from the user. When working at the low-level we will need to write the function to get the next line ourselves.

It is important to understand the cause of the performance degradation in order to fully understand why buffering will improve performance.

## 17.2 Buffering Algorit

The first step when developing an algorithm is to understand the problem. We will assume the file is already open and available for reading. For our buffering problem, we wish to provide a *myGetLine()* function to a calling routine.

The routine might be called as follo

```
status = myGetLine(fileDescriptor, textLine, MAXLEN)
```

The file opening and associated error checking is required but would be addressed separately and is outlined in Chapter 13, System Services. Once a file descriptor is obtained, it can be made available to the *myGetLine()* function. It is shown here as an explicit parameter for clarity.

As you may already be aware there is no special end of file code or character. We must infer the end of file based on the number of characters actually read. The file read system service will return the actual number of characters read. If the process requests 100,000 characters to be read and less than 100,000 are read, the end of file has been reached and all characters have been read. Further attempts at reading the file will generate an error. While it is easy to recognize the end of file and thus the last time the file needs to be read, the remaining characters in the buffer need to be processed. So while the program may know the end of file has been found, the program must continue until the buffer is depleted.

The calling routine should not need to know any of the details regarding the buffering, buffer management, or file operations. If a line of text from the file is available, the *myGetLine()* function will return the line from the file and a TRUE status. The line is returned via the passed argument. An argument for the maximum length of the text line is also provided to ensure that the text line array is not overwritten. If a line is not available, possibly due to all lines having been already returned or an unrecoverable error, the function will return FALSE.



read error, the function should return FALSE. An unrecoverable read error might occur if the storage medium goes off-line (drive removed), the file is deleted from another console, or a hardware error occurs. Such errors will make continued reading impossible. In such an event, an error message will be displayed and a FALSE returned which will halt the program (with incomplete results).

In general, the function would read characters, up to and including the LF, from the large buffer and place them in the line buffer, named **textLine** in the above example



This is easily done when there are characters available in the buffer. It will require an extra step to ensure that the buffer has characters. The file must be read if the buffer is empty, either due to the very first call or when all characters in the buffer have been returned. It is possible to check if characters are available by checking the current location of the next character to read in the buffer against the buffer size. The current location cannot be allowed to exceed the buffer size (or we would be accessing a location past the end of the buffer array). This is fairly straightforward when the buffer is filled entirely. The buffer may be partially filled due to a file size smaller than the buffer size or on the last file read after a series of file reads. To address this, the end of the buffer, initially set to the buffer size, must be set based on the actual number of characters read. There is a special case when the number of characters in the file is an exact multiple of the buffer size. If this occurs, the returned number of characters read is 0. An additional check is required to address this possibility.



Now that the problem and all the associated subtle issues are understood, we can take the next step of developing an algorithm. In general this process is typically iterative. That is, a first cut is developed, reviewed and refined. This would occur multiple times until a comprehensive algorithm is obtained. While some try to rush through this step, it is a mistake. Saving minutes on the algorithm development step means extra hours of debugging.



Based on this earlier example, the passed arguments include

```
; file descriptor → file descriptor for open file
;           as required for read system service
; text line → starting address of
; max length → maximum length of the array for
;           the text line
```



Some variables are required and defined as follows

```
; BUFFER_SIZE → parameter for size of buffer
; currIndex → index for current location
```



```
;      the buffer, initially set to BUFFER_SIZE
; buffMaximum → current maximum size of buffer
;           initially set to BUFFER_SIZE
; eofFlag → boolean to mark if the end of file
;           has been found, initially set to false
```

Based on an understanding of the problem, a number of different algorithmic approaches are possible. Using the arguments and local variables, one such approach is presented for reference.

```
; myGetLine(fileDescriptor, textLine, maxLength)
;   repeat {
;       if current index buffer maxi... [yellow]
;           read buffer (buffer si... [yellow]
;           if er... [yellow]
;               handle read er... [yellow]
;               display error mess... [yellow]
;               exit routine (with fal... [yellow]
;           reset point... [yellow]
;           if chars read < characters request r... [yellow]
;               set eofFlag = T... [yellow]
;               get one character from buffer at current in... [yellow]
;               place character in text line buf... [yellow]
;               increment current in... [yellow]
;               if character is... [yellow]
;                   exit with t... [yellow]
```

This algorithm outline does not verify that the text line buffer is not overwritten nor does it handle the case when the file size is an exact multiple of the buffer size. Refining, implementing, and testing the algorithm is left to the reader as an exercise.

As presented, this algorithm will require statically declared variables. Stack dynamic variables cannot be used here since the information is required to be maintained between successive calls.

The variables and algorithm are provided as program comments. The algorithm is the most important part of the program documentation and will not only help when writing the code but in the debugging process. Whatever time you think you are saving by skipping the documentation, it will cost far, far more time when debugging.

## 17.3 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 17.3.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) What is the end of line character or characters for Linux and Windows?
- 2) Based on the explanations in this chapter, what is I/O buffering?
- 3) In reference to a high-level language, where are the I/O buffering routines located?
- 4) What is the advantage of hiding the I/O buffering complexity from the programmer?
- 5) What is the key advantage of performing I/O buffering (as opposed to reading one character at a time)?
- 6) Why is it difficult to use the file read system service to read one text line from a file?
- 7) In terms of the memory hierarchy, why is buffering advantageous?
- 8) In terms of system overhead, why is buffering advantageous?
- 9) Why does the file read buffering algorithm require statically declared variables?
- 10) How is the end of file recognized by the program?
- 11) Provide one, of many, reasons that a file read request might return an error even when the file has been opened successfully.
- 12) What must be done to address the case when the file size is an exact multiple of the buffer size?
- 13) Why is the maximum length of the text line passed as an argument?
- 14) How does the presented algorithm ensure that the very first call to the `myGetLine()` function will read the buffer correctly?

### 17.3.2 Suggested Proj

Below are some suggested projects based on this chapter.

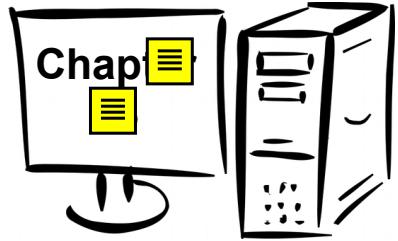
- 1) Refine the presented algorithm to address maximum text line check possibility that the file size is an even multiple of the buffer size. This includes formalizing the variable names and looping constructs (many possible choices).
- 2) Create a simple main program that will test the *myGetLine()* function by opening a file, calling the *myGetLine()* function, and displaying the lines to the console. Test the program with a series of different size files including ones smaller, larger, and much larger than the selected buffer size. Capture the program output and compare the captured output to the original file to ensure your program is correct.
- 3) Create a program that will read two file names from the command line, read lines from the first file, add line numbers, and write the modified line (with the line number) to the second file. For example, your add lines program might be initiated with:

```
./addLine inFile.txt newFile.
```

where the *inFile.txt* exists and contains standard ASCII text. If the *inFile.txt* file does not exist an error should be generated and the program terminated. The program should open/create the file *newFile.txt* and write the lines, with the line numbers, to the *newFile.txt* file. The output file should be created, deleting an old version if one exists. Use a text editor to verify that the line numbers track correctly in the output file.



*To err is human; to make a real mess, you need a computer.*



## 18.0 Floating-Point Instructions

This chapter provides a basic overview for a subset of the x86-64 floating-point instructions. Only the most basic instructions are covered.

The text focuses on the x86-64 floating-point operations, which are not the same as the 32-bit floating-point operations.

The instructions are presented in the following order:

- Data Movements
- Integer / Floating-point Conversion Instructions
- Arithmetic Instructions
- Floating-point Control Instructions

A complete listing of the instructions covered in this text is located in Appendix B for reference.

It should be noted that the floating-point arithmetic operations do not require the use of a specific register and do not change types (sizes). This makes the floating-point instructions easier to use.

### 18.1 Floating-Point Values

Floating-point values are typically represented as either single precision (32-bits) or double precision (64-bits). In C and C++ single precision floating-point variables are typically declared as **float** type and double precision floating-point variables are declared as **double** type. As noted in the following sections, assembly language instructions will use an **s** (lower-case letter S) qualifier to refer to single precision and a **d** (lower-case letter D) qualifier to refer to double precision.



## 18.2 Floating-Point Registers



There are a set of dedicated registers, referred to as XMM registers, used to support floating-point operations. Floating-point operations must use the floating-point registers. The XMM registers are 128-bits and 256-bits on the later processors. Initially, we will only use the lower 32 or 64-bits.



There are 16 XMM registers, named **xmm0** through **xmm15**. Refer to Chapter 2 for an explanation and summary of the CPU registers.



## 18.3 Data Movement



Typically, data must be moved into a CPU floating-point register in order to be operated upon. Once the calculations are completed, the result may be copied from the register and placed into a variable. There are a number of simple formulas in the example program that perform these steps. This basic data movement operations are performed with the move instruction.



The general form of the move instruction:



```
movss    <dest>, <src>
movsd    <dest>, <src>
```



For the **movss** instruction, a single 32-bit source operand is copied into the destination operand. For the **movsd** instruction, a single 64-bit source operand is copied into the destination operand. The value of the source operand is unchanged. The destination and source operand must be of the correct size for the instruction (32 or 64-bits). Neither operand can be an immediate value. Both operands, cannot be memory, however one can be. If a memory to memory operation is required, two instructions must be used.



The move instruction loads one value, using the lower 32 or 64-bits, into or out of the register. Other move instructions are required to load multiple values.



The floating-point move instructions are summarized as follows:



Instruction	Explanation
<b>movss</b> <dest>, <src>	Copy 32-bit source operand to the 32-bit destination operand. <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , operands cannot be an immediate value.



Instruction	Explanation
Examples:	<pre>movss    xmm0, dword [x] movss    dword [f1tSVar], x movss    xmm3, xmm2</pre>
<b>movsd &lt;dest&gt;, &lt;src&gt;</b>	<p>Copy 64-bit source operand to the 64-bit destination operand.</p> <p>Note 1, both operands cannot be memory locations.</p> <p>Note 2, operands cannot be an immediate value.</p>
Examples:	<pre>movsd    xmm0, qword [y] movsd    qword [f1tDVar], x movsd    xmm3, xmm2</pre>

A more complete list of the instructions is located in Appendix A.

For example, assuming the following data declarations:

```
fSVar1 dd ?  
fSVar2 dd ?  
fDVar1 dq ?  
fDVar2 dq ?
```

To perform the basic operations:

```
fSVar2 = fSVar2 ; single precision variable  
fDVar2 = fDVar1 ; double precision variable
```

The following instructions could be used:

```
movss    xmm0, dword [fSVar1] ; fSVar2 = fSVar1  
movss    dword [fSVar2], xmm0  
  
movsd    xmm1, qword [fDVar1] ; fDVar2 = fDVar1  
movsd    qword [fDVar2], xmm1  
  
movss    xmm2, xmm0 ; xmm2 = xmm0 (32-bit)  
movsd    xmm3, xmm1 ; xmm3 = xmm1 (64-bit)
```

For some instructions, including those above, the explicit type specification (e.g., *byte*, *word*, *dword*, *qword*) can be omitted as the other operand will clearly define the size. It is included for consistency and good programming practices.



## 18.4 Integer / Floating-Point Conversion Instructions



If integer values are required during floating-point calculations, the integers must be converted into floating-point values. If single precision and double precision floating-point values are required for a series of calculations, they must be converted to single or double so that the operations are performed on a consistent size/type.



Refer to Chapter 3 for a more detailed explanation of the representation details for floating-point values. It is assumed the reader understands the representation details and recognizes the requirement to ensure consistent formats before performing floating-point operations.



This basic data conversion operations are performed with the convert instructions.



The floating-point conversion instructions are summarized as follows:

Instruction	Explanation
<b>cvtss2sd</b> <RXdest>, <src>	Convert 32-bit floating-point source operand to the 64-bit floating-point destination operand. <i>Note 1</i> , destination operand must be floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtss2sd    xmm0, dword [fltsVar] cvtss2sd    xmm3, eax cvtss2sd    xmm3, xmm2</pre>
<b>cvtsd2ss</b> <RXdest>, <src>	Convert 64-bit floating-point source operand to the 32-bit floating-point destination operand. <i>Note 1</i> , destination operand must be floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtsd2ss    xmm0, qword [fldDVar] cvtsd2ss    xmm1, rax cvtsd2ss    xmm3, xmm2</pre>



Instruction	Explanation
<b>cvtss2si</b> <reg>, <src>	Convert 32-bit floating-point source operand to the 32-bit integer destination operand. <i>Note 1</i> , destination operand must be register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtss2si    xmm1, xmm0 cvtss2si    eax, xmm0 cvtss2si    eax, dword [fltSV]</pre>
<b>cvtsd2si</b> <reg>, <src>	Convert 64-bit floating-point source operand to the 32-bit integer destination operand. <i>Note 1</i> , destination operand must be register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtsd2si   xmm1, xmm0 cvtsd2si   eax, xmm0 cvtsd2si   eax, qword [fltDV]</pre>
<b>cvtsi2ss</b> <RXdest>, <src>	Convert 32-bit integer source operand to the 32-bit floating-point destination operand. <i>Note 1</i> , destination operand must be floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtsi2ss   xmm0, eax cvtsi2ss   xmm0, dword [fltDV]</pre>
<b>cvtsi2sd</b> <RXdest>, <src>	Convert 32-bit integer source operand to the 64-bit floating-point destination operand. <i>Note 1</i> , destination operand must be floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<pre>cvtsi2sd   xmm0, eax cvtsi2sd   xmm0, dword [fltDV]</pre>

A more complete list of the instructions is located in Appendix

## 18.5 Floating-Point Arithmetic Instructions

The floating-point arithmetic instructions perform arithmetic operations such as add, subtract, multiplication, and division on single or double precision floating-point values. The following sections present the basic arithmetic operations.

### 18.5.1 Floating-Point Addition

The general form of the floating-point addition instructions are as follows:

```
addss      <RXdest>, <s>
addsd      <RXdest>, <s>
```

Where operation is as follows:

$$<\text{RXdest}> = <\text{RXdest}> + <\text{s}>$$

Specifically, the source and destination operands are added and the result is placed in the destination operand (over-writing the previous value). The destination operand must be a floating-point register. The source operand may not be an immediate value. The value of the source operand is unchanged. The destination and source operand must be of the same size (double-words or quadwords). If a memory to memory addition operation is required, two instructions must be used.

For example, assuming the following data declarations:

<b>fSNum1</b>	<b>dd</b>	42
<b>fSNum2</b>	<b>dd</b>	1
<b>fSAns</b>	<b>dd</b>	1
<b>fDNum3</b>	<b>dq</b>	200
<b>fDNum4</b>	<b>dq</b>	72
<b>fDAns</b>	<b>dq</b>	2

To perform the basic operations:

```
fSAns = fSNum1 + fSNum2
fDAns = fDNum3 + fDNum4
```

The following instructions could be used:

```
; fSAns = fSNum1 + fSNum2
movss    xmm0, dword [fSNum1]
addss    xmm0, dword [fSNum2]
movss    dword [fSAns], xmm0
```

```

; fDAns = fDNum3 + fDNum4
movsd    xmm0, qword [fDNum3]
addsd    xmm0, qword [fDNum4]
movsd    qword [fDAns], xmm0

```

For some instructions, including those above, the explicit type specification (e.g., *dword*, *qword*) can be omitted as the other operand or the instruction itself clearly defines the size. It is included for consistency and good programming practices.

The floating-point addition instructions are summarized as follows:

Instruction	Explanation
<b>addss</b> <RXdest>, <src>	Add two 32-bit floating-point operands, (<RXdest> + <src>) and place the result in <RXdest> (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>addss</b> xmm0, xmm3 <b>addss</b> xmm5, dword [fSV]
<b>addsd</b> <RXdest>, <src>	Add two 64-bit floating-point operands, (<RXdest> + <src>) and place the result in <RXdest> (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>addsd</b> xmm0, xmm3 <b>addsd</b> xmm5, qword [fDV]

A more complete list of the instructions is located in Appendix E.

### 18.5.2 Floating-Point Subtraction

The general form of the floating-point subtraction instructions are as follows:

```

subss    <RXdest>, <src>
subsd    <RXdest>, <src>

```

## Chapter 18.0 ◀ Floating-Point Instructions

Where operation is as follows

$$\text{<RXdest>} = \text{<RXdest>} - \text{<s>}$$

Specifically, the source and destination operands are subtracted and the result is placed in the destination operand (over-writing the previous value). The destination operand must be a floating-point register. The source operand may not be an immediate value. The value of the source operand is unchanged. The destination and source operand must be of the same size (double-words or quadwords). If a memory to memory addition operation is required, two instructions must be used.

For example, assuming the following data declarations

<b>fSNum1</b>	<b>dd</b>	42	
<b>fSNum2</b>	<b>dd</b>	1	
<b>fSAns</b>	<b>dd</b>		
<b>fDNum3</b>	<b>dq</b>	200	
<b>fDNum4</b>	<b>dq</b>	72	
<b>fDAns</b>	<b>dq</b>		

To perform the basic operations

$$\begin{aligned} \text{fSAns} &= \text{fSNum1} - \text{fSNum2} \\ \text{fDAns} &= \text{fDNum3} - \text{fDNum4} \end{aligned}$$

The following instructions could be used

```
; fSAns = fSNum1 - fSNum2
movss      xmm0, dword [fSNum1]
subss      xmm0, dword [fSNum2]
movss      dword [fSAns], xmm0

; fDAns = fDNum3 - fDNum4
movsd      xmm0, qword [fDNum3]
subsd      xmm0, qword [fDNum4]
movsd      qword [fDAns], xmm0
```

For some instructions, including those above, the explicit type specification (e.g., *dword*, *qword*) can be omitted as the other operand or the instruction itself clearly defines the size. It is included for consistency and good programming practices.

The floating-point subtraction instructions are summarized as follows

Instruction	Explanation
<b>subss</b> <RXdest>, <src>	<p>Subtract two 32-bit floating-point operands, (<b>&lt;RXdest&gt;</b> - <b>&lt;src&gt;</b>) and place the result in <b>&lt;RXdest&gt;</b> (over-writing previous value).</p> <p><i>Note 1</i>, destination operands must be a floating-point register.</p> <p><i>Note 2</i>, source operand cannot be an immediate.</p>
Examples:	<b>subss xmm0, xmm3</b> <b>subss xmm5, dword [fSV]</b>
<b>subsd</b> <RXdest>, <src>	<p>Subtract two 64-bit floating-point operands, (<b>&lt;RXdest&gt;</b> - <b>&lt;src&gt;</b>) and place the result in <b>&lt;RXdest&gt;</b> (over-writing previous value).</p> <p><i>Note 1</i>, destination operands must be a floating-point register.</p> <p><i>Note 2</i>, source operand cannot be an immediate.</p>
Examples:	<b>subsd xmm0, xmm3</b> <b>subsd xmm5, qword [fDV]</b>

A more complete list of the instructions is located in Appendix E.

### 18.5.3 Floating-Point Multiplication

The general form of the floating-point multiplication instructions are as follows:

**mulss** <RXdest>, <src>  
**mulsd** <RXdest>, <src>

Where operation is as follows:

$$\text{<RXdest>} = \text{<RXdest>} * \text{<src>}$$

Specifically, the source and destination operands are multiplied and the result is placed in the destination operand (over-writing the previous value). The destination operand must be a floating-point register. The source operand may not be an immediate value. The value of the source operand is unchanged. The destination and source operand must be of the same size (double-words or quadwords). If a memory to memory addition operation is required, two instructions must be used.

## Chapter 18.0 ◀ Floating-Point Instructions

For example, assuming the following data declarations:

fSNum1	dd	42.1
fSNum2	dd	1.234
fSAns	dd	
fDNum3	dq	200.732
fDNum4	dq	2.2
fDAns	dq	

To perform the basic operations:

```
fSAns = fSNum1 * fSNum2  
fDAns = fDNum3 * fDNum4
```

The following instructions could be used:

```
; fSAns = fSNum1 * fSNum2  
movss      xmm0, dword [fSNum1]  
mulss      xmm0, dword [fSNum2]  
movss      dword [fSAns], xmm0  
  
; fDAns = fDNum3 * fDNum4  
movsd      xmm0, qword [fDNum3]  
mulsd      xmm0, qword [fDNum4]  
movsd      qword [fDAns], xmm0
```

For some instructions, including those above, the explicit type specification (e.g., *dword*, *qword*) can be omitted as the other operand or the instruction itself clearly defines the size. It is included for consistency and good programming practices.

The floating-point multiplication instructions are summarized as follows:

Instruction	Explanation
<b>mulss &lt;RXdest&gt;, &lt;src&gt;</b>	Multiply two 32-bit floating-point operands, ( <i>&lt;RXdest&gt;</i> * <i>&lt;src&gt;</i> ) and place the result in <i>&lt;RXdest&gt;</i> (over-writing previous value). Note 1, destination operands must be a floating-point register. Note 2, source operand cannot be an immediate.

Instruction	Explanation
Examples:	<code>mulss xmm0, xmm3</code> <code>mulss xmm5, dword [fSV]</code>
<code>mulsd &lt;RXdest&gt;, &lt;src&gt;</code>	Multiply two 64-bit floating-point operands, ( <code>&lt;RXdest&gt;</code> * <code>&lt;src&gt;</code> ) and place the result in <code>&lt;RXdest&gt;</code> (over-writing previous value). Note 1, destination operands must be a floating-point register. Note 2, source operand cannot be an immediate.
Examples:	<code>mulsd xmm0, xmm3</code> <code>mulsd xmm5, qword [fDV]</code>

A more complete list of the instructions is located in Appendix E.

#### 18.5.4 Floating-Point Division

The general form of the floating-point division instructions are as follows:

`divss <RXdest>, <src>`  
`divsd <RXdest>, <src>`

Where operation is as follows:

$$<\text{RXdest}> = <\text{RXdest}> / <\text{src}>$$

Specifically, the source and destination operands are divided and the result is placed in the destination operand (over-writing the previous value). The destination operand must be a floating-point register. The source operand may not be an immediate value. The value of the source operand is unchanged. The destination and source operand must be of the same size (double-words or quadwords). If a memory to memory addition operation is required, two instructions must be used.

For example, assuming the following data declarations:

<code>fSNum1</code>	<code>dd</code>	42
<code>fSNum2</code>	<code>dd</code>	1
<code>fSAns</code>	<code>dd</code>	
<code>fDNum3</code>	<code>dq</code>	200
<code>fDNum4</code>	<code>dq</code>	72
<code>fDAns</code>	<code>dq</code>	2

## Chapter 18.0 ◀ Floating-Point Instructions

To perform the basic operations

```
fSAns = fSNum1 / fSNum2  
fDAns = fDNum3 / fDNum4
```

The following instructions could be used

```
; fSAns = fSNum1 / fSNum2  
movss      xmm0, dword [fSNum1]  
divss      xmm0, dword [fSNum2]  
movss      dword [fSAns], xmm0  
  
; fDAns = fDNum3 / fDNum4  
movsd      xmm0, qword [fDNum3]  
divsd      xmm0, qword [fDNum4]  
movsd      qword [fDAns], xmm0
```

For some instructions, including those above, the explicit type specification (e.g., *dword*, *qword*) can be omitted as the other operand or the instruction itself clearly defines the size. It is included for consistency and good programming practices.

The floating-point division instructions are summarized as follows

Instruction	Explanation
<b>divss</b> <RXdest>, <src>	Divide two 32-bit floating-point operands, (<RXdest> / <src>) and place the result in <RXdest> (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>divss</b> xmm0, xmm3 <b>divss</b> xmm5, dword [fSV]
<b>divsd</b> <RXdest>, <src>	Divide two 64-bit floating-point operands, (<RXdest> / <src>) and place the result in <RXdest> (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.

Instruction	Explanation
Examples:	<code>divsd xmm0, xmm3 divsd xmm5, qword [fDV]</code>

A more complete list of the instructions is located in Appendix E.

### 18.5.5 Floating-Point Square Root

The general form of the floating-point square root instructions are as follows:

`sqrtss <RXdest>, <src>  
sqrtsd <RXdest>, <src>`

Where operation is as follows:

$$<\text{dest}> = \sqrt{<\text{src}>}$$

Specifically, the square root of the source operand is placed in the destination operand (over-writing the previous value). The destination operand must be a floating-point register. The source operand may not be an immediate value. The value of the source operand is unchanged. The destination and source operand must be of the same size (double-words or quadwords). If a memory to memory addition operation is required, two instructions must be used.

For example, assuming the following data declarations:

<code>fSNum1</code>	<code>dd</code>	<code>121</code>
<code>fSAns</code>	<code>dd</code>	
<code>fDNum3</code>	<code>dq</code>	<code>172935.</code>
<code>fDAns</code>	<code>dq</code>	

To perform the basic operations:

$$\begin{aligned} fSAns &= \sqrt{fSNum1} \\ fDAns &= \sqrt{fDNum3} \end{aligned}$$

The following instructions could be used:

```
; fSAns = sqrt (fSNum1)
sqrtss      xmm0, dword [fSNum1]
movss       dword [fSAns], xmm0
```

## Chapter 18.0 ◀ Floating-Point Instructions

```
; fDAns = sqrt(fDNum3)
sqrtsd    xmm0, qword [fDNum3]
movsd    qword [fDAns], xmm0
```

For some instructions, including those above, the explicit type specification (e.g., *dword*, *qword*) can be omitted as the other operand or the instruction itself clearly defines the size. It is included for consistency and good programming practices.

The floating-point addition instructions are summarized as follows:

Instruction	Explanation
<b>sqrtss</b> <RXdest>, <src>	Take the square root of the 32-bit floating-point source operand and place the result in destination operand (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>sqrtss</b> xmm0, xmm3 <b>sqrtss</b> xmm7, dword [fSV]
<b>sqrtsd</b> <RXdest>, <src>	Take the square root of the 64-bit floating-point source operand and place the result in destination operand (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>sqrtsd</b> xmm0, xmm3 <b>sqrtsd</b> xmm7, qword [fDV]

A more complete list of the instructions is located in Appendix E.

## 18.6 Floating-Point Control Instructions

The control instructions refer to programming constructs such as IF statements and looping. The integer comparison instruction, **cmp**, as described in Chapter 7 will not work for floating-point values.

The floating-point comparison instructions compare two floating-point values. Like the integer comparisons, the result of the comparison is stored in the **rFlag** register and neither operand is changed. Immediately after the comparison, the **rFlag** register is accessed to determine the results by using a conditional jump instruction. While all floating-point comparisons are signed, an unsigned conditional jump is used (**ja/jae/jb/jbe**). Program labels (i.e., a target of a conditional jump) are the same.

There are two forms of the floating-point comparison, ordered and unordered. The ordered floating-point comparisons can cause a number of exceptions. The unordered floating-point comparisons can only cause an exception for a *S-NaN* (signaling not a number), more generically referred to as a *Nan* (not a number) as described in Chapter 3, Data Representation.

The GNU C/C++ compiler favors the unordered floating-point compare instruction. Since they are similar, this text will focus only on the unordered version.

### 18.6.1 Floating-Point Comparison Instructions

The general form of the floating-point comparison instructions are as follows:

```
ucomiss    <RXsrc>, <src>
ucomisd    <RXsrc>, <src>
```

Where **<RXsrc>** and **<src>** are compared as floating-point values and must be the same size. The results of the comparison are placed in the **rFlag** register. Neither operand is changed. The **<RXsrc>** operand must be one of the **xmm** registers. The **<src>** register can be a **xmm** register or a memory location, but may not be an immediate value. One of the unsigned conditional jump instructions can be used to read the **rFlag** register.

The conditional control instructions include the jump equal (**je**) and jump not equal (**jne**). The unsigned conditional control instructions include the basic set of comparison operations; jump below than (**jb**), jump below or equal (**jbe**), jump above than (**ja**), and jump above or equal (**jae**).

The general form of the signed conditional instructions along with an explanatory comment are as follows:

```
je      <label>          ; if <op1> == <op2>
jne     <label>          ; if <op1> != <op2>
```

## Chapter 18.0 ◀ Floating-Point Instructions

```
jb    <label>          ; unsigned, if <op1> < <op2>
jbe   <label>          ; unsigned, if <op1> <= <op2>
ja    <label>          ; unsigned, if <op1> > <op2>
jae   <label>          ; unsigned, if <op1> >= <op2>
```

For example, given the following pseudo-code for floating-point variables:

```
if (fltNum > fltMax)
    fltMax = fltNum
```

And, given the following data declarations:

```
fltNum      dq
fltMax      dq      5
```

Then, assuming that the values are updating appropriately within the program (not shown), the following instructions could be used:

```
movsd      xmm1, qword [fltNum]
ucomisd    xmm1, qword [fltMax]      ; if fltNum <= fltMax
jbe       notNewFltMax             ; skip set new
movsd      qword [fltMax], xmm1
notNewFltMax
```

As with integer comparisons, the floating-point compare and conditional jump provide functionality for jump or not jump. As such, if the condition from the original IF statement is false, the code to update the **fltMax** should not be executed. Thus, when false, in order to skip the execution, the conditional jump will jump to the target label immediately following the code to be skipped (not executed). While there is only one line in this example, there can be many lines of code.

A more complex example might be as follows:

```
if (x != 0.0)
    ans = x / y;
    errFlg = FALSE
} else
    errFlg = TRUE
```

This basic compare and conditional jump does not provide an IF-ELSE structure. It must be created. Assuming the **x** and **y** variables are signed double-words that will be set during the program execution, and the following declarations:

TRUE	equ	
FALSE	equ	
fltZero	dq	
x	dq	
y	dq	
ans	dq	
errFlg	db	

The following code could be used to implement the above IF-ELSE statement:

```

movsd      xmm1, qword [ ]           ; stack
ucomisd   xmm1, qword [fltZero]      ; if statement
je         doE, [ ]                 ; stack
divsd      xmm1, qword [ ]           ; stack
movsd      dword [ans], [ ]          ; stack
mov        byte [errFlg], FA        ; stack
jmp       skipE, [ ]                ; stack

doEl      mov    byte [errFlg], T  ; stack
skipEl    [ ]                      ; stack

```

Floating-point comparisons can be very tricky due to the inexact nature of the floating-point representations and rounding errors. For example, the value 0.1 added 10 times should be 1.0. However, implementing a program to perform this summation and checking the result, will show that;

0.1 ≠

which can be very confusing for inexperienced programmers.

For more information regarding the details of floating-point, refer to the popular article *What Every Computer Scientist Should Know About Floating-Point Arithmetic*<sup>50</sup>.

The floating-point comparison instructions are summarized as follows:

---

50 See: [http://docs.oracle.com/cd/E19957-01/806-3568/ncg\\_goldberg.html](http://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html)



Instruction	Explanation
<b>ucomiss</b> <RXsrc>, <src>	Compare two 32-bit floating-point operands, (<RXsrc> + <src>). Results are placed in the <b>rFlag</b> register. Neither operand is changed. <i>Note 1</i> , <RXsrc> operands must be a floating-point register. <i>Note 2</i> , source operand may be a floating-point register or memory, but not be an immediate.
Examples:	<b>ucomiss</b> xmm0, xmm3 <b>ucomiss</b> xmm5, dword [fSV]
<b>ucomisd</b> <RXsrc>, <src>	Compare two 64-bit floating-point operands, (<RXsrc> + <src>). Results are placed in the <b>rFlag</b> register. Neither operand is changed. <i>Note 1</i> , <RXsrc> operands must be a floating-point register. <i>Note 2</i> , source operand may be a floating-point register or memory, but not be an immediate.
Examples:	<b>ucomisd</b> xmm0, xmm3 <b>ucomisd</b> xmm5, qword [fSV]

A more complete list of the instructions is located in Appendix E.



## 18.7 Floating-Point Calling Convention



The standard calling conventions detailed in Chapter 12, Functions still fully apply. This section addresses the usage of the floating-point registers when calling floating-point functions.



When using floating-point registers, none of the registers are preserved across a floating-point function call.



The first eight (8) floating-point arguments are passed in floating-point registers **xmm0** – **xmm7**. Any additional arguments are placed on the stack in backwards order in the manner described in Chapter 12, Functions. A value returning floating-point function will return the result in **xmm0**.



Since none of the floating-point registers are preserved, the code must be written carefully.



## 18.8 Example Program, Sum and Average

This example is a simple assembly language program to calculate the sum and average for a list of floating-point values.

```
; Floating-Point Example Program
; ****
section    .data
; -----
; Define constants
NULL        equ 0          ; end of string
TRUE         equ 1
FALSE        equ 0
EXIT_SUCCESS equ 0          ; Successful operation
SYS_exit     equ 60         ; system call code for termination
; --
fltLst      dq  21.34,   6.15,   9.12, 10.05,  7.5
             dq  1.44, 14.50,   3.32, 75.71, 11.1
             dq 17.23, 18.25, 13.65, 24.24,  8.8
length       dd  ?
lstSum       dq  ?
lstAve       dq  ?
; ****
section    .text
global _start
_start:    mov rbp, [esp+4]
; -----
; Loop to find floating-point sum
            mov ecx, [length]
            mov rbx, fltLst
            mov rsi, 0
            movsd  xmm1, qword [lstS]
sum:        movsd  xmm0, qword [rbx+rsi*8]           ; get fltLst
```

## Chapter 18.0 ◀ Floating-Point Instructions

```
addsd      xmm1, xmm0
inc        rsi
loop       sub [list icon]
movsd      qword [1stSum], xmm1
; update [list icon]
; save [list icon]

; -----
; Compute average of entire list [list icon]

cvtsi2sd  xmm0, dword [leng]
cvtsd2si  dword [length], xmm0
divsd     xmm1, xmm0
movsd      qword [1stAve], xmm1
; -----
; Done, terminate program [list icon]

la [list icon]
    mov     rax, SYS_exit
    mov     rbx, EXIT_SUCCESS
    sysc   [list icon]
; exit w/succ [list icon]
```

The debugger can be used to examine the results and verify correct execution of the program.

### 18.9 Example Program, Absolute Value

This example is a simple assembly language program to find the absolute value of a floating-point value in order to demonstrate floating-point comparisons. Recall that if a value is negative, it must be made positive and if the value is already positive, nothing should be done.

```
; Floating-Point Absolute Value Example [list icon]

section    .data
; -----
; Define constants [list icon]
TRUE       equ 1
FALSE      equ 0
```

```

SUCCESS      equ     0          ; successful operation
SYS_exit    equ     60         ; call code for termination

; -----
; Define some test variables
dZero        dq     0
dNegOne     dq     -1
fltVal       dq     -8.0

; ****
section      .text
global _start
_start:      db 0x43

; -----
; Perform absolute value function on fltVal
movsd        xmm0, qword [fltVal]
ucomisd     xmm0, qword [dZero]
jae          isP
mulsd        xmm0, qword [dNegOne]
movsd        qword [fltVal], xmm0
isP:         db 0x4C

; -----
; Done, terminate program
la         db 0x4E
mov        rax, SYS_exit
mov        rbx, EXIT_SUCCESS
sysc      db 0x4F
; exit w/success

```

In this example, the final result for `|fltVal|` was saved to memory. Depending on the context, this may not be required.

## 18.10 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 18.10.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) List the floating-point registers.
- 2) How many bytes are used by single precision floating-point values and how many bytes are used for double precision floating-point values?
- 3) Explain why 0.1 added 10 times does not equal 1.0.
- 4) Where is the result of a value returning floating-point function such as `sin` returned?
- 5) For a value returning floating-point function, which floating-point registers need to be preserved across the function call?

### 18.10.2 Suggested Projects

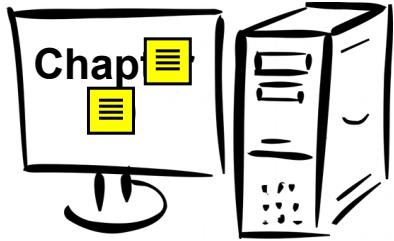
Below are some suggested projects based on this chapter.

- 1) Implement the example program to find the sum and average for a list of floating-point values. Use the debugger to execute the program and verify the final results.
- 2) Implement the floating-point absolute value function as two macros, one `fAbs` for 32-bit floating-point values and `fAbsd` for 64-bit floating-point values. Create a simple main program that uses each macro three times on various different values. Use the debugger to execute the program and verify the results.
- 3) Implement a program to perform the summation of floating-point numbers.



Compare the results of the summation to the value 1.0 and display a message “Are Same” if the summation result equals 1.0 and the message “Are Not Same” if the result of the summation does not equal 1.0. Use the debugger as needed to debug the program. When working, execute the program without the debugger and verify that the expected results are displayed to the console.

`#define QUESTION ((bb) || !(`  
- Shakespeare



## 19.0 Parallel Processing

In the context of computing, parallel processing<sup>51</sup>, or more generically concurrency<sup>52</sup> refers to multiple processes appearing to execute simultaneously.

In broad terms, concurrency implies multiple different (not necessarily related) processes simultaneously making progress. This can be accomplished in multiple ways. For example, process A could be executing on core 0 while process B is simultaneously executing on core 1 thus executing in parallel. Another possibility is that the execution of process A and B could be interleaved on a single core where process A might execute for a period of time and then be paused while process B is then executed for a period of time and then paused while process A is resumed. This continues until either or both are completed. If this interleaved execution is performed often and fast enough it will appear that each is executing simultaneously.

The term parallel processing implies that processes are executing simultaneously. These processes may be unrelated or working together as a coordinated unit to solve a single complex problem.

This chapter provides an overview of some basic approaches to parallel processing as applied to a single problem. Assume that a large problem can be divided into various sub-problems and those sub-problems can be executed independently. The sub-solutions would be brought together to provide a final solution for the problem. If the sub-problems can be executed simultaneously, a final solution might be found significantly faster than if the sub-problems are executed in series (one after another). The potential rewards for implementing concurrency are significant but so are the challenges. Unfortunately, not all problems can be easily divided into such sub-problems.

<sup>51</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Parallel\\_processing](http://en.wikipedia.org/wiki/Parallel_processing)

<sup>52</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Concurrency\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Concurrency_(computer_science))



The basic approaches to parallel processing are distributed computing<sup>53</sup> and multiprocessing<sup>54</sup> (also referred to as threaded computations). Each approach is explained with an emphasis on the technical issues associated with shared memory multiprocessing. The larger topic of learning to create parallel algorithms is outside the scope of this text.



## 19.1 Distributed Computing



Distributed computing or distributed processing refers to the general idea of taking a large problem, dividing it into multiple sub-problems, and executing the sub-problems on different computers connected through a network. Typically, a master or main server node will distribute the sub-problems to various available computation nodes. When completed, the computational nodes will send the intermediate results back to the master. As needed, the master may send additional sub-problems to the computational nodes. The master will also track and combine the intermediate results into a final solution. The details of how this is actually performed vary significantly and are directly related to the specifics of the problem.



There are many large scale examples of distributed computing projects<sup>55</sup>. One such project is Folding@home<sup>56</sup> which is a large distributed computing project for disease research that simulates protein folding, computational drug design, and other types of molecular dynamics. The project uses the idle processing resources of well over 100,000 different personal computers owned by volunteers who have installed the software on their systems.



This approach has the advantage of scaling to very large number of distributed computers. The disadvantages are associated with the communication limitations associated with the network.



## 19.2 Multiprocessing



As noted in Chapter 2, Architecture Overview, most current CPU chips include multiple cores. The CPU cores all have equal access to the main memory resource. Multiprocessing is a form of parallel processing that specifically refers to using multiple cores to perform simultaneous execution of multiple processes.



<sup>53</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Distributed\\_computing](http://en.wikipedia.org/wiki/Distributed_computing)

<sup>54</sup> For more information, refer to: <http://en.wikipedia.org/wiki/Multiprocessing>

<sup>55</sup> For more information, refer to: [http://en.wikipedia.org/wiki/List\\_of\\_distributed\\_computing\\_projects](http://en.wikipedia.org/wiki/List_of_distributed_computing_projects)

<sup>56</sup> For more information, refer to: <http://en.wikipedia.org/wiki/Folding@home>



Focusing on a single large project, a main or initial process might generate subprocesses, referred to as threads<sup>57</sup>. When the initial process is started and placed in memory by the system loader<sup>58</sup>, the Operating System creates a new process including the required Operating System data structures, memory allocations, and page/swap file allotments. A thread is often referred to as a light-weight process since it will use the initial process's allocations and address space thus making it quicker to create. It will also be quicker to terminate as the deallocations are not performed. Since the threads share memory with the initial process and any other threads, this presents the potential for very fast communication between simultaneously executing threads. It also has the added requirement that the communications in the form of memory writes be carefully coordinated to ensure results are not corrupted. Such a problem is referred to as a race condition<sup>59</sup>. Race conditions are addressed in the following section.

The threaded approach will not scale to the extent of the distributed approach. CPU chips have a limited number of cores which places an upper bound on the number of simultaneous computations that can occur. This limit does not apply to distributed computing. The limitations regarding network communication speeds do not apply to threaded computations.

### 19.2.1 POSIX Threads

POSIX Threads<sup>60</sup>, commonly called pThreads, is a widely available thread library on Ubuntu and many other operating systems. The example in this section will use the pThreads thread library Application Programmer Interface (API) for creating and joining threads.

The initial or main process is created during the load process. The main process may create additional threads with the `pthread_create()` library function. While there is no specific limit to the number of threads that can be created, there is a limit to the number of cores available thus creating a practical limit for the number of threads.

The initial or main process can see if the thread has completed with the `pthread_join()` library function. If the thread has not completed, the join function will wait until it does. Ideally, in order to maximize the overall parallel operations, the main process would perform other computations while the thread or threads are executing and only check for thread completion when the other work has been completed.

<sup>57</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Thread\\_\(computing\)](http://en.wikipedia.org/wiki/Thread_(computing))

<sup>58</sup> For more information, refer to Chapter 5, Tool Chain

<sup>59</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Race\\_condition](http://en.wikipedia.org/wiki/Race_condition)

<sup>60</sup> For more information, refer to: [http://en.wikipedia.org/wiki/POSIX\\_Threads](http://en.wikipedia.org/wiki/POSIX_Threads)

There are other pThread functions to address mutexes<sup>61</sup>, condition variables, and synchronization<sup>62</sup> between threads which are not addressed in this very brief overview.

It should be noted that there are other approaches to threading not addressed here.

### 19.2.2 Race Condition

A race condition is a generic term referring to when multiple threads simultaneously write to the same location at the same time. Threaded programming is typically done in a high-level language. However, fully understanding the specific cause of a race condition is best done at the assembly language level.

A simple example is provided to purposely create a race condition and examine the problem in detail. This example is not computationally useful.

Assuming we wish to compute the following formula MAX times, where MAX is a defined constant. For example;

$$\text{myValue} \leftarrow \left( \frac{\text{myValue}}{X} \right) + Y$$

We could write a high-level language program something along the lines

```
for (int i=0; i < MAX; i++)
    myValue = (myValue / X) + Y
```

If we wish to speed-up this computation, perhaps because MAX is extremely large, we might create two thread functions, each to perform MAX/2 computations. Each of the thread functions would have shared access to the variables **myValue**, **X**, and **Y**. Thus the code might be;

```
for (int i=0; i < MAX/2; i++)
    myValue = (myValue / X) + Y
```

This code would be repeated in each of the thread functions. It may not be obvious, but assuming both threads are simultaneously executing, this will cause a race condition on the **myValue** variable. Specifically, each of the two threads are attempting to update the variable simultaneously and some of the updates to the variable may be lost.

<sup>61</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Mutual\\_exclusion](http://en.wikipedia.org/wiki/Mutual_exclusion)

<sup>62</sup> For more information, refer to: [http://en.wikipedia.org/wiki/Synchronization\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Synchronization_(computer_science))

To further simplify this example, we will assume that *X* and *Y* are both set to 1. As such, the result of each calculation would be to increment *myValue* by 1. If *myValue* is initialized to 0, repeating the calculation MAX times, should result in *myValue* being set to MAX. This simplification will allow easy verification of the results. It is not computationally useful in any meaningful way.

Implementing this in assembly would first require a main program that creates each of the two threads. Then, each of the thread functions would need to be created. In this very simple example, each thread function would be the same (performing MAX/2 iterations and that MAX is an even number).

Assuming one of the thread functions is named *threadFunction0()* and given the below pThread thread data structure;

```
pthreadID0      dd      0, 0, 0, 0,
```

The following code fragment would create and start *threadFunction0()* executing

```
; pthread_create(&pthreadID0, NULL,  
;                           &threadFunction0, NULL)  
    mov    rdi, pthreadID0  
    mov    rsi, NULL  
    mov    rdx, threadFunction0  
    mov    rcx, NULL  
    call   pthread_create
```

This would need to be performed for each thread function

The following code fragment will check if a thread function is complete

```
; pthread_join(pthreadID0, NULL).  
    mov    rdi, qword [pthreadID0]  
    mov    rsi, NULL  
    call   pthread_join
```

If the thread function is not done, the join call will wait until it is complete

The thread function, *threadFunction0()*, itself might contain the following code

```
; --
```

```
global threadFunction0  
threadFunction0:
```

```
; Perform MAX / 2 iterations to update myValue
```

```

mov    rcx, M [ ]
shr    rcx, 1
mov    r10, qword [ ] [ ]
mov    r11, qword [ ] [ ]

incLoop [ ]
; myValue = (myValue / x) + [ ]

mov    rax, qword [myValue] [ ]
cqo
div   [ ] [ ]
add   rax, [ ] [ ]
mov    qword [myValue], [ ] [ ]
loop  incLoop [ ]
[ ]

```

The code for the second thread function would be sim

If both threads are simultaneously executing, they are both trying to update the ***myValue*** variable. For example, assuming that thread function 0 is executing on core 0 and thread function 1 is executing on core 1 (arbitrarily chosen), the following execution trace is possible;

Step	Code: Core 0, Thread 0	Code: Core 1, Thread 1
1	mov rax, qword [myValue]	
2	cqo	mov rax, qword [myValue]
3	div qword [x]	
4	add rax, qword [y]	div qword [ ]
5	mov qword [myValue], rax	add rax, qword [ ]
6		mov qword [myValue], [ ]

As a reminder, each core has its own set of registers. Thus, the core 0 **rax** register is different than the core 1 **rax** register.

If the variable ***myValue*** is currently at 730, two thread executions should increase it to 732. On core 0 code, at step 1 the 730 is copied into core 0, **rax**. On core 1, at step 2, the 730 is copied into core 1, **rax**. As execution progresses, steps 2-4 are performed

core 0 **rax** is incremented from 730 to 731. During this time, on core 1, steps 1-3 are completed and the 730 is also incremented to 731. As the next step, step 5, is executed on core 0, the 731 is written to the **myValue** variable. As the next step, step 6 is executed on core 1, the value 731 is again written to the **myValue** variable. Two execution traces should have incremented the variable twice. However, since the value was obtained in core 1 before core 0 was able to write the final value, one of the increments was lost or duplicated. The end result of this overlap is that the final value of **myValue** will not be correct. And, since the amount of overlapping execution is not predictable, the degree of incorrectness is not easily predicted and may vary between different executions.

For example, if the value of MAX is fairly small, such as 10,000, there would not likely be any overlap. Each thread would start and complete so quickly, that the chances of overlapping execution would be very small. The problem still exists but would likely appear mostly correct on such executions, making it easy to ignore occasional anomalous output. As MAX is increased, the odds of overlapping execution increase.

Such problems can be very challenging to debug and require an understanding of the low-level operations and the technical architecture.

## 19.3 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 19.3.1 Quiz Questions

Below are some quiz questions based on this chapter.

- 1) Explain the difference between concurrency and parallel processing.
- 2) Name the two common approaches to parallel computation.
- 3) In distributed processing, where might the parallel computations take place?
- 4) Provide the names of two examples of large distributed computing projects. Include a one-sentence description of each.
- 5) In multiprocessing, where might the parallel computations take place?
- 6) Provide one advantage and one disadvantage of the distributed computation approach to parallel processing.
- 7) Provide one advantage and one disadvantage of the multiprocessing approach to parallel processing.
- 8) Explain what a *race condition*.

- 9) Will a race condition occur when a shared variable is read by multiple simultaneously executing threads? Explain why or why not.
- 10) Will a race condition occur when a shared variable is written by multiple simultaneously executing threads (without any coordination)? Explain why or why not.

### 19.3.2 Suggested Proj

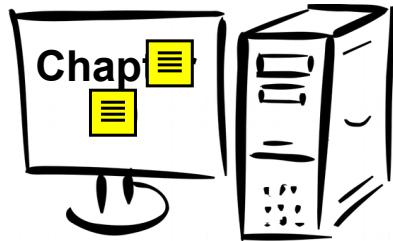
Below are some suggested projects based on this chapter.

- 1) Implement the outlined example program and create two thread functions where each thread function computes the formula MAX/2 times. Set MAX to 1,000,000,000 (one billion).
  1. Initially, structure the main function to call the first thread function and wait until it completes until creating the second thread function and waiting until it completes. This will force the threads to execute sequentially (not in parallel). Include a function to convert the integer into a string and display the result to the console. Use the debugger as necessary to debug the program. When working, execute the program without the debugger and verify that the displayed results are the same as MAX.
  2. Use the Unix time<sup>63</sup> command to establish a base execute time. Record the total elapsed time.
  3. Restructure the program so that both threads are created and then wait for both to complete. This will allow the execution of the threads to occur in parallel. Use the debugger as necessary to debug the program. When working, execute the program without the debugger and note the final value of MAX. Ensure a full understanding of why the displayed value for MAX is incorrect. Additionally, use the Unix time command on the modified program to verify that it uses less elapsed time.
- 2) Update the program from the previous question to resolve the race condition. Use the debugger as necessary to debug the program. One very simple way to accomplish this is to use temporary variables for each thread and combine them after both thread functions have completed and display the final combined result. When working, execute the program without the debugger and verify that the displayed results for results is the same as MAX.

---

63 For more information, refer to: [http://en.wikipedia.org/wiki/Time\\_\(Unix\)](http://en.wikipedia.org/wiki/Time_(Unix))

*If a program is useful, it must be changed.  
If a program is useless, it must be documented.*



## 20.0 Interrupt

In a general sense, an interrupt<sup>64</sup> is a pause or hold in the current flow. For example, if you are talking on the phone and the doorbell rings, the phone conversation is placed on hold, and the door answered. After the salesperson is sent away, the phone conversation is resumed (where the conversation left off).

In computer programming, an interrupt is also a pause, or hold, of the currently executing process. Typically, the current process is interrupted so that some other work can be performed. An interrupt is usually defined as an event that alters the sequence of instructions executed by a processor. Such events correspond to signals generated by software and/or hardware. For example, most Input/Output (I/O) devices generate an interrupt in order to transmit or receive data. Software programs can also generate interrupts to initiate I/O as needed, request OS services, or handle unexpected conditions.

Handling interrupts is a sensitive task. Interrupts can occur at any time; the kernel tries to get the interrupt addressed as soon as possible. Additionally, an interrupt can be interrupted by another interrupt.

### 20.1 Multi-user Operating System

A modern multi-user Operating System (OS) supports multiple programs executing, or appearing to be executing, simultaneously by sharing resources as necessary. The OS is responsible for managing and sharing the resources. These resources include the CPU cores, primary memory (i.e., RAM), secondary storage (i.e., disk or SSD), display screen, keyboard, and mouse. For example, multiple programs must share the available CPU resources (core or cores as applicable).

The interrupt mechanism is the primary means that the OS uses in order to provide the resource sharing. Consequently, understanding how interrupts are processed by

<sup>64</sup> For more information, refer to: <http://en.wikipedia.org/wiki/Interrupt>



computer provides insight into how the operating system is able to provide multi-processing functions. When an interrupt occurs, the current process is interrupted (i.e., placed on hold), the interrupt is handled (which depends on the specific reason for the interrupt), and then eventually the process is resumed. The OS may choose to perform other tasks or processes before the original process is resumed. The interrupt is handled by a special software routine called an Interrupt Service Routine (also called Interrupt Handler, Device Driver, etc.). By using interrupts and quickly switching between various processes, the OS is able to provide the illusion that all processes are executing simultaneously.

Not all code can be interrupted. For example, due to the nature of some operating system kernel functions, there are regions in the kernel which must not be interrupted at all. This includes updating some sensitive operating system data structures.

## 20.1.1 Interrupt Classification



To better understand interrupts and interrupt processing, some background on the timing and categories of interrupts is useful.



## 20.1.2 Interrupt Timing



The timing of interrupts may occur synchronously or asynchronously. These terms are fairly common terms in computer processing and are explained in the following sections.



### 20.1.2.1 Asynchronous Interrupts



In the context of computer interrupts, an asynchronously occurring interrupt means that the interrupt may occur at an arbitrary time with respect to program execution. Asynchronous interrupts are unpredictable relative to any specific location within the executing process. For example, an external hardware device might interrupt the currently executing process at an unpredictable location.



### 20.1.2.2 Synchronous Interrupts



Synchronously occurring interrupts typically occur while under CPU control and are caused by or on behalf of the currently executing process. The synchronous nature is related to where the interrupt occurs and not a specific clock time or CPU cycle time. Synchronous interrupts typically reoccur at the same location (assuming nothing has changed to resolve the original cause).



## 20.1.3 Interrupt Catego

Interrupts are typically categorized as hardware or softw

### 20.1.3.1 Hardware Interr

Hardware interrupts are typically generated by hardware. Hardware interrupts can be issued by

- I/O devices (keyboard, network adapter, )
- Interval timer
- Other CPUs (on multiprocessor syste)

Hardware interrupts are asynchronously occurring. An example of a hardware interrupt is when a key is typed on the keyboard. The OS cannot know ahead of time when, or even if, the key will be pressed. To handle this situation, the keyboard support hardware will generate an interrupt. If the OS is executing an unrelated program, that program is temporarily interrupted while the key is processed. In this example, that specific processing consists of storing the key in a buffer and returning to the interrupted process. Ideally, this brief interruption will have little impact in the interrupted pro

#### 20.1.3.1.1 Except

An exception is a term for an interrupt that is caused by the current process and needs the attention of the kernel. Exceptions are synchronously occurring. In this context synchronous implies that the exception will occur in a predictable or repeatable man

Exceptions are typically divided into categories as follo

- Fa
- Tr
- Ab

An example of a fault is a page fault which is a request for loading part of the program from disk storage into memory. The interrupted process restarts with no loss of continuity.

A trap is typically used for debugging. The process restarts with no loss of continu

An abort is typically an indication that a serious error condition occurred and must be handled. This includes division by zero, attempt to access an invalid memory address, or attempt to execute an invalid/illegal instruction. An illegal instruction might be an instruction that is only allowed to be executed by privileged/authorized processes. An invalid instruction might be caused by attempting to execute a data item (which will not make sense). Based on the severity of the error condition, the process is o

terminated. If the process is not terminated, another routine may be executed to attempt to resolve the problem and re-execute the original routine (but not necessarily from the interrupted location). The C/C++/Java try/catch block is an example of this.

It must be noted that there is not an absolute agreed upon definition for these terms. Some texts use slightly different terminology.

### 20.1.3.2 Software Interr

A software interrupt is produced by the CPU while processing instructions. This is typically a programmed exception explicitly requested by the programmer. Such interrupts are typically synchronously occurring and often used to request system services from the OS. For example, requesting system services such as I/O.

## 20.2 Interrupt Types and Lev

Interrupts have various types and privileges associated with them. The following sections provide an explanation of the types and privileges. Interrupted processes may execute at a lower privilege than the interrupt processing code. In order for interrupts to be effective, the OS must securely handle this privilege escalation and deescalation securely and quickly.

### 20.2.1 Interrupt Ty

The two different types or kinds of interrupts

- Maskable interr
- Non-maskable interr

Maskable interrupts are typically issued by I/O devices. As the name 'maskable' implies, maskable interrupts can be ignored, or masked, for a short time period. This allows the associated interrupt processing to be delayed.

Non-maskable interrupts (NMI's) must be handled immediately. This includes some OS functions and critical malfunctions such as hardware failures. Non-maskable interrupts are always processed by the CPU.

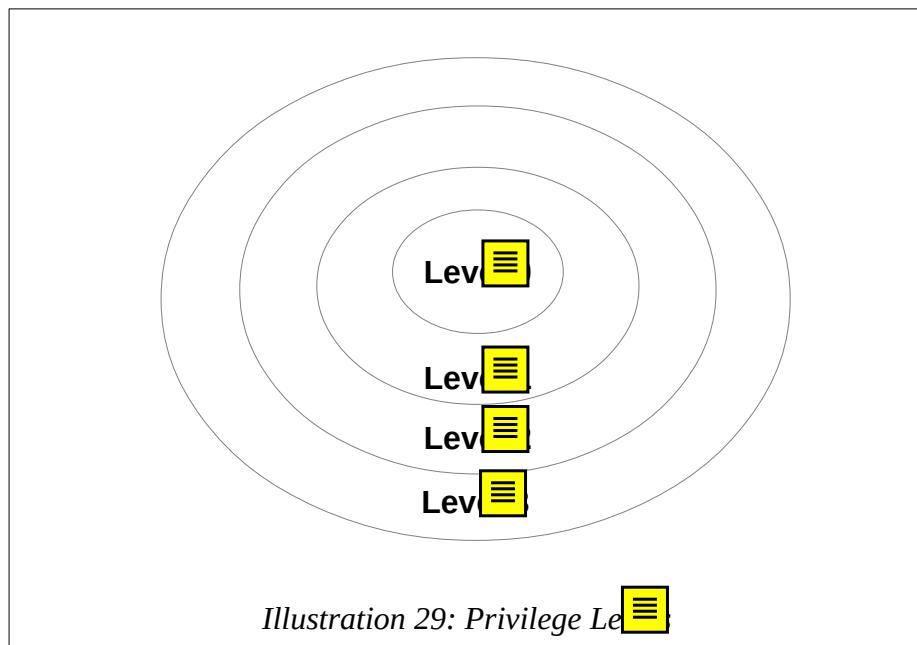
### 20.2.2 Privilege Le

Privilege Levels refer to the privilege level at which the interrupt code executes. This may be a higher privilege level than the interrupted code is executing. The processor executes code in one of four privilege levels as follows:

Level	Description
<b>Level 0</b>	Full access to all hardware resources (no restrictions). Used by only the lowest level OS functions.
<b>Level 1</b>	Somewhat restricted access to hardware resources. Used by library routines and software that interacts with hardware.
<b>Level 2</b>	More restricted access to hardware resources. Used by library routines and software that has limited access to some hardware.
<b>Level 3</b>	No direct access to hardware resources. Application programs run at this level.

Should an application program executing at level 3 be interrupted by a hardware interrupt for the keyboard, the keyboard interrupt handler must execute at level 0.

The following diagram shows the relationship of the levels:



The operating system interrupt processing mechanism will handle this privilege elevation and restoration in a secure manner. That requires that the interrupt source and privileges be verified as part of the interrupt handling mechanism.

## 20.3 Interrupt Process

When an interrupt occurs, it must be handled or processed securely, quickly, and correctly. The general idea is that when the currently executing process is interrupted it must be placed on hold, and the appropriate interrupt handling code found and executed. The specific interrupt processing required depends on the cause or purpose of the interrupt. Once the interrupt is serviced, the original process execution will eventually be resumed.

### 20.3.1 Interrupt Service Routine (ISR)

The code that is executed in response to an interrupt is typically called an Interrupt Service Routine or ISR. The code is sometimes referred to as interrupt handler, handler service routine, or ISR code. For consistency, this document will use the term ISR.

ISR code is challenging to develop due to the issues related to the concurrency and race conditions. Additionally, it is difficult to isolate problems and debug ISR code.

### 20.3.2 Processing Steps

The general steps involved in processing an interrupt are outlined in the following sections.

#### 20.3.2.1 Suspending

Execution of the current program is suspended. As a minimum, this requires saving the **rip** and **rFlags** registers to system stack. The remaining registers are likely to be preserved (as a further step), depending on the specific interrupt. The **rFlags** flag register must be preserved immediately since the interrupt may have been generated asynchronously and those registers will change as successive instructions are executed. This multi-stage process ensures that the program context can be fully restored.

#### 20.3.2.2 Obtaining ISR Address

The ISR addresses are stored in a table referred to as an Interrupt Descriptor Table (IDT). For each ISR, the IDT contains the ISR address and some additional information including task gate (priority and privilege information) for the ISR. Each entry in the IDT is a total of 8 bytes each for a total of 16 bytes per IDT entry. There is a maximum of 256 (0-255) possible entries in the IDT.

---

65 Note, for Windows this data structure is referred to as the Interrupt Vector Table (IVT).

To obtain the starting address of an ISR, the interrupt number is multiplied by 16 (since each entry is 16 bytes) which is used as offset into the IDT where the ISR address is obtained (for that interrupt).

The start of the IDT is pointed to by a dedicated register, IDTR, which is only accessible by the OS and requires level 0 privilege to access.

The addresses of the ISR routines are specific to the OS version and the specific hardware complement of the system. The IDT is created when the system initially boots and reflects the specific system configuration. This is critical in order for the OS to work correctly and consistently on different system hardware configurations.

### 20.3.2.3 Jump to

Once the ISR address is obtained from the IDT some validation is performed. This includes ensuring the interrupt is from a legal/valid source and if a privilege level change is required and allowed. Once the verifications have been completed successfully, the address of the ISR from the IDT is placed in the **rip** register, thus effecting a jump to the ISR routine.

### 20.3.2.4 Suspension Execute

At this point, depending on the specific ISR, a complete process context switch may be performed. A process context switch involves saving the entire set of CPU registers for the interrupted process.

In Linux-based OS's, ISR are typically divided into two parts, referred to as the top-half and bottom-half. Other OS's refer to these as the First-Level Interrupt Handler (FLIH) and the Second-Level Interrupt Handlers (SLIH).

The top-half or FLIH is executed immediately and is where any critical activities are performed. The activities are specific to the ISR, but might include acknowledging the interrupt, resetting hardware (if necessary), and recording any information only available at the time of interrupt. The top-half may perform some blocking of other interrupts (which needs to be minimized).

The bottom-half is where any processing activities (if any) are performed. This helps ensure that the top-half is completed quickly and that any non-critical processing is deferred to a more convenient time. If a bottom-half exists, the top-half will create and schedule the execution of the bottom-half.

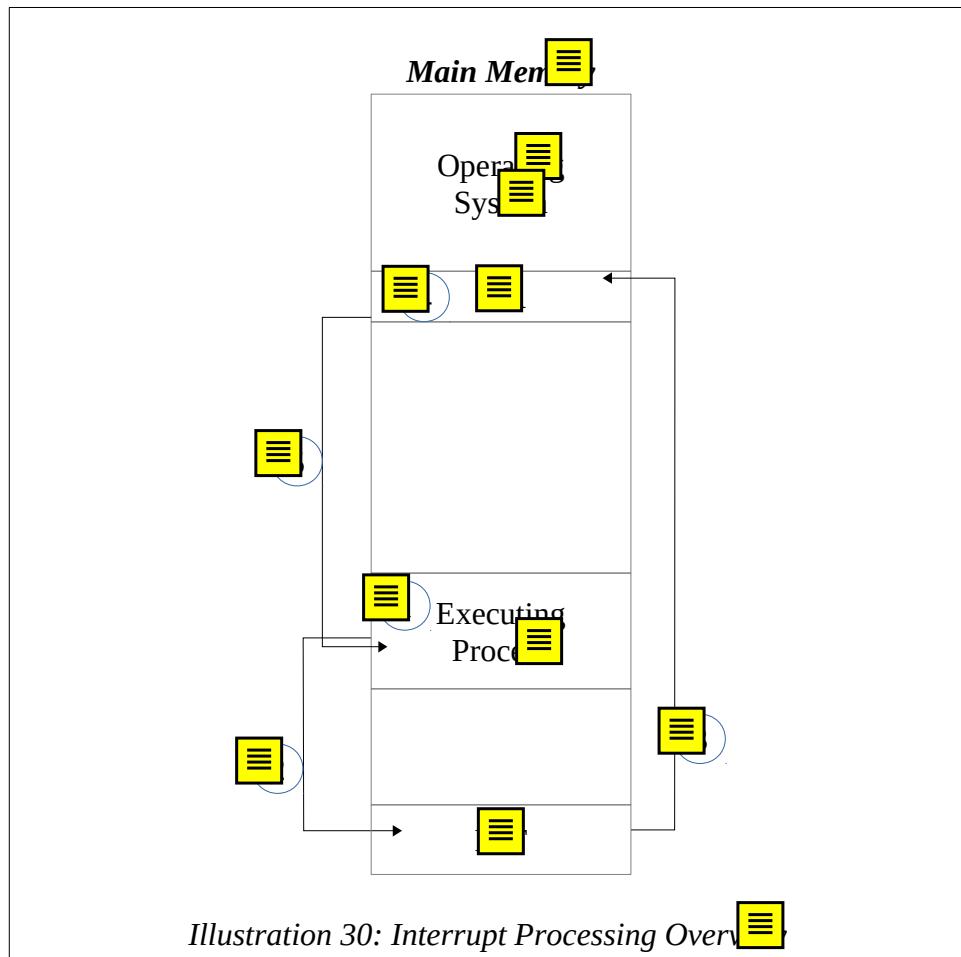
Once the top-half completes, the OS scheduler will select a new pro

### 20.3.2.5 Resump

When the OS is ready to resume the interrupted process, the program context is restored and an **iret** instruction is executed (to pop **rFlags** and **rip** registers, thus completing the restoration).

## 20.4 Suspension Interrupt Processing Sum

The following diagram presents an overview of the general flow used for processing interrupts by the system.



The steps are detailed as follows:

1. Execution of current program is suspended:
  - save **rip** and **rFlags** registers to stack
2. Obtain starting address of Interrupt Service Routine (ISR):
  - interrupt number multiplied by base address
  - used as offset into Interrupt Descriptor Table (IDT)
  - obtains ISR address (for that interrupt) from IDT
3. Jump to Interrupt Service Routine:
  - set **rip** to address from IDT
4. Interrupt Service Routine Execution:
  - save context (i.e., any additional registers altered)
  - process interrupt (specific to interrupt generated)
  - schedule any later data processing activity
5. Interrupted Process Resumption:
  - resume scheduling based on OS scheduler
  - restore context
  - perform **iret** (to pop **rFlags** and **rip** registers)

This interrupt processing mechanism allows a dynamic, run-time lookup for the ISR address.

## 20.5 Exercises

Below are some quiz questions and suggested projects based on this chapter.

### 20.5.1 Quiz Questions

Below are some quiz questions based on this chapter.

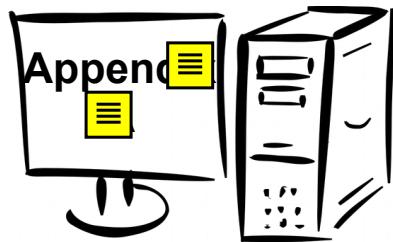
- 1) What is the operating system responsible for? Name some of the resources it manages.
- 2) What is an interrupt?
- 3) What is an exception?
- 4) What is an ISR and what is its function?

- 5) Where (name of the data structure) does the operating system obtain the address when an interrupt occurs and what is contained in it?
- 6) When an interrupt occurs, how is the appropriate offset into the IDT calculated?
- 7) What is the difference between the **iret** and **ret** instructions?
- 8) Why does the OS use the interrupt mechanism instead of just performing standard function calls?
- 9) What is meant by asynchronously occurring interrupts?
- 10) What is meant by synchronously occurring interrupts?
- 11) When an interrupt occurs, the **rip** and **rFlags** registers are pushed on the stack. Much like the call statement, the **rip** register is pushed to save the return address. Explain why is the **rFlag** register pushed on the stack.
- 12) Name two hardware interrupt types.
- 13) List one way for a program to generate an exception.
- 14) What is the difference between a maskable and non-maskable interrupt?

### 20.5.2 Suggested Projects

Below are some suggested projects based on this chapter.

- 1) Write a program to obtain and list the contents of the IDT. This will require writing an integer to ASCII/Hex program in order to display the applicable addresses in hex. Use the debugger as necessary to debug the program. When working on this project, execute the program without the debugger to display results.



## 21.0 Appendix A – ASCII Table

This appendix provides a copy of the ASCII Table for reference.

Char	Dec	Hex
NUL	0	0x00
SOH	1	0x01
STX	2	0x02
ETX	3	0x03
EOT	4	0x04
ENQ	5	0x05
ACK	6	0x06
BEL	7	0x07
BS	8	0x08
TAB	9	0x09
LF	10	0x0A
VT	11	0x0B
FF	12	0x0C
CR	13	0x0D
SO	14	0x0E
SI	15	0x0F
DLE	16	0x10
DC1	17	0x11
DC2	18	0x12
DC3	19	0x13

Char	Dec	Hex
spc	32	0x20
!	33	0x21
"	34	0x22
#	35	0x23
\$	36	0x24
%	37	0x25
&	38	0x26
'	39	0x27
(	40	0x28
)	41	0x29
*	42	0x2A
+	43	0x2B
,	44	0x2C
-	45	0x2D
.	46	0x2E
/	47	0x2F
0	48	0x30
1	49	0x31
2	50	0x32
3	51	0x33

Char	Dec	Hex
@	64	0x40
A	65	0x41
B	66	0x42
C	67	0x43
D	68	0x44
E	69	0x45
F	70	0x46
G	71	0x47
H	72	0x48
I	73	0x49
J	74	0x4A
K	75	0x4B
L	76	0x4C
M	77	0x4D
N	78	0x4E
O	79	0x4F
P	80	0x50
Q	81	0x51
R	82	0x52
S	83	0x53

Char	Dec	Hex
'	96	0x60
a	97	0x61
b	98	0x62
c	99	0x63
d	100	0x64
e	101	0x65
f	102	0x66
g	103	0x67
h	104	0x68
i	105	0x69
j	106	0x6A
k	107	0x6B
l	108	0x6C
m	109	0x6D
n	110	0x6E
o	111	0x6F
p	112	0x70
q	113	0x71
r	114	0x72
s	115	0x73

## Appendix A – ASCII Table

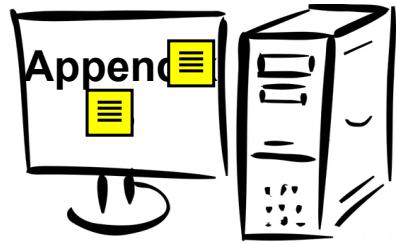
<b>DC4</b>	<b>20</b>	<b>0x14</b>
<b>NAK</b>	<b>21</b>	<b>0x15</b>
<b>SYN</b>	<b>22</b>	<b>0x16</b>
<b>ETB</b>	<b>23</b>	<b>0x17</b>
<b>CAN</b>	<b>24</b>	<b>0x18</b>
<b>EM</b>	<b>25</b>	<b>0x19</b>
<b>SUB</b>	<b>26</b>	<b>0x1A</b>
<b>ESC</b>	<b>27</b>	<b>0x1B</b>
<b>FS</b>	<b>28</b>	<b>0x1C</b>
<b>GS</b>	<b>29</b>	<b>0x1D</b>
<b>RS</b>	<b>30</b>	<b>0x1E</b>
<b>US</b>	<b>31</b>	<b>0x1F</b>

<b>4</b>	<b>52</b>	<b>0x34</b>
<b>5</b>	<b>53</b>	<b>0x35</b>
<b>6</b>	<b>54</b>	<b>0x36</b>
<b>7</b>	<b>55</b>	<b>0x37</b>
<b>8</b>	<b>56</b>	<b>0x38</b>
<b>9</b>	<b>57</b>	<b>0x39</b>
<b>:</b>	<b>58</b>	<b>0x3A</b>
<b>;</b>	<b>59</b>	<b>0x3B</b>
<b>&lt;</b>	<b>60</b>	<b>0x3C</b>
<b>=</b>	<b>61</b>	<b>0x3D</b>
<b>&gt;</b>	<b>62</b>	<b>0x3E</b>
<b>?</b>	<b>63</b>	<b>0x3F</b>

<b>T</b>	<b>84</b>	<b>0x54</b>
<b>U</b>	<b>85</b>	<b>0x55</b>
<b>V</b>	<b>86</b>	<b>0x56</b>
<b>W</b>	<b>87</b>	<b>0x57</b>
<b>X</b>	<b>88</b>	<b>0x58</b>
<b>Y</b>	<b>89</b>	<b>0x59</b>
<b>Z</b>	<b>90</b>	<b>0x5A</b>
<b>[</b>	<b>91</b>	<b>0x5B</b>
<b>\</b>	<b>92</b>	<b>0x5C</b>
<b>]</b>	<b>93</b>	<b>0x5D</b>
<b>^</b>	<b>94</b>	<b>0x5E</b>
<b>_</b>	<b>95</b>	<b>0x5F</b>

<b>t</b>	<b>116</b>	
<b>u</b>	<b>117</b>	
<b>v</b>	<b>118</b>	
<b>w</b>	<b>119</b>	
<b>x</b>	<b>120</b>	
<b>y</b>	<b>121</b>	
<b>z</b>	<b>122</b>	
<b>{</b>	<b>123</b>	
<b> </b>	<b>124</b>	
<b>}</b>	<b>125</b>	
<b>~</b>	<b>126</b>	
<b>DEL</b>	<b>127</b>	

For additional information and a more complete listing of the ASCII codes (including the extended ASCII characters), refer to <http://www.asciitable.com/>



## 22.0 Appendix B – Instruction Set Summary

This appendix provides a listing and brief description of the instructions covered in this text. This set of instructions is a subset of the complete instruction set. For a complete listing of the instructions, refer to the references noted in chapter 1.

### 22.1 Notation

The following table summarizes the notational conventions used in this appendix.

Operand Notation	Description
<code>&lt;reg&gt;</code>	Register operand. The operand must be a register (e.g., <code>al</code> , <code>bl</code> , etc.).
<code>&lt;reg8&gt;, &lt;reg16&gt;, &lt;reg32&gt;</code>	Register operand with specific size requirement. For example, <code>reg8</code> means a byte sized register (e.g., <code>al</code> , <code>bl</code> , etc.) only and <code>reg32</code> means a double-word sized register (e.g., <code>eax</code> , <code>ebx</code> , etc.) only.
<code>&lt;dest&gt;</code>	Destination operand. The operand may be a register or memory. Since it is a destination operand, the contents will be overwritten with the new result (based on the specific instruction).
<code>&lt;RXdest&gt;</code>	Floating-point destination register operand. The operand must be a floating-point register. Since it is a destination operand, the contents will be overwritten with the new result (based on the specific instruction).
<code>&lt;src&gt;</code>	Source operand. Operand value is unchanged after the instruction.
<code>&lt;imm&gt;</code>	Immediate value. May be specified in decimal, hex, octal or binary.

Operand Notation	Description
<mem>	Memory location. May be a variable name or an indirect reference (i.e., a memory address).
<op> or <operand>	Operand, register or memory.
<op8>, <op16>, <op32>, <op64>	Operand, register or memory, with specific size requirement. For example, <b>op8</b> means a byte sized operand only and <b>reg32</b> means a double-word sized operand only.
<label>	Program label.

## 22.2 Data Movement Instructions

Below is a summary of the basic data movement and addressing instructions.

Instruction	Description
<b>mov</b> <dest>, <src>	Copy source operand to the destination operand. <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operands cannot be an immediate.
<b>lea</b> <reg64>, <mem>	Place address of <mem> into register.

## 22.3 Data Conversion Instructions

Below is a summary of the basic data conversion instructions.

Instruction	Description
<b>movzx</b> <dest>, <src>	Unsigned widening conversion. <i>Note 1</i> , both operands cannot be memory.
<b>movzx</b> <reg16>, <op8>	<i>Note 2</i> , destination operands cannot be an immediate.
<b>movzx</b> <reg32>, <op8>	<i>Note 3</i> , immediate values not allowed.
<b>movzx</b> <reg32>, <op16>	
<b>movzx</b> <reg64>, <op16>	
<b>movzx</b> <reg64>, <op32>	
<b>cbw</b>	Convert byte in <b>al</b> into word in <b>ax</b> . <i>Note</i> , only works for <b>al</b> to <b>ax</b> register.



Instruction	Explanation
<b>cwd</b>	Convert word in <b>ax</b> into double-word in <b>dx:ax</b> . Note, only works for <b>ax</b> to <b>dx:ax</b> registers.
<b>cwde</b>	Convert word in <b>ax</b> into double-word in <b>eax</b> . Note, only works for <b>ax</b> to <b>eax</b> register.
<b>cdq</b>	Convert double-word in <b>eax</b> into quadword in <b>edx:eax</b> . Note, only works for <b>eax</b> to <b>edx:eax</b> register.
<b>cdqe</b>	Convert double-word in <b>eax</b> into quadword in <b>rax</b> . Note, only works for <b>rax</b> register.
<b>cqo</b>	Convert quadword in <b>rax</b> into word in double-quadword in <b>rdx:rax</b> . Note, only works for <b>rax</b> to <b>rdx:rax</b> register.
<b>movsx</b> <dest>, <src>  <b>movsx</b> <reg16>, <op8> <b>movsx</b> <reg32>, <op8> <b>movsx</b> <reg32>, <op16> <b>movsx</b> <reg64>, <op8> <b>movsx</b> <reg64>, <op16> <b>movsxd</b> <reg64>, <op16>	Signed widening conversion (via sign extension).  Note 1, both operands cannot be memory. Note 2, destination operands cannot be an immediate.  Note 3, immediate values not allowed

## 22.4 Integer Arithmetic Instructions



Below is a summary of the basic integer arithmetic instructions.



Instruction	Explanation
<b>add</b> <dest>, <src>	Add two operands, (<dest> + <src>) and place the result in <dest> (over-writing previous value).  Note 1, both operands cannot be memory. Note 2, destination operand cannot be an immediate.
<b>inc</b> <operand>	Increment <operand> by 1. Note, <operand> cannot be an immediate.
<b>adc</b> <dest>, <src>	Add two operands, (<dest> + <src>) and



## Appendix B – Instruction Set Summary

Instruction	Explanation
	<p>previous carry (stored in the carry bit in the <b>rFlag</b> register) and place the result in &lt;<b>dest</b>&gt; (over-writing previous value).</p> <p><i>Note 1</i>, both operands cannot be memory.</p> <p><i>Note 2</i>, destination operand cannot be an immediate.</p>
Examples:	<pre>adc    rcx, qword [dVva1] adc    rax, 42</pre>
<b>sub</b> < <b>dest</b> >, < <b>src</b> >	<p>Subtract two operands, (&lt;<b>dest</b>&gt; - &lt;<b>src</b>&gt;) and place the result in &lt;<b>dest</b>&gt; (over-writing previous value).</p> <p><i>Note 1</i>, both operands cannot be memory.</p> <p><i>Note 2</i>, destination operand cannot be an immediate.</p>
<b>dec</b> < <b>operand</b> >	<p>Decrement &lt;<b>operand</b>&gt; by 1.</p> <p><i>Note</i>, &lt;<b>operand</b>&gt; cannot be an immediate.</p>
<b>mul</b> < <b>src</b> > <b>mul</b> < <b>op8</b> > <b>mul</b> < <b>op16</b> > <b>mul</b> < <b>op32</b> > <b>mul</b> < <b>op64</b> >	<p>Multiply A register (<b>al</b>, <b>ax</b>, <b>eax</b>, or <b>rax</b>) times the &lt;<b>src</b>&gt; operand.</p> <p>Byte: <b>ax</b> = <b>al</b> * &lt;<b>src</b>&gt;</p> <p>Word: <b>dx:ax</b> = <b>ax</b> * &lt;<b>src</b>&gt;</p> <p>Double: <b>edx:eax</b> = <b>eax</b> * &lt;<b>src</b>&gt;</p> <p>Quad: <b>rdx:rax</b> = <b>rax</b> * &lt;<b>src</b>&gt;</p> <p><i>Note</i>, &lt;<b>src</b>&gt; operand cannot be an immediate.</p>
<b>imul</b> < <b>src</b> > <b>imul</b> < <b>dest</b> >, < <b>src/imm32</b> > <b>imul</b> < <b>dest</b> >, < <b>src</b> >, < <b>imm</b> >  <b>imul</b> < <b>op8</b> > <b>imul</b> < <b>op16</b> > <b>imul</b> < <b>op32</b> > <b>imul</b> < <b>op64</b> > <b>imul</b> < <b>reg16</b> >, < <b>op16/imm</b> > <b>imul</b> < <b>reg32</b> >, < <b>op32/imm</b> > <b>imul</b> < <b>reg64</b> >, < <b>op64/imm</b> > <b>imul</b> < <b>reg16</b> >, < <b>op16</b> >, < <b>imm</b> > <b>imul</b> < <b>reg32</b> >, < <b>op32</b> >, < <b>imm</b> > <b>imul</b> < <b>reg64</b> >, < <b>op64</b> >, < <b>imm</b> >	<p>Signed multiply instruction</p> <p>For single operand:</p> <p>Byte: <b>ax</b> = <b>al</b> * &lt;<b>src</b>&gt;</p> <p>Word: <b>dx:ax</b> = <b>ax</b> * &lt;<b>src</b>&gt;</p> <p>Double: <b>edx:eax</b> = <b>eax</b> * &lt;<b>src</b>&gt;</p> <p>Quad: <b>rdx:rax</b> = <b>rax</b> * &lt;<b>src</b>&gt;</p> <p><i>Note</i>, &lt;<b>src</b>&gt; operand cannot be an immediate.</p> <p>For two operands:</p> <p>&lt;<b>reg16</b>&gt; = &lt;<b>reg16</b>&gt; * &lt;<b>op16/imm</b>&gt;</p> <p>&lt;<b>reg32</b>&gt; = &lt;<b>reg32</b>&gt; * &lt;<b>op32/imm</b>&gt;</p> <p>&lt;<b>reg64</b>&gt; = &lt;<b>reg64</b>&gt; * &lt;<b>op64/imm</b>&gt;</p>



Instruction	Explanation
	<p>For three operations:</p> $<\text{reg16}> = <\text{op16}> * <\text{imm}>$ $<\text{reg32}> = <\text{op32}> * <\text{imm}>$ $<\text{reg64}> = <\text{op64}> * <\text{imm}>$
<b>div</b> <>  <b>div</b> <> <b>div</b> <> <b>div</b> <>	<p>Unsigned divide A/D register (<b>ax</b>, <b>dx:ax</b>, <b>cx:dx</b>, <b>bx:cx</b>, or <b>rdx:rax</b>) by the &lt;<b>src</b>&gt; operand.</p> <p>Byte: <b>al</b> = <b>ax</b> / &lt;<b>src</b>&gt;, rem in <b>ah</b></p> <p>Word: <b>ax</b> = <b>dx:ax</b> / &lt;<b>src</b>&gt;, rem in <b>dx</b></p> <p>Double: <b>eax</b> = <b>eax</b> / &lt;<b>src</b>&gt;, rem in <b>edx</b></p> <p>Quad: <b>rax</b> = <b>rax</b> / &lt;<b>src</b>&gt;, rem in <b>rdx</b></p> <p>Note, &lt;<b>src</b>&gt; operand cannot be an immediate.</p>
<b>idiv</b> <>  <b>idiv</b> <> <b>idiv</b> <> <b>idiv</b> <>	<p>Signed divide A/D register (<b>ax</b>, <b>dx:ax</b>, <b>cx:dx</b>, <b>bx:cx</b>, or <b>rdx:rax</b>) by the &lt;<b>src</b>&gt; operand.</p> <p>Byte: <b>al</b> = <b>ax</b> / &lt;<b>src</b>&gt;, rem in <b>ah</b></p> <p>Word: <b>ax</b> = <b>dx:ax</b> / &lt;<b>src</b>&gt;, rem in <b>dx</b></p> <p>Double: <b>eax</b> = <b>eax</b> / &lt;<b>src</b>&gt;, rem in <b>edx</b></p> <p>Quad: <b>rax</b> = <b>rax</b> / &lt;<b>src</b>&gt;, rem in <b>rdx</b></p> <p>Note, &lt;<b>src</b>&gt; operand cannot be an immediate.</p>

## 22.5 Logical, Shift, and Rotate Instructions



Below is a summary of the basic logical, shift, arithmetic shift, and rotate instructions.



Instruction	Explanation
<b>and</b> < <b>dest</b> >, < <b>src</b> >	<p>Perform logical AND operation on two operands, (&lt;<b>dest</b>&gt; and &lt;<b>src</b>&gt;) and place the result in &lt;<b>dest</b>&gt; (over-writing previous value).</p> <p>Note 1, both operands cannot be memory.</p> <p>Note 2, destination operand cannot be an immediate.</p>
<b>or</b> < <b>dest</b> >, < <b>src</b> >	<p>Perform logical OR operation on two operands, (&lt;<b>dest</b>&gt;    &lt;<b>src</b>&gt;) and place the result in &lt;<b>dest</b>&gt; (over-writing previous value).</p> <p>Note 1, both operands cannot be memory.</p> <p>Note 2, destination operand cannot be an immediate.</p>



## Appendix B – Instruction Set Summary

Instruction	Explanation
<b>xor</b> <dest>, <src>	Perform logical XOR operation on two operands, ( <b>&lt;dest&gt;</b> ^ <b>&lt;src&gt;</b> ) and place the result in <b>&lt;dest&gt;</b> (over-writing previous value). <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , destination operand can not be an immediate.
<b>not</b> <op>	Perform a logical not operation (one's complement on the operand 1's→0's and 0's→1's). <i>Note</i> , operand cannot be an immediate
<b>shl</b> <dest>, <imm> <b>shl</b> <dest>, cl	Perform logical shift left operation on destination operand. Zero fills from right (as needed). The <imm> or the value in <b>cl</b> register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.
<b>shr</b> <dest>, <imm> <b>shr</b> <dest>, cl	Perform logical shift right operation on destination operand. Zero fills from left (as needed). The <imm> or the value in <b>cl</b> register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.
<b>sal</b> <dest>, <imm> <b>sal</b> <dest>, cl	Perform arithmetic shift left operation on destination operand. Zero fills from right (as needed). The <imm> or the value in <b>cl</b> register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.
<b>sar</b> <dest>, <imm> <b>sar</b> <dest>, cl	Perform arithmetic shift right operation on destination operand. Sign fills from left (as needed). The <imm> or the value in <b>cl</b> register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.



Instruction	Explanation
<b>rol</b> <dest>, <imm> <b>rol</b> <dest>, cl	Perform rotate left operation on destination operand. The <imm> or the value in cl register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.
<b>ror</b> <dest>, <imm> <b>ror</b> <dest>, cl	Perform rotate right operation on destination operand. The <imm> or the value in cl register must be between 1 and 64. <i>Note</i> , destination operand cannot be an immediate.

## 22.6 Control Instructions

Below is a summary of the basic control instructions.



Instruction	Explanation
<b>cmp</b> <op1>, <op2>	Compare <op1> with <op2>. Results are stored in the rFlag register. <i>Note 1</i> , operands are not changed. <i>Note 2</i> , both operands cannot be memory. <i>Note 3</i> , <op1> operand cannot be an immediate.
<b>je</b> <label>	Based on preceding comparison instruction, jump to <label> if <op1> == <op2>. Label must be defined exactly once.
<b>jne</b> <label>	Based on preceding comparison instruction, jump to <label> if <op1> != <op2>. Label must be defined exactly once.
<b>jl</b> <label>	For signed data, based on preceding comparison instruction, jump to <label> if <op1> < <op2>. Label must be defined exactly once.



## Appendix B – Instruction Set Summary

Instruction	Explanation
jle <label>	For signed data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> \leq <\text{op2}>$ . Label must be defined exactly once.
jg <label>	For signed data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> > <\text{op2}>$ . Label must be defined exactly once.
jge <label>	For signed data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> \geq <\text{op2}>$ . Label must be defined exactly once.
jb <label>	For unsigned data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> < <\text{op2}>$ . Label must be defined exactly once.
jbe <label>	For unsigned data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> \leq <\text{op2}>$ . Label must be defined exactly once.
ja <label>	For unsigned data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> > <\text{op2}>$ . Label must be defined exactly once.
jae <label>	For unsigned data, based on preceding comparison instruction, jump to <label> if $<\text{op1}> \geq <\text{op2}>$ . Label must be defined exactly once.
loop <label>	Decrement rcx register and jump to <label> if rcx is $\neq 0$ . <i>Note</i> , label must be defined exactly once.



## 22.7 Stack Instructions

Below is a summary of the basic stack instructions.

Instruction	Explanation
<code>push &lt;op64&gt;</code>	Push the 64-bit operand on the stack. Adjusts <code>rsp</code> accordingly. Operand is unaltered.
<code>pop &lt;op64&gt;</code>	Pop the 64-bit operand from the stack. Adjusts <code>rsp</code> accordingly. The operand may not be an immediate value. Operand is overwritten.

## 22.8 Function Instructions

Below is a summary of the basic instructions for implementing function calls.

Instruction	Explanation
<code>call &lt;funcName&gt;</code>	Calls a function. Push the 64-bit <code>rip</code> register and jump to the <code>&lt;funcName&gt;</code> .
<code>ret</code>	Return from a function. Pop the stack into the <code>rip</code> register, effecting a jump to the line after the call.

## 22.9 Floating-Point Data Movement Instructions

Below is a summary of the basic instructions for floating-point data movement instructions.

Instruction	Explanation
<code>movss &lt;dest&gt;, &lt;src&gt;</code>	Copy 32-bit source operand to the 32-bit destination operand. <i>Note 1</i> , both operands cannot be memory. <i>Note 2</i> , operands cannot be an immediate.
Examples:	<code>movss xmm0, dword [x]</code> <code>movss dword [fltVar], xmm1</code> <code>movss xmm3, xmm2</code>



Instruction	Explanation
<code>movsd &lt;dest&gt;, &lt;src&gt;</code>	<p>Copy 64-bit source operand to the 64-bit destination operand.</p> <p><i>Note 1</i>, both operands cannot be memory locations.  <i>Note 2</i>, operands cannot be an immediate value.</p>
Examples:	<pre>movsd xmm0, qword [y] movsd qword [fldtVar], xmm1 movsd xmm3, xmm2</pre>

## 22.10 Floating-Point Data Conversion Instructions

Below is a summary of the basic instructions for floating-point data conversion instructions.

Instruction	Explanation
<code>cvtss2sd &lt;RXdest&gt;, &lt;src&gt;</code>	<p>Convert 32-bit floating-point source operand to the 64-bit floating-point destination operand.</p> <p><i>Note 1</i>, destination operand must be floating-point register.  <i>Note 2</i>, source operand cannot be an immediate value.</p>
Examples:	<pre>cvtss2sd xmm0, dword [fltsVar] cvtss2sd xmm3, eax cvtss2sd xmm3, xmm2</pre>
<code>cvtsd2ss &lt;RXdest&gt;, &lt;src&gt;</code>	<p>Convert 64-bit floating-point source operand to the 32-bit floating-point destination operand.</p> <p><i>Note 1</i>, destination operand must be floating-point register.  <i>Note 2</i>, source operand cannot be an immediate value.</p>
Examples:	<pre>cvtsd2ss xmm0, qword [fltdVar] cvtsd2ss xmm1, rax cvtsd2ss xmm3, xmm2</pre>
<code>cvtss2si &lt;reg&gt;, &lt;src&gt;</code>	<p>Convert 32-bit floating-point source operand to the 32-bit integer destination operand.</p> <p><i>Note 1</i>, destination operand must be register.  <i>Note 2</i>, source operand cannot be an immediate value.</p>



Instruction	Explanation
Examples:	<pre>cvtss2si    xmm1, xmm0 cvtss2si    eax, xmm0 cvtss2si    eax, dword [fltsV]</pre>
<b>cvtsd2si &lt;reg&gt;, &lt;src&gt;</b>	<p>Convert 64-bit floating-point source operand to the 32-bit integer destination operand.</p> <p><i>Note 1</i>, destination operand must be register.</p> <p><i>Note 2</i>, source operand cannot be an immediate.</p>
Examples:	<pre>cvtss2si    xmm1, xmm0 cvtss2si    eax, xmm0 cvtss2si    eax, qword [fltDV]</pre>
<b>cvtsi2ss &lt;RXdest&gt;, &lt;src&gt;</b>	<p>Convert 32-bit integer source operand to the 32-bit floating-point destination operand.</p> <p><i>Note 1</i>, destination operand must be floating-point register.</p> <p><i>Note 2</i>, source operand cannot be an immediate.</p>
Examples:	<pre>cvtsi2ss    xmm0, eax cvtsi2ss    xmm0, dword [fltDV]</pre>
<b>cvtsi2sd &lt;RXdest&gt;, &lt;src&gt;</b>	<p>Convert 32-bit integer source operand to the 64-bit floating-point destination operand.</p> <p><i>Note 1</i>, destination operand must be floating-point register.</p> <p><i>Note 2</i>, source operand cannot be an immediate.</p>
Examples:	<pre>cvtsi2sd    xmm0, eax cvtsi2sd    xmm0, dword [fltDV]</pre>

## 22.11 Floating-Point Arithmetic Instructions



Below is a summary of the basic instructions for floating-point arithmetic instructions.



## Appendix B – Instruction Set Summary

Instruction	Explanation
<b>addss</b> <RXdest>, <src>	Add two 32-bit floating-point operands, ( $<\text{RXdest}> + <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>addss</b> xmm0, xmm3 <b>addss</b> xmm5, dword [fSV]
<b>addsd</b> <RXdest>, <src>	Add two 64-bit floating-point operands, ( $<\text{RXdest}> + <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>addsd</b> xmm0, xmm3 <b>addsd</b> xmm5, qword [fDV]
<b>subss</b> <RXdest>, <src>	Subtract two 32-bit floating-point operands, ( $<\text{RXdest}> - <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>subss</b> xmm0, xmm3 <b>subss</b> xmm5, dword [fSV]
<b>subsd</b> <RXdest>, <src>	Subtract two 64-bit floating-point operands, ( $<\text{RXdest}> - <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.

Instruction	Explanation
Examples:	<code>subsd xmm0, xmm3</code> <code>subsd xmm5, qword [fDV]</code>
<b>mulss &lt;RXdest&gt;, &lt;src&gt;</b>	Multiply two 32-bit floating-point operands, ( $<\text{RXdest}> * <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<code>mulss xmm0, xmm3</code> <code>mulss xmm5, dword [fSV]</code>
<b>mulsd &lt;RXdest&gt;, &lt;src&gt;</b>	Multiply two 64-bit floating-point operands, ( $<\text{RXdest}> * <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<code>mulsd xmm0, xmm3</code> <code>mulsd xmm5, qword [fDV]</code>
<b>divss &lt;RXdest&gt;, &lt;src&gt;</b>	Divide two 32-bit floating-point operands, ( $<\text{RXdest}> / <\text{src}>$ ) and place the result in $<\text{RXdest}>$ (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<code>divss xmm0, xmm3</code> <code>divss xmm5, dword [fSV]</code>

## Appendix B – Instruction Set Summary

Instruction	Explanation
<b>divsd</b> <RXdest>, <src>	Divide two 64-bit floating-point operands, (<RXdest> / <src>) and place the result in <RXdest> (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>divsd</b> xmm0, xmm3 <b>divsd</b> xmm5, qword [fDV]
<b>sqrtss</b> <RXdest>, <src>	Take the square root of the 32-bit floating-point source operand and place the result in destination operand (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>sqrtss</b> xmm0, xmm3 <b>sqrtss</b> xmm7, dword [fSV]
<b>sqrtsd</b> <RXdest>, <src>	Take the square root of the 64-bit floating-point source operand and place the result in destination operand (over-writing previous value). <i>Note 1</i> , destination operands must be a floating-point register. <i>Note 2</i> , source operand cannot be an immediate.
Examples:	<b>sqrtsd</b> xmm0, xmm3 <b>sqrtsd</b> xmm7, qword [fDV]

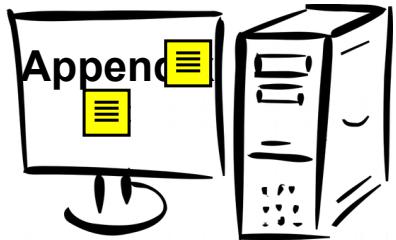
## 22.12 Floating-Point Control Instructions

Below is a summary of the basic instructions for floating-point control instructions.

Instruction	Explanation
<b>ucomiss</b> <RXsrc>, <src>	Compare two 32-bit floating-point operands, (<RXsrc> + <src>). Results are placed in the <b>rFlag</b> register. Neither operand is changed. <i>Note 1</i> , <RXsrc> operands must be a floating-point register. <i>Note 2</i> , source operand may be a floating-point register or memory, but not be an immediate.
Examples:	<b>ucomiss</b> xmm0, xmm3 <b>ucomiss</b> xmm5, dword [fSV]
<b>ucomisd</b> <RXsrc>, <src>	Compare two 64-bit floating-point operands, (<RXsrc> + <src>). Results are placed in the <b>rFlag</b> register. Neither operand is changed. <i>Note 1</i> , <RXsrc> operands must be a floating-point register. <i>Note 2</i> , source operand may be a floating-point register or memory, but not be an immediate.
Examples:	<b>ucomisd</b> xmm0, xmm3 <b>ucomisd</b> xmm5, dword [fSV]

## Appendix B – Instruction Set Summary





## 23.0 Appendix C – System Services

This appendix provides a listing and brief description of a subset of the system service calls. This list is for 64-bit Ubuntu systems. A more complete list can be obtained from multiple web sources.

### 23.1 Return Codes

The system call will return a code in the **rax** register. If the value returned is less than 0, that is an indication that an error has occurred. If the operation is successful, the value returned will depend on the specific system service. Refer to the Error Codes section for additional information regarding the values of the error codes.

### 23.2 Basic System Services

The following table summarizes the more common system services.

Call ID ( <code>int</code> )	System Service	Description
0	<code>SYS_read</code>	<b>Read characters</b>  <code>rdi</code> = file descriptor (of where to read from) <code>rsi</code> = address of where to store characters <code>rdx</code> = count of characters to read
		If unsuccessful, returns negative value. If successful, returns count of characters actually read.
1	<code>SYS_write</code>	<b>Write characters</b>  <code>rdi</code> = file descriptor (of where to write to)

## Appendix C – System Services

Call #	System Service	Description
		<b>rsi</b> = address of characters to write <b>rdx</b> = count of characters to write
		If unsuccessful, returns negative value. If successful, returns count of characters actually written.
2	<b>SYS_open</b>	<b>Open a file</b> <b>rdi</b> = address of NULL terminated file name <b>rsi</b> = file status flags (typically O_RDONLY)
		If unsuccessful, returns negative value. If successful, returns file descriptor.
3	<b>SYS_close</b>	<b>Close an open file</b> <b>rdi</b> = file descriptor of open file to close
		If unsuccessful, returns negative value
8	<b>SYS_lseek</b>	<b>Reposition the file read/write file offset</b> <b>rdi</b> = file descriptor (of where to write) <b>rsi</b> = offset <b>rdx</b> = offset type
		If unsuccessful, returns negative value
57	<b>SYS_fork</b>	<b>Fork current process</b>
59	<b>SYS_execve</b>	<b>Execute a program</b> <b>rdi</b> = Address of NULL terminated string for name of program to execute.
60	<b>SYS_exit</b>	<b>Terminate executing process</b> <b>rdi</b> = exit status (typically 0)

Call #	System Service	Description
85	SYS_creat	Open/Create a file. rdi = address of NULL terminated file name rsi = file mode
		If unsuccessful, returns negative value. If successful, returns file descriptor.
96	SYS_gettimeofday	Get date and time of day. rdi = address of time value structure rsi = address of time zone structure
		If unsuccessful, returns negative value. If successful, returns information in the passed structures.

### 23.3 File Modes

When performing file operations, the file mode provides information to the operating system regarding the file access permissions that will be allowed.

When opening an existing file, one of the following file modes must be specified:

Mode	Value	Description
O_RDONLY	0	Read only. Allow reading from the file, but do not allow writing to the file. Most common operation.
O_WRONLY	1	Write only. Typically used when information is to be appended to a file.
O_RDWR	2	Allow simultaneous reading and writing.

When creating a new file, the file permissions must be specified. Below are the complete set of file permissions. As is standard for Linux file systems, the permission values are specified in Octal.

## Appendix C – System Service Permissions

Mode	Value	Description
S_IRWXU	00700q	User (file owner) has read, write, and execute permission.
S_IRUSR	00400q	User (file owner) has read permission.
S_IWUSR	00200q	User (file owner) has write permission.
S_IXUSR	00100q	User (file owner) has execute permission.
S_IRWXG	00070q	Group has read, write, and execute permission.
S_IRGRP	00040q	Group has read permission.
S_IWGRP	00020q	Group has write permission.
S_IXGRP	00010q	Group has execute permission.
S_IRWXO	00007q	Others have read, write, and execute permission.
S_IROTH	00004q	Others have read permission.
S_IWOTH	00002q	Others have write permission.
S_IXOTH	00001q	Others have execute permission.

The text examples only address permissions for the user or owner of the file.

### 23.4 Error Codes

If a system service returns an error, the value of the return code will be negative. The following is a list of the error code. The code value is provided along with the Linux symbolic name. High-level languages typically use the name which is not used at the assembly level and is only provided for reference.

Error Code	Symbolic Name	Description
-1	EPERM	Operation not permitted.
-2	ENOENT	No such file or directory.
-3	ESRCH	No such process.
-4	EINTR	Interrupted system call.



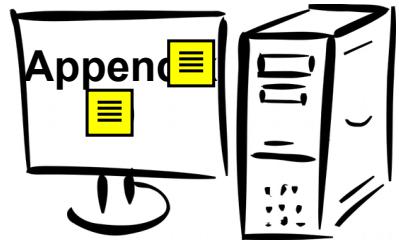
-5	EIO	I/O E
-6	ENXIO	No such device or add
-7	E2BIG	Argument list too l
-8	ENOEXEC	Exec format e
-9	EBADF	Bad file num
-10	ECHILD	No child pro
-11	EAGAIN	Try ag
-12	ENOMEM	Out of mem
-13	EACCES	Permission de
-14	EFAULT	Bad add
-15	ENOTBLK	Block device requ
-16	EBUSY	Device or resource b
-17	EEXIST	File ex
-18	EXDEV	Cross-device
-19	ENODEV	No such dev
-20	ENOTDIR	Not a directo
-21	EISDIR	Is a directo
-22	EINVAL	Invalid argum
-23	ENFILE	File table overf
-24	EMFILE	Too many open f
-25	ENOTTY	Not a typewri
-26	ETXTBSY	Text file b
-27	EFBIG	File too larg
-28	ENOSPC	No space left on dev
-29	ESPIPE	Illegal s
-30	EROFS	Read-only file sys
-31	EMLINK	Too many li
-32	EPIPE	Broken p



## Appendix C – System Services

-33	EDOM	Math argument out of domain of function.
-34	ERANGE	Math result not representable.

Only the most common error codes are shown. A complete list can be found via the Internet or by looking on the current system includes files. For Ubuntu, this is typically located in **/usr/include/asm-generic/errno-base.h**.



## 24.0 Appendix D – Quiz Question Answers

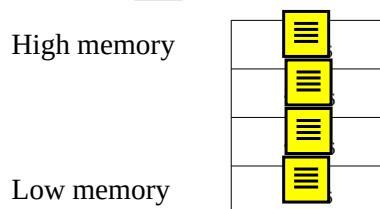
This appendix provides answers for the quiz questions in each chapter.

### 24.1 Quiz Question Answers, Chapter 24

There are no quiz questions for Chapter 24.

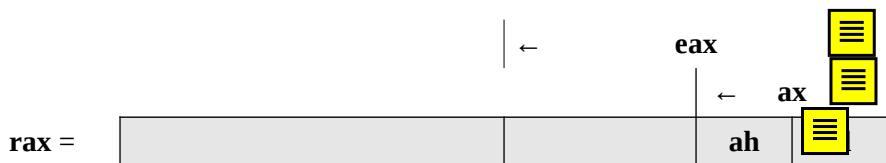
### 24.2 Quiz Question Answers, Chapter 24

- 1) See Section 2.1, Illustration 1, Computer Architectures.
- 2) Bus or Interconnect.
- 3) Secondary storage.
- 4) Primary storage or main memory (RAM).
- 5) Keeps a copy of the data closer to the CPU, eliminating the extra time required to access the RAM via the bus.
- 6) 4 bytes.
- 7) 1 byte.
- 8) The LSB is  $40_{16}$  and the MSB is  $00_{16}$ .
- 9) The answer is as follows:



## Appendix D – Quiz Question Answers

10) The layout of the registers is:



11) The answers are as follows:

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.

12) The **rip** register:



13) The **rsp** register:



14) The **rax** register is: 0000000000000000



15) The answers are as follows:

1. E
2. CDE
3. 89ABCDE
4. 0123456789ABCDE

### 24.3 Quiz Question Answers, Chapter 24

1) The answers are as follows:

1. -128 to +
2. 0 to



3. -32,768 to +32,
4. 0 to 65,
5. -2,147,483,648 to +2,147,483,
6. 0 to 4,294,967,
- 2) The answers are as follo
- 1.
- 2.
- 3.
- 4.
- 3) The answers are as follo
1. 0,
2. 0,
3. 0,
4. 0,
- 4) The answers are as follo
1. 0xF,
2. 0x0,
3. 0xF,
4. 0xF,
- 5) The answers are as follo
1. 0xFFFFF,
2. 0xFFFFF,
3. 0x00000,
4. 0xFFFFFI,
- 6) The answers are as follo
- 1.
- 2.



## Appendix D – Quiz Question Answers

3.   

  4. 
- 7)  $0.5_{10}$  is represented as 
- 8) The answers are as follows 
1. -1 
  2. +1 
  3. 
  4. 
- 9) The answers are as follows 
1. 0x41340 
  2. 0xC1890 
  3. 0x41AF0 
  4. 0xBF400 
- 10) The answers are as follows 
1. 0 
  2. 0 
  3. 0 
  4. 0 
  5. 0 
- 11) The answers are as follows 
1. “World” = 0x57 0x6F 0x72 0x6C 0 
  2. “123” = 0x31 0x32 0 
  3. “Yes!?” = 0x59 0x65 0x73 0x21 0 

## 24.4 Quiz Question Answers, Chapter 24

- 1) y 
- 2) With the ; (semicolon) 



- 3) Section d
- 4) Section
- 5) Section t
- 6) The answers are as folld
  1. bNum db
  2. wNum dw 10
  3. dwNum dd 2126
  4. qwNum dq 10000000
- 7) The answers are as folld
  1. bArr resb
  2. wArr resw 3
  3. dwArr resd
  4. qArr resq 5
- 8) The declarations 

```
global _$  
_s
```

## 24.5 Quiz Question Answers, Chapter 24

- 1) The relationship is 1:1 (one to one)
- 2) Creation of symbol table, macro expansion, and evaluation of constants and expressions
- 3) Final generation of code, create list file if requested, create object files
- 4) Combine one or more object files into a single executable, update all relocation addresses, search user and system libraries, create cross reference file if requested, and create final executable file.
- 5) Attempt to open executable file (verifying existence and permissions), read header information, ask operating system to create new process, if successful, read rest of executable file and load into memory (where specified by operating system), and inform operating system when load is completed. Note, the loader does not run the process.

## Appendix D – Quiz Question Answers

- 6) Examples might include:  
1. BUFSIZE  
2. MAX + OFFS

Note, assumes that upper case implies defined constant. Many examples possible.

- 7) See Section 5.1, Illustration 4, Overview: Assemble, Link, Load  
8) At run-time  
9) The symbol name and the symbol address

### 24.6 Quiz Question Answers, Chapter 24

- 1) By typing: ddd <progName>  
2) The “-g” option  
3) Executes the program, always starting from the beginning  
4) The continue command continues to the next breakpoint  
5) Via the menu option Status → Registers  
6) The first is the register name, the second is the hex representation of the value, and the third is the decimal representation of the value (excluding some registers such as rip and rsp which are always shown in hex).  
7) There are multiple ways to exit the debugger including typing exit in the command window, clicking the x (upper left corner), or using the menu options File → Exit.  
8) There are multiple ways to set a breakpoint including double-clicking on a line, typing b <lineNumber>, or typing b <labelName> (if a label exists on the desired line).  
9) The debugger command to read commands from a file is; source <fileName>  
10) The green arrow points to the next instruction to be executed  
11) The answers are as follows:  
1. x/db &bV  
2. x/dh &wV  
3. x/dw &dV



4. x/dg &qV
5. x/30db &bA
6. x/50dh &wA
7. x/75dw &dA

12) The answers are as follo

1. x xb &bV
2. x xh &wV
3. x xw &dV
4. x xg &qV
5. x/30xb &bA
6. x/50xh &wA
7. x/75xw &dA

13) The command is: x/ug

14) The command is: x/5ug

## 24.7 Quiz Question Answers, Chapter 24



1) The answers are as follo

1. L
2. L
3. Illegal, 354 does not fit into a
4. L
5. Illegal, sizes do not m
6. Illegal, cannot change the val
7. L
8. Legal, while legal it would probably result in an incorrect v
9. L
10. L



## Appendix D – Quiz Question Ans

11. Legal, while legal it would probably result in an incorrect value.  
12. Illegal, cannot move memory to memory.  
13. Illegal, cannot move memory to memory.  
14. Legal.  
15. Illegal, **r16** is not a valid register.  
16. Legal.
- 2) The answers are as follows:
1. Copies the byte value at **bVar1** into the **rsi** register treating as an unsigned value thus setting the upper 56 bits to zero.  
2. Copies the byte value at **bVar1** into the **rsi** register treating as a signed value thus sign extending the upper 56-bits (1's for negative, 0's for positive).
- 3) The answers are as follows:
1. `mov a`
  2. `cl`
- 4) The answers are as follows:
1. `movzx eax`
  2. `cwd`
- 5) The answers are as follows:
1. `mov d`
  2. `cl`
- 6) The **cwd** instruction only converts the signed value in **ax** into a sign value in **dx:ax** (and nothing else). The **movsx** instruction copies the word source operand into the double-word destination operand.
- 7) On the first instruction, the destination operand size must be explicitly specified since the source operand, an immediate value of 1, does not have an inherent size associated with it. On the second instruction, the destination operand size can be determined from the source operand (since the **eax** register is a double word in this case).



- 8) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rbx** = 0x0000 0000 0000 0000
- 9) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rbx** = 0x0000 0000 0000 0000
- 10) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rbx** = 0xFFFF FFFF FFFF FFFF
- 11) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rdx** = 0x0000 0000 0000 0000
- 12) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rdx** = 0x0000 0000 0000 0000
- 13) The answers are as follows (grouped in sets of 4 for clarity)
1. **rax** = 0x0000 0000 0000 0000
  2. **rdx** = 0x0000 0000 0000 0000
- 14) The answers are as follows
1. The destination operand cannot be an immediate value
  2. An immediate operand is not allowed since the size/type cannot be determined
  3. The **mov** instruction does not allow a memory to memory operation
  4. An address requires 64-bits which will not fit into the **ax** register
- 15) The **idiv** instruction will divide by **edx:eax** and **edx** was not
- 16) The operands for the divide a signed (-500), but an unsigned divide is used
- 17) The word divide used will place the result in the **dx:ax** registers, but the **dx** register is used to obtain the remainder

## Appendix D – Quiz Question Answers

- 18) The three-operand multiply instructions are only allowed for a limited set of signed multiplication operations.

### 24.8 Quiz Question Answers, Chapter 24

- 1) The first instruction places the value from qVar1 into the **rdx** register. The second instruction places the address of qVar1 into the **rdx** register.
- 2) The answers are as follows:
  1. Immediate
  2. Memory
  3. Immediate
  4. Illegal, destination operand cannot be an immediate value
  5. Register
  6. Memory
  7. Memory
  8. Illegal, source and destination operands not the same size
- 3) The answers are as follows (grouped in sets of 4 for clarity).
  1. **eax** = 0x0000 0000 0000 0000
  - 4) The answers are as follows (grouped in sets of 4 for clarity).
    1. **eax** = 0x0000 0000 0000 0000
    2. **edx** = 0x0000 0000 0000 0000
    - 5) The answers are as follows (grouped in sets of 4 for clarity).
      1. **eax** = 0x0000 0000 0000 0000
      2. **ebx** = 0x0000 0000 0000 0000
      3. **rcx** = 0x0000 0000 0000 0000
      4. **rsi** = 0x0000 0000 0000 0000
      - 6) The answers are as follows (grouped in sets of 4 for clarity).
        1. **rax** = 0x0000 0000 0000 0000
        2. **rcx** = 0x0000 0000 0000 0000



3. **edx** = 0x0000 0 [ANSWER]
  4. **rsi** = 0x0000 0000 0000 0 [ANSWER]
- 7) The answers are as follows (grouped in sets of 4 for clarity) [ANSWER]
1. **eax** = 0x0000 0 [ANSWER]
  2. **rcx** = 0x0000 0000 0000 0 [ANSWER]
  3. **edx** = 0x0000 0 [ANSWER]
  4. **rsi** = 0x0000 0000 0000 0 [ANSWER]
- 8) The answers are as follows (grouped in sets of 4 for clarity) [ANSWER]
1. **eax** = 0x0000 0 [ANSWER]
  2. **edx** = 0x0000 0 [ANSWER]
  3. **rcx** = 0x0000 0000 0000 0 [ANSWER]
  4. **rsi** = 0x0000 0000 0000 0 [ANSWER]

## 24.9 Quiz Question Answers, Chapter 24

- 1) The **rsp** register [ANSWER]
  - 2) First, **rsp** = **rsp** - 8 and then **rax** register is copied to [**rsp**] (in that order) [ANSWER]
  - 3) 8 bytes [ANSWER]
- 4) The answers are as follows (grouped in sets of 4 for clarity) [ANSWER]
1. **r10** = 0x0000 0000 0000 0 [ANSWER]
  2. **r11** = 0x0000 0000 0000 0 [ANSWER]
  3. **r12** = 0x0000 0000 0000 0 [ANSWER]
- 5) The array is reversed in memory [ANSWER]
- 6) Memory is used more efficiently [ANSWER]

## 24.10 Quiz Question Answers, Chapter 24

- 1) An unambiguous, ordered sequence of steps involved in solving a problem [ANSWER]
- 2) The answer is as follows [ANSWER]
  1. Understand the Problem [ANSWER]



## Appendix D – Quiz Question Answers

2. Create the Algorithm
  3. Implement the Program
  4. Test/Debug the Program
- 3) No, the steps are applicable to any language or complex problem (even beyond programming)
- 4) An assemble-time error
  - 5) An assemble-time error
  - 6) An assemble-time error
  - 7) A run-time error

### 24.11 Quiz Question Answers, Chapter 24

- 1) At the top (above the data, BSS, and text sections)
- 2) Once for each time the macro is invoked
- 3) The %% will ensure that a unique label name is generated each time the macro is used
- 4) If the %% is omitted on a label, the label will be copied, as is, and thus appear to be duplicated
- 5) Yes. This might be used to exit a macro to an error handling code block within the macro
- 6) The macro argument substitution occurs at assemble-time

### 24.12 Quiz Question Answers, Chapter 24

- 1) Linkage and Argument Transmission
- 2) The **call** and the **ret** instructions
- 3) Call-by-value
- 4) Call-by-reference
- 5) Once, regardless of how many times it is called
- 6) The current **rip** is placed on the stack and the **rip** is changed to the address of the called function
- 7) Save and restore the contents of the callee preserved registers

- 8) In: **rdi, rsi, rdx, rcx, r8**, and   

- 9) In: **edi, esi, edx, ecx, r8d**, and   

- 10) That a function may change the value in the register without needing to save   
 restore 
- 11) Two of: **rax, rcx, rdx, rsi, rdi, r8, r9, r10**, and   

- 12) Call frame, function call frame, or activation record   

- 13) A leaf function does not call other functions   

- 14) Clear the passed arguments off the stack   

- 15) Twenty-four (24) since three arguments at 8 bytes each ( $8 \times 3 = 24$ )   

- 16) The offset is **[rbp+16]** regardless of which saved registers are pushed   

- 17) Available memory   

- 18) Call-by-reference   

- 19) The 7<sup>th</sup> argument is at **[rbp+16]** and the 8<sup>th</sup> argument is at **[rbp+32]**   

- 20) Memory efficiency since stack dynamic local variables only use memory when needed (when the function is being executed)   


### 24.13 Quiz Question Answers, Chapter 24

- 1) The **rax** register   

- 2) The operating system   

- 3) The call code is **SYS\_write** (1). The 1<sup>st</sup> argument in **rdi** is the output location to **STDOUT**, the 2<sup>nd</sup> argument in **rsi** is the starting address of the characters to output, and the 3<sup>rd</sup> argument in **rdx** is the number of characters to write.   

- 4) It is unknown how many characters will be entered   

- 5) The **rax** register will contain the file descriptor   

- 6) The **rax** register will contain an error code   

- 7) In: **rdi, rsi, rdx, r10, r8, and r9**, and   


### 24.14 Quiz Question Answers, Chapter 24

- 1) The statement is: **extern func1, func2**   


## Appendix D – Quiz Question Answers

- 2) The statement is: **extern func1, fu** [ANSWER]
- 3) The assembler will generate an e [ANSWER]
- 4) Link-ti [ANSWER]
- 5) The linker will generate an unsatisfied external reference e [ANSWER]
- 6) Yes. However, in the debugger, the code would not be displa [ANSWER]

## 24.15 Quiz Question Answers, Chapter 24

- 1) The buffer overflow exploit is typically called *stack smash* [ANSWER]
- 2) The C function does not check the array bounds of the input argument [ANSWER]
- 3) [ANSWER]
- 4) Typing a very large number of characters when input is requested and, if the program cras [ANSWER]
- 5) A series of **nop** instructions designed to make the target of a buffer overf [ANSWER]
- 6) Many possible answers. Delete a file, open a network connection, kill a pro [ANSWER]
- 7) Use of canaries, implementation of Data Execution Prevention (DEP), and us [ANSWER]  
Data Address Space Layout Randomizat [ANSWER]

## 24.16 Quiz Question Answers, Chapter 24

- 1) The operating system. Specifically, the load [ANSWER]
- 2) The program being execut [ANSWER]
- 3) The name of the executable [ANSWER]
- 4) The **argc** refers to argument count and **argv** refers to the argument ve [ANSWER]  
(starting address of the table of addresses for the string representing each argument). [ANSWER]
- 5) In the **rdi** register [ANSWER]
- 6) In the **rsi** register [ANSWER]
- 7) The spaces are removed by the operating system so the program does not hav [ANSWER]  
do anything [ANSWER]



- 8) No. The program is required to check and determine if that is an e



## 24.17 Quiz Question Answers, Chapter



- 1) The end of line character for Linux is linefeed (LF) and the end of line character for Windows is carriage return, line feed (CR, LF)
- 2) Store a subset of the information for quick access
- 3) They are in the language I/O library functions (i.e., cout, cin, etc.)
- 4) Simplify the program
- 5) I/O performance improvement
- 6) The system service functions require a specific number of characters to read which is not known ahead of time for “one line” of input
- 7) Keeps a subset of the information at the next higher level in the hierarchy (which is faster than the next lower level)
- 8) Reduces the overhead associated bus contention and memory latency from excessive system reads
- 9) The variable values must be retained between function calls
- 10) The end of file must be inferred from the number of characters actually read
- 11) Many reasons possible, including file being deleted in another Window after it was open or the drive (USB) being removed after the operation
- 12) The actual number of characters read will be 0 which must be checked explicitly
- 13) To ensure the passed line buffer array is not overwritten
- 14) By initializing the variables to indicate that all buffer characters have been read



## 24.18 Quiz Question Answers, Chapter



- 1) The registers are: **xmm0, xmm1, xmm2, . . . , xmm7**
- 2) Single precision is 32-bit and double precision is 64-bit
- 3) Cumulative rounding error associated with the inexact representation of 0. binary
- 4) Float functions return the value in **xmm0**
- 5) None of the floating-point registers are preserved





## 24.19 Quiz Question Answers, Chapter

- 1) Concurrency implies multiple different (not necessarily related) processes simultaneously making progress. Parallel processing implies that processes are executing simultaneously.
- 2) Distributed computing and multiprocessor
- 3) On different computers connected via a network
- 4) Many possible answers, including Folding@Home and SETI@Home. Internet search can provide a more complete list.
- 5) On different cores in the CPU
- 6) Distributed computing allows a very large number of compute nodes but requires communications over a network which has inherent communication delays.
- 7) Multiprocessing allows very fast communications between processes via shared memory but supports only a limited amount of simultaneous executing threads related to the number of cores available.
- 8) Multiple threads simultaneously writing to a shared variable with no coordination
- 9) No. No problem exists since the variable is not being changed.
- 10) Yes. Since the variable is being changed, one thread may alter the value before the other has obtained the value.

## 24.20 Quiz Question Answers, Chapter

- 1) The operating system is responsible for managing the resources. The resources include CPU cores, primary memory, secondary storage, display screen, keyboard, and mouse.
- 2) An event that alters the sequence of instructions executed by a process.
- 3) An interrupt that is caused by the current process and needs attention of the kernel.
- 4) An ISR is an Interrupt Service Routine which is executed when an interrupt occurs to service (perform required actions) that interrupt.
- 5) Interrupt Descriptor Table (IDT) which contains the addresses of the Interrupt Service Routines (ISRs) and the gate information.

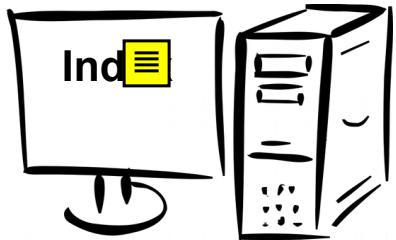




- 6) The interrupt number is multiplied by
- 7) The **ret** instruction will pop the return address from the stack and place it in **rip** register. The **iret** instruction will pop the return address and the preserved flag register contents from the stack and place it in the **rip** register and **rFlag** registers.
- 8) The call requires the target address. Since the ISR addresses may change due to hardware changes or software updates, the interrupt mechanism performs a run time look up for the ISR address.
- 9) That the interrupt timing, when or even if the interrupt might occur, cannot be predicted in the context of the executing code
- 10) That the interrupt timing can be predicted in the context of the executing code. This is typical of system service calls or exceptions such as division by zero
- 11) Each instruction changes the **rFlag** register. After the interrupt is completed, the flag register must be restored to its original value to ensure that the interrupted process is able to resume.
- 12) Many possible answers, including I/O devices such as keyboard and mouse, network adapter, secondary storage devices, or other peripheral devices
- 13) Many possible answers, including dividing by zero
- 14) A maskable interrupt may be ignored briefly where a non-maskable interrupt must be handled immediately

Appendix D – Quiz Question Ans





## 25.0 Alphabetical Ind<sup>█</sup>

---

0x.....	70	automatics.....	168
53.....	280	base address.....	128
activation record.....	174	Base Pointer Register.....	12
add with carry.....	81	Basic System Services.....	315
Addresses and Values.....	73, 125	biased exponent.....	25
Addressing Modes.....	125	BSS.....	17
American Standard Code for Information Interchange.....	28	BSS Section.....	35
Architecture Overview.....	7	buffer.....	250
argc.....	240	Buffering Algorithm.....	251
Argument Count.....	241	byte addressable.....	15, 128
Argument Transmission.....	171	cache memory.....	9
Argument Vector Table.....	241	Cache Memory.....	14, 19
argv.....	240	call code.....	191
Arithmetic Logic Unit.....	9	call frame.....	174
ASCII.....	28	Call Frame.....	174
Assemble and Link.....	223	call-by-reference.....	171
Assemble Commands.....	43	call-by-value.....	171
Assemble/Link Script.....	50	callee.....	171
Assemble/Link/Load Overview.....	41	Callee.....	177, 180
Assembler.....	43	caller.....	171
Assembler Directives.....	47	Caller.....	177, 179
Assembler Error.....	157	Calling Convention.....	171
Assembly Main.....	220	Calling System Services.....	191
assembly source file.....	33	Central Processing Unit.....	9
Asynchronous Interrupts.....		character.....	28
		code generation.....	

Code Injection.....	234	Destination operand.....	70, 299
Code Injection Protections.....	235	Direct Memory Access.....	250
Code to Inject.....	231	Displaying Register Contents.....	60
Command Line Arguments.....	239	Displaying Stack Contents.....	65
Comments.....	33	distributed computing.....	280
Compile, Assemble, and Link.....	225	Distributed Computing.....	280
concurrency.....	279	distributed processing.....	280
Conditional Control Instructions.....	109	DMA.....	250
Console Input.....	197	dq.....	35
Console Output.....	193	dt.....	35
constant expression.....	46	dw.....	34
Conversion Instructions.....	74	dynamic linking.....	49
CPU register.....	10	Dynamic Linking.....	49
Data Address Space Layout		epilogue.....	171
Randomization.....	236	Error Codes.....	318
Data Execution Prevention.....	236	Error Terminology.....	157
Data Movement.....	71, 258	Example Program.....	37
Data representation.....	21	Example Program, Absolute Value.....	276
Data Section.....	34	Example Program, List Summation.....	129
Data Stack Smashing Protector (or Canaries).....	235	Example Program, Pyramid Areas and Volumes.....	131
db.....	34	Example Program, Sum and Average. ....	275
dd.....	35	Example, C++ Main / Assembly Function	
DDD Configuration Settings.....	57	.....	224
DDD Debugger.....	55	Example, Console Input.....	198
DDD/GDB Commands Summary.....	62	Example, Console Output.....	194
DDD/GDB Commands, Examples.....	63	Example, File Read.....	211
ddq.....	35	Example, File Write.....	205
Debugger.....	52	Example, Statistical Function 1 (leaf). ..	176
Debugger Command, <i>next</i> .....	168	Example, Statistical Function2 (non-leaf)	
Debugger Command, <i>step</i> .....	168	.....	178
Debugger Commands.....	168	Example, Sum and Average.....	220
Debugger Commands File (interactive).....	65	Exceptions.....	289
Debugger Commands File (non- interactive).....	66	Executing Programs.....	58
Debugging Macros.....	165	extern.....	219
Defining Constants.....	34	Extern Statement.....	219
dereferencing.....		external reference.....	48
		File Control Block.....	

file descriptor.....	193, 197, 202, 204
File Modes.....	317
File Open.....	202
File Open Operations.....	202
File Open/Create.....	203
File Operations Examples.....	205
File Read.....	204
File Write.....	205
First Pass.....	46
Flag Register.....	12
Floating-Point Addition.....	262
Floating-Point Arithmetic Instructions.	262
Floating-Point Calling Conventions....	274
Floating-Point Comparison.....	271
Floating-Point Control Instructions.....	271
Floating-point destination register operand .....	70, 299
Floating-Point Division.....	267
Floating-Point Instructions.....	257
Floating-Point Multiplication.....	265
Floating-Point Registers.....	258
Floating-point Representation.....	24
Floating-Point Square Root.....	269
Floating-Point Subtraction.....	263
Floating-Point Values.....	257
forward reference.....	45
Function Declaration.....	169
Function Source.....	222
Functions.....	167
General Purpose Registers.....	10
GPRs.....	10
Hardware Interrupt.....	289
heap.....	17, 144
I/O buffering.....	249
IEEE 32-bit Representation.....	25
IEEE 64-bit Representation.....	27
IEEE 754 32-bit floating-point standard	24
IF statement.....	24
Immediate Mode Addressing.....	126
Immediate value.....	70, 299
index.....	128
indirection.....	127
input buffer.....	250
Input/Output Buffering.....	249
Instruction Pointer Register.....	12
Instruction Set Overview.....	69
Integer / Floating-Point Conversion	
Instructions.....	260
Integer Arithmetic Instructions.....	78
integer numbers.....	21
Interfacing with a High-Level Language .....	224
Interrupt Categories.....	289
Interrupt Classification.....	288
Interrupt Processing.....	292
Interrupt Service Routine (ISR).....	292
Interrupt Timing.....	288
Interrupt Types.....	290
Interrupt Types and Levels.....	290
Interrupts.....	287
Iteration.....	115
Jump Out of Range.....	112
Jump to ISR.....	293
Labels.....	109
leaf function.....	174
Least Significant Byte.....	15
light-weight process.....	281
Linkage.....	170
Linker.....	47
Linking Multiple Files.....	48
Linking Process.....	48
list file.....	43
List File.....	43
List Summation.....	129
little-endian.....	15p., 146, 233
Loader.....	24

## Alphabetical Index

Logic Error.....	157	operation.....	69
logical AND operation.....	103	ordered floating-point comparisons.....	271
logical NOT operation.....	103	parallel processing.....	279
logical OR operation.....	103	Parallel Processing.....	279
logical XOR operation.....	103	Parameter Passing.....	172
machine language.....	17, 43, 45, 232	Parsing Command Line Arguments.....	239
Macro Definition.....	162	pop operation.....	141
Macros.....	161	POSIX Threads.....	281
Main Memory.....	15	preserved register.....	173
memory hierarchy.....	17, 249	Primary Storage.....	7, 19
Memory Hierarchy.....	17	Privilege Levels.....	290
memory layout.....	17, 143	Process Stack.....	141
Memory Mode Addressing.....	126	Processing Steps.....	292
Most Significant Byte.....	15	processor registers.....	9
Multi-Line Macros.....	162	Program Development.....	151
Multi-user Operating System.....	287	Program Format.....	33
Multiple Source Files.....	219	prologue.....	171
multiprocessing.....	280	push operation.....	141
Multiprocessing.....	280	race condition.....	281
NaN.....	27, 271	Race Conditions.....	282
Narrowing conversions.....	74	Random Access Memory (RAM).....	7
Narrowing Conversions.....	74	RBP .....	12
Newline Character.....	192	Red Zone.....	176
Next / Step.....	60	register.....	10
nexti.....	165	Register Mode Addressing.....	126
non-volatile memory.....	8	Register operand.....	70, 299
NOP slide.....	234	Register Usage.....	173
normalized scientific notation.....	25	resb.....	36
not a number.....	27, 271	resd.....	36
Not a Number (NaN).....	27	resdq.....	36
Notation.....	299	resq.....	36
Notational Conventions.....	69	Resumption.....	294
Numeric Values.....	33	resw.....	36
object file.....	43	Return Codes.....	315
Obtaining ISR Address.....	292	rFlags.....	12
offset.....	128	RIP.....	12
Operand Notation.....	70, 299	RSP.....	12
operands.....		Run / Continue.....	12

Run-time Error.....	157	Starting DDD.....	55
S-NaN.....	271	stepi.....	165
saved register.....	173	string.....	29
Second Pass.....	46	Summary.....	186
secondary storage.....	8	Suspension.....	292
Secondary Storage.....	7, 19	Suspension Execute ISR.....	293
section .bss.....	35	Suspension Interrupt Processing Summary .....	294
section .data.....	34	symbol table.....	46
section .text.....	36	Synchronous Interrupts.....	288
segment fault.....	231	system call.....	191
Setting Breakpoints.....	57	System Services.....	191
short-jump.....	112	Text Section.....	36
sign extension.....	105	thread.....	281
sign-extend.....	76	Tool Chain.....	41
signed.....	21	Two-Pass Assembler.....	45
Signed Conversions.....	76	two's complement.....	22p.
Single Instruction Multiple Data.....	13	Unconditional Control Instructions.....	109
Single-Line Macros.....	161	Understanding a Stack Buffer Overflow .....	230
Software Interrupts.....	290	Unicode.....	29
source file.....	43	uninitialized data.....	17, 144
Source operand.....	70, 299	unordered floating-point comparisons.....	271
stack buffer overflow.....	229	unsafe function.....	231
Stack Buffer Overflow.....	229	unsigned.....	21
Stack Dynamic Local Variables.....	168	Unsigned Conversions.....	74
Stack Example.....	147	Updated Linking Instructions.....	167
stack frame.....	174	Using a Macro.....	162
Stack Implementation.....	143	volatile memory.....	8
Stack Instructions.....	142	Why Buffer?.....	249
Stack Layout.....	143	Widening conversions.....	74
Stack Operations.....	145	Widening Conversions.....	74
Stack Pointer Register.....	12	%define.....	
stack smashing.....	229		
Stack-Based Local Variables.....	183		
Standard Calling Convention.....			