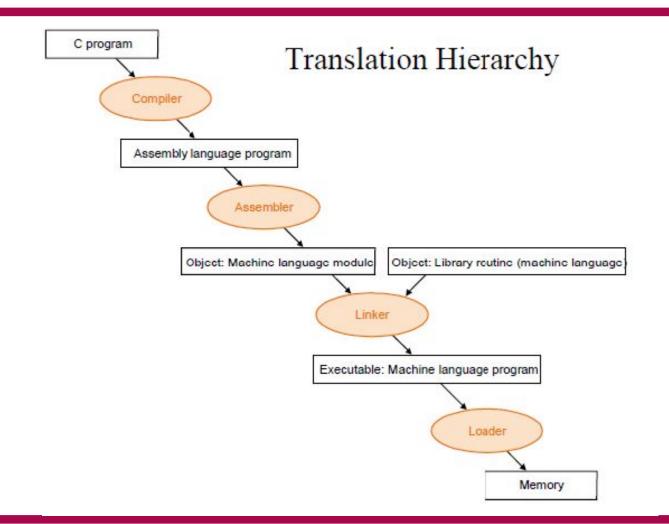
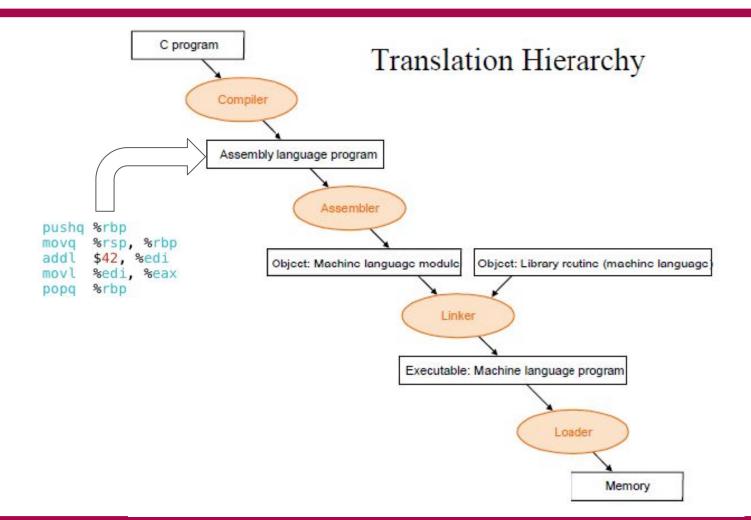
# A Beginners guide to Assembly

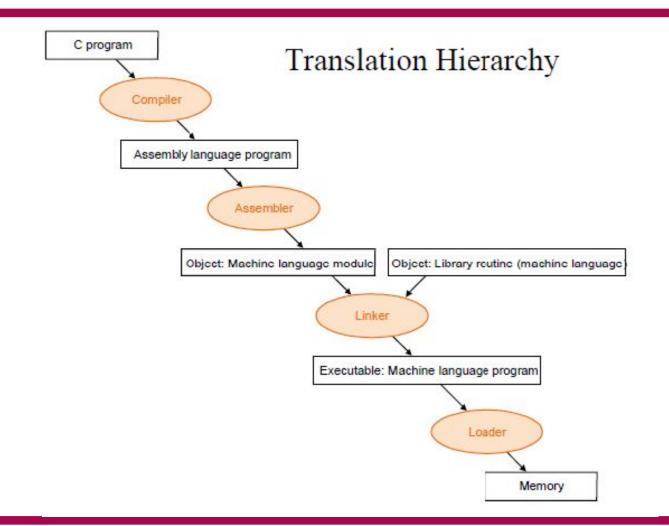
By Tom Glint and Rishiraj CS301 | Fall 2020

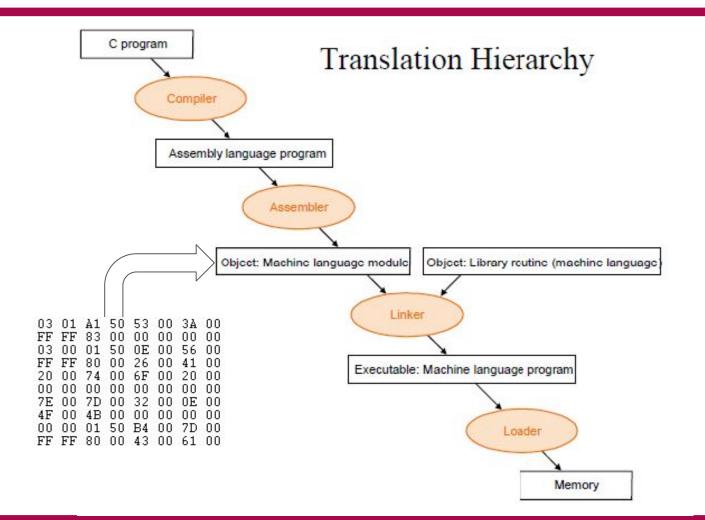
Contributors: Varun and Shreyas

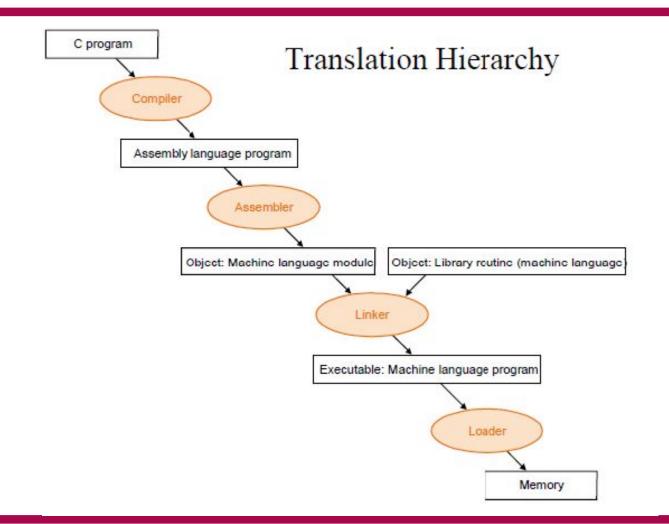


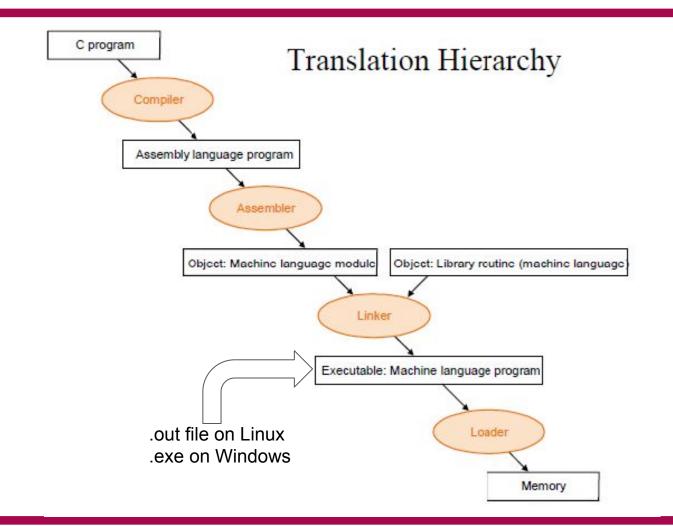


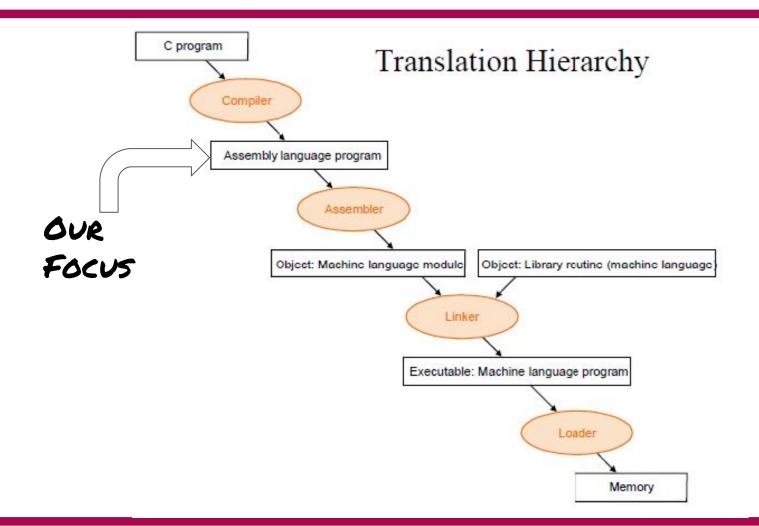












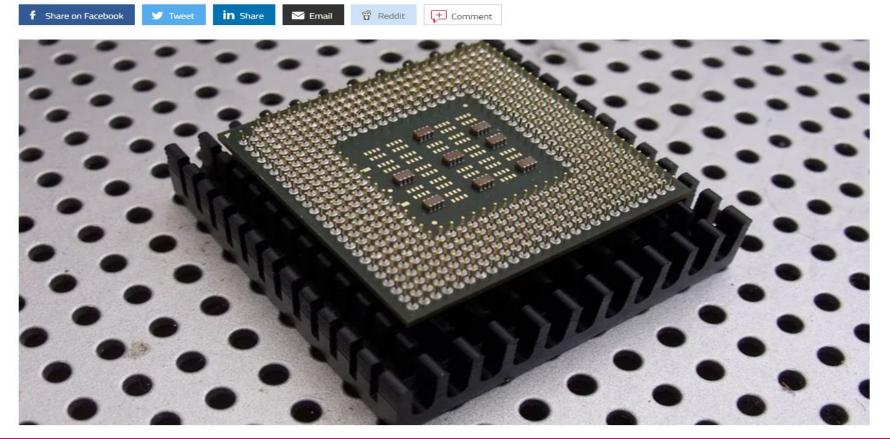
### Prominent ISAs





# ARM

# IIT-Madras Develops 'India's First Microprocessor', Shakti By Indo-Asian News Service | Updated: 2 November 2018 16:03 IST



```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
               %rbp
       pushq
               %rsp, %rbp
       movq
       movl
               $3000, -8(%rbp)
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

### Some Basics

- % indicates register names. Example : %rbp
- \$ indicates constants Example: \$100
- Accessing register values:
  - %rbp : Access value stored in register rbp
  - (%rbp) : Treat value stored in register rbp as a pointer. Access the value stored at address pointed by the pointer. Basically \*rbp
  - 4(%rbp): Access value stored at address which is 4 bytes after the address stored in rbp. Basically \*(rbp + 4)

```
main:
#include<stdio.h>
                                                        %гьр
                                                 pushq
                                                        %rsp, %rbp
                                                 movq
int main(){
                                                 movl
                                                         $3000, -8(%rbp)
        int x = 3000, y;
                                                 addl
                                                         $3, -8(%rbp)
        X = X + 3;
                                                 movl
                                                        $100, -4(%rbp)
                                                         -4(%rbp), %eax
                                                 movl
        y = 100;
                                                        %гьр
                                                 popq
        return y;
                                                 ret
```

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
               %rbp
       pushq
               %rsp, %rbp
       pvom
       movl
                $3000, -8(%rbp)
       addl
                $3, -8(%rbp)
       movl
                $100, -4(%rbp)
                -4(%rbp), %eax
       movl
               %гьр
       popq
        ret
```

For each function call, new space is created on the stack to store local variables and other data. This is known as a stack frame. To accomplish this, you will need to write some code at the beginning and end of each function to create and destroy the stack frame

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
               %rbp
       pushq
               %rsp, %rbp
       pvom
       movl
               $3000, -8(%rbp)
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

**rbp** is the frame pointer. In our code, it gets a snapshot of the stack pointer (**rsp**) so that when rsp is changed, local variables and function parameters are still accessible from a constant offset from **rbp**.

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
       pushq
               %rbp
               %rsp, %rbp
       movq
               $3000, -8(%rbp)
      → movl
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

move immediate value 3000 to (%rbp-8)

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
       pushq
               %rbp
               %rsp, %rbp
       movq
               $3000, -8(%rbp)
       movl
      ⇒ addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

add immediate value 3 to (%rbp-8)

# <u>An intriguing Example!</u>

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
       pushq
               %rbp
               %rsp, %rbp
       movq
               $3000, -8(%rbp)
       movl
       addl
               $3, -8(%rbp)
      → movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

Move immediate value 100 to (%rbp-4)

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
       pushq
               %rbp
               %rsp, %rbp
       movq
               $3000, -8(%rbp)
       movl
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
      → movl
               %гьр
       popq
       ret
```

Move (%rbp-4) to auxiliary register

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
       pushq
               %rbp
               %rsp, %rbp
       movq
       movl
               $3000, -8(%rbp)
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
      popq
       ret
```

Pop the base pointer to restore state

# <u>An intriguing Example!</u>

```
#include<stdio.h>
int main(){
    int x = 3000,y;
    x = x + 3;
    y = 100;
    return y|;
}
```

```
main:
               %rbp
       pushq
               %rsp, %rbp
       pvom
               $3000, -8(%rbp)
       movl
       addl
               $3, -8(%rbp)
       movl
               $100, -4(%rbp)
               -4(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

The calling convention dictates that a function's return value is stored in %eax, so the above instruction sets us up to return y at the end of our function.

# Operation Suffixes

- b = byte (8 bit)
- s = single (32-bit floating point)
- w = word (16 bit)
- I = long (32 bit integer or 64-bit floating point)
- q = quad (64 bit)
- t = ten bytes (80-bit floating point)

# How to get assembly code?

### Two ways:

- While Compiling
  - Use -S flag with gcc. WIll create a .s file containing assembly

- Using Binary
  - O Use objdump. Will show the assembly in terminal.

# Understanding the output

- The output will have assembly, but there is more information!
- You will see lots of Directives like:
  - o .file
  - o .text
  - o .global name

# Understanding the output

- The output will have assembly, but there is more information also!.
- You will see lots of Directives like:
  - o .file
  - .text
  - global name



To disable these, use the gcc option

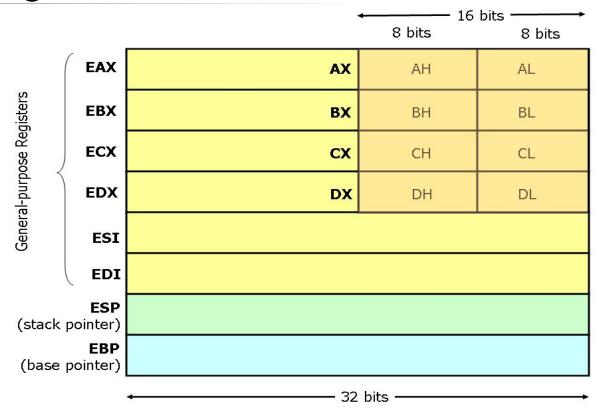
137

-fno-asynchronous-unwind-tables



Note, I know this is a really old thread, but this is the top result on google for cfi\_startproc, so many people probably come here to disable that output.

# x86 Register Set



# x86 Register Set: A few more

- Registers starting with "r"
  - Same as "e" registers but 64 bits wide
- EIP: The Instruction Pointer or the Program Counter

# An Example with Loops!

```
#include<stdio.h>
int main(){
    int x = 0;
    for(int i=0;i<10;i++){
        x = x + 1;
    }
    return x;
}</pre>
```

```
main:
               %гьр
       pushq
               %rsp, %rbp
       movq
       movl
               $0, -8(%rbp)
       movl
               $0, -4(%rbp)
       jmp
               .L2
.L3:
       addl
               $1, -8(%rbp)
       addl
               $1, -4(%rbp)
.L2:
       cmpl
               $9, -4(%rbp)
       jle
               .L3
               -8(%rbp), %eax
       movl
               %гьр
       popq
       ret
```

# System Calls in Assembly

```
kernel:

int 80h; //Call kernel
ret

open:

push dword mode
push dword flags
push dword path
mov eax, 5
call kernel
add esp, byte 12
ret
```

# System Calls in Assembly

```
int 80h; //Call kernel
ret

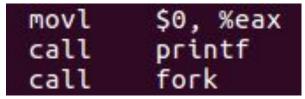
open:

push dword mode
push dword flags
push dword path
mov eax, 5

call kernel
add esp, byte 12
ret
```

## A bit different!

A simple fork program



# Embedding Assembly in C

```
__asm__( "instruction 1", "instruction 2", ...)
Example:
  asm (
           "movl %edx, %eax\n\t"
           "addl $2, %eax\n\t"
```

# Embedding Assembly in C

```
#include <stdio.h>
int main(int argc, char *argv[]) {
  int x = 5;
  printf("x = %d\n", x);
  _asm__("add $10, %0":"=m"(x));
  printf("x = %d\n", x);
  return 0;
}
```

# Where will I use assembly?



### Where will I use assembly?

- To write Compilers and Device Drivers
- To write viruses and for malware analysis
- Used while programming Real Time Embedded systems
- Implementing Locks for Concurrency.
   We will cover this in the third module of the course!

### References

- Chapter 11. x86 Assembly Language Programming, FreeBSD, <a href="https://www.freebsd.org/doc/en\_US.ISO8859-1/books/developers-handbook/x86.html">https://www.freebsd.org/doc/en\_US.ISO8859-1/books/developers-handbook/x86.html</a>
- Easy x86-64, <a href="http://ian.seyler.me/easy-x86-64/">http://ian.seyler.me/easy-x86-64/</a>
- Introduction to the GNU/Linux assembler and linker for Intel Pentium processors, <a href="https://www.cs.usfca.edu/~cruse/cs210s07/lesson01.ppt">https://www.cs.usfca.edu/~cruse/cs210s07/lesson01.ppt</a>
- Is there a way to insert assembly code into C?, <a href="https://stackoverflow.com/questions/61341/is-there-a-way-to-insert-assembly-code-int-o-c">https://stackoverflow.com/questions/61341/is-there-a-way-to-insert-assembly-code-int-o-c</a>

## More Topics

- GCC
- Clang
- GCC vs Clang
  - (More: <a href="https://opensource.apple.com/source/clang/clang-23/clang/tools/clang/www/comparison.html">https://opensource.apple.com/source/clang/clang-23/clang/tools/clang/www/comparison.html</a>)
- Debugger
- How to use a debugger (the practical way)

### **GCC**

- The GNU Compiler Collection (GCC) is a compiler system produced by the GNU Project supporting various programming languages.
- GCC is a key component of the GNU toolchain and the standard compiler for most projects related to GNU and Linux, including the Linux kernel

# Clang

- Clang is a compiler front end for the C, C++, Objective-C and Objective-C++
  programming languages, as well as the OpenMP, OpenCL, RenderScript, CUDA
  ...
- It uses the LLVM compiler infrastructure as its back end and has been part of the LLVM release cycle since LLVM 2.6

### **LLVM**

- The LLVM compiler infrastructure project is a set of compiler and toolchain technologies, which can be used to develop a front end for any programming language and a back end for any instruction set architecture.
- LLVM is designed around a language-independent intermediate representation (IR) that serves as a portable, high-level assembly language that can be optimized with a variety of transformations over multiple passes.

#### Benefits of Using GCC

- supports languages that clang does not aim to, such as Java, Ada, FORTRAN, etc.
- front-ends are very mature and already support C++. clang's support for C++ is nowhere near what GCC supports.
- supports more targets than LLVM.
- popular and widely adopted.
- GCC does not require a C++ compiler to build it.

#### Benefits of Clang

- The Clang ASTs and design are intended to be easily understandable by anyone who is familiar with the languages involved and who has a basic understanding of how a compiler works.
- Clang is designed as an API from its inception, allowing it to be reused by source analysis tools, refactoring, IDEs (etc) as well as for code generation

#### Benefits of Clang

- Clang aims to provide extremely clear and concise diagnostics (error and warning messages), and includes support for expressive diagnostics.
- Clang inherits a number of features from its use of LLVM as a backend, including support for a bytecode representation for intermediate code, pluggable optimizers, link-time optimization support, Just-In-Time compilation, ability to link in multiple code generators, etc.

```
description of the state o
```

```
glint@Mark4:~$ gcc main.cpp
main.cpp:3:20: error: expected ';' before 'int'
    3 | using namespace std
    4
       int main(){
glint@Mark4:~$ clang main.cpp
main.cpp:3:20: error: expected ';' after namespace name
using namespace std
1 error generated.
glint@Mark4:~$
```

### Debugger

- A debugger or debugging tool is a computer program used to test and debug other programs (the "target" program).
- The main use of a debugger is to run the target program under controlled conditions
  - that permit the programmer to track its operations in progress
  - monitor changes in computer resources
  - most often memory areas used by the target program or the computer's operating system that may indicate malfunctioning code.

#### **GNU GDB**

• The GNU Debugger (GDB) is a portable debugger that runs on many Unix-like systems and works for many programming languages, including Ada, C, C++, Objective-C, Free Pascal, Fortran, Go ...

# How to use GNU GDB (the impractical basics)

- https://u.osu.edu/cstutorials/2018/09/28/how-to-debug-c-program-using-gdb-in-6-simple-steps/
- Might end up as interview questions!

## How to use GNU GDB (the practical basics)

- Use an IDE or an advanced text editor with support for debugging
- Something that works on almost all OS and is open source and user extensible is Visual Studio Code.
  - Requirements
    - The source code Eg. main.cpp
    - A compiler Eg gcc
    - A debugger like gdb
    - An editor with support for debugging like Visual Studio Code.

