COMP1521

Week 1 - Memory & C revision

Introductions

Name

Degree and Year

Highlight from the break

Favourite MIPS syscall!!

Preference for chocolate: Twirl, Cadbury Milk or Crunchie?

Overview

Admin

Programming tools (Q5)

Memory (Q2, Q3, Q4)

Compilation (Q10)

C revision (Q6, Q8, Q9)

Admin

Assessment:

- Weekly labs (15%)
- Weekly tests (10%) (beginning in Week 3)
- Assignments (2 * 15%)
- Final (45%)

Challenge lab exercises (time for maths):

- There are nine weeks of labs. Each week has up to `2.0` possible marks
  - `1.6` is allocated evenly between the normal exercises
  - `0.4` is allocated evenly between the challenge exercises
- Getting full marks for all normal exercises would award you `1.6 * 9 = 14.4/15`
- So, you will need to do exactly `1.5` weeks of challenge exercises in order to be able to achieve a full `15/15` for the labs. Marks above `15` are capped at `15`.

Lab marking:

- Style is not marked for labs

Admin

Submission of work

- All work is submitted on the command line using give
- All required commands are provided in lab/assignment specs

Admin

Course website tour: https://cgi.cse.unsw.edu.au/~cs1521/25T1/

Help sessions:

- Help sessions will start running in wk3/4.
- Show up for either debugging help or help with content

Admin

Cat GPT and friends

Please don’t use Chat GPT to generate code for you (asking questions is fine of course). Why?

- This course teaches fundamental programming skills which you should learn yourself.
- Chat GPT produces poor code for the MIPS component of the course (first half).
- If use of Chat GPT (or similar) is detected you risk significant penalties.
- You can’t use it in the final
- If you use it I will be sad :(

Tools

Man pages (manual pages)

$ man 3 [name]

(3 = C library functions, see “man man” for more detail)

Contains:

- Arguments
- Return values
- What it does
- Sometimes example of usage

```
DESCRIPTION
       Never use this function.
```

```
SED(1)              User Commands              SED(1)

NAME
   sed - stream editor for filtering and
   transforming text

SYNOPSIS
   sed [OPTION]... {script-only-if-no-other-
   script} [input-file]...

DESCRIPTION
   Sed is a stream editor.  A stream editor is
   used to perform basic text transformations
   on an input stream (a file or input from a
   pipeline).  While in some ways similar to an
   editor which permits scripted edits (such as
   ed), sed works by making only one pass over
   the input(s), and is consequently more
   efficient.  But it is sed's ability to
   filter text in a pipeline which particularly
   distinguishes it from other types of
   editors.

   -n, --quiet, --silent

      suppress automatic printing of pattern
      space

   --debug

      annotate program execution

   -e script, --expression=script

      add the script to the commands to be
      executed
:
 ESC    ⇆    CTRL   ALT    —     ↓     ↑
```

Tute Questions

5. Write a c program `count_chars.c` that uses getchar to read in characters until the user enters Ctrl-D and then prints the total number of characters entered.

   Use `man 3 getchar` to look at the manual entry.

Makefiles

Instructions stored in Makefile

Convenient way to compile C programs.

Compile a single program:

“make program_name”

Compile all programs:

“make”

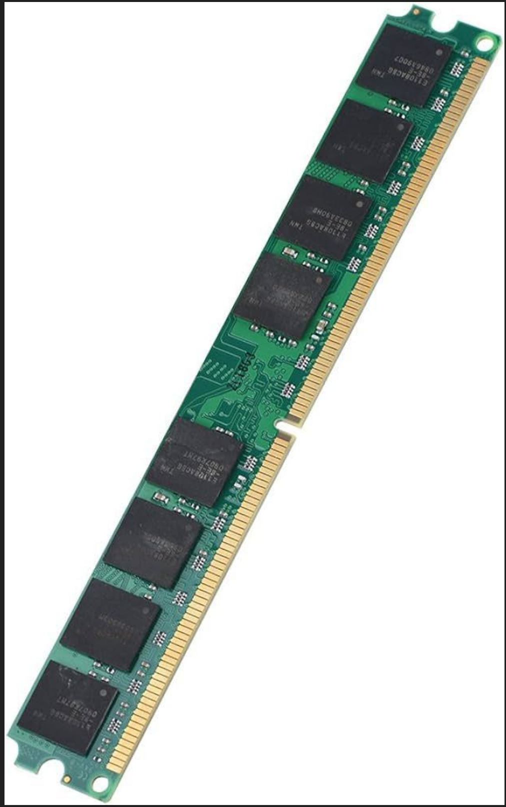
Pretty much just runs dcc for you

Not assessable, but useful for the exercises (we provide Makefiles for labs).

Memory in C

What is RAM?

Just a big 1D array that your computer can read and write to.



What is a process?

A smaller section of the RAM array.

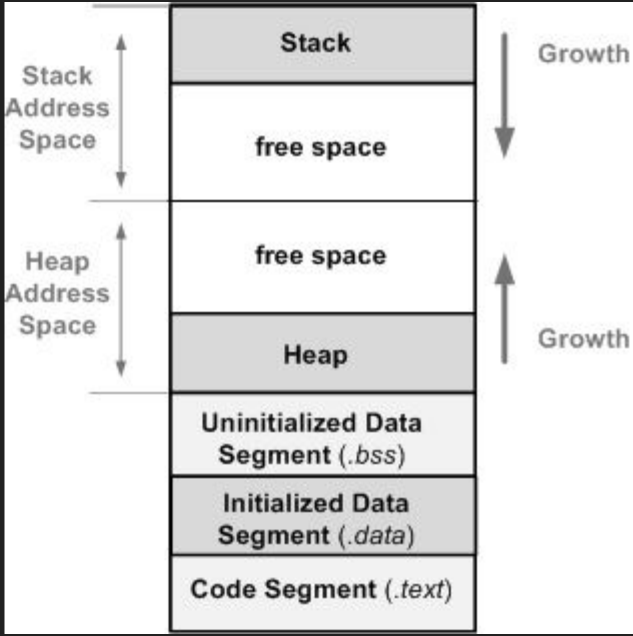
This array has some different segments:

Stack (local vars)

Heap (malloc or other dynamic allocated)

Data (global variables)

Text (code)



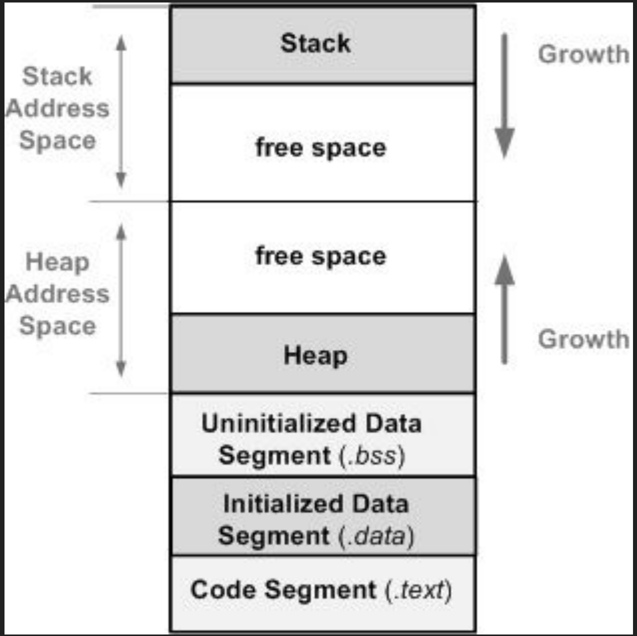
Stack segment

A stack frame describes a small part of the process array which is used to store information about a function call, like local variables.

A stack frame is always located inside the stack segment.

Every time you call a function, the computer creates a new stack frame to keep track of its variables.

Whenever you return from a function, its stack frame is DESTROYED :(

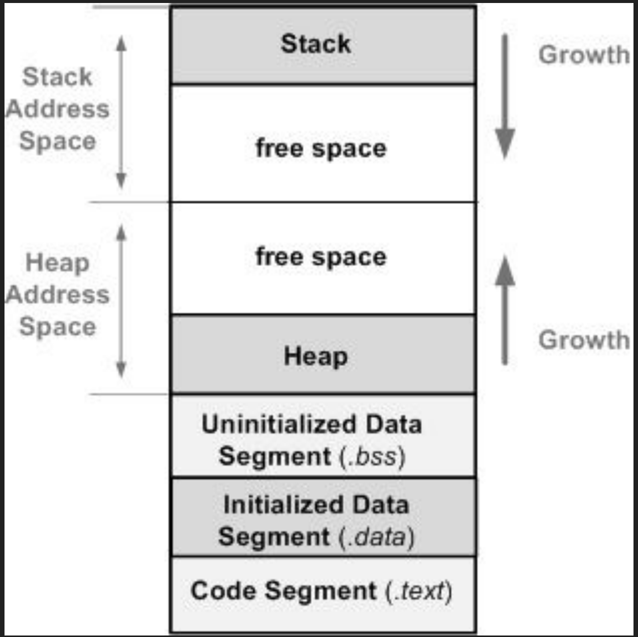


Heap segment

The heap is a segment used to contain memory which is allocated by the programmer.

If you call malloc, the computer will set aside some heap memory and give you a pointer to it.

Heap memory is only destroyed at the programmer's request (like with free)

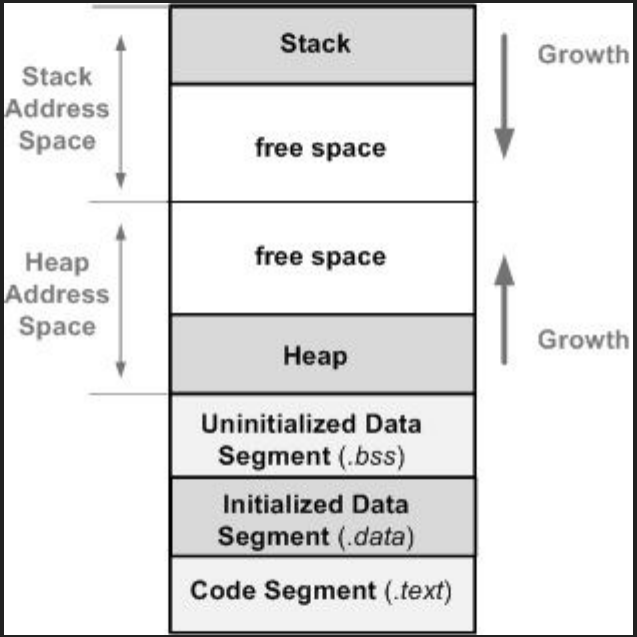


Data segment

The data segment mainly stores global variables.

Kinda like the heap, but the data segment is static instead of dynamic

(so the data segment won't let you add more stuff at runtime, just whatever globals you defined originally)

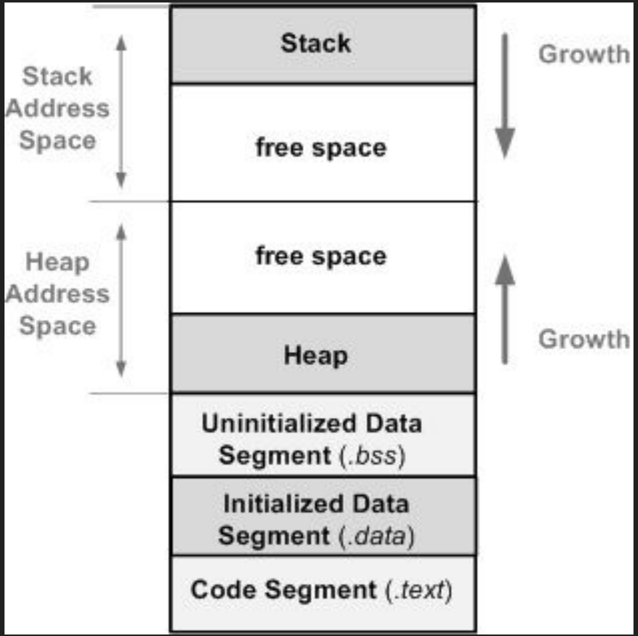


Text segment

Stores a binary representation of your process's code.

This binary representation is produced by a compiler.

There is an "entry" address inside the text segment where the CPU will start reading the binary code from your program, and running it using the CPU hardware.



What is a pointer / memory address.

A pointer is the exact same thing as a memory address.

A memory address is just the value of an index into the RAM array.

For example, the address NULL (0x0) is the index 0.

```c
#include <stdio.h>

int main(void) {
    int pointer = 1337;
    // Tell C that it's actually a pointer (to the address 1337) using type casting
    int *trust_me = (int*) pointer;
    // Dereference it and pray
    printf("%d\n", *trust_me);
    // Segmentation fault :(
```

Tute Questions

2. The following snippet of C code declares two variables, of the variables `s1` and `s2` with some subtle differences between them.

```c
#include <stdio.h>

char *s1 = "abc";

int main(void) {
  char *s2 = "def";
  // ...
}
```

- What does the memory layout of a typical program look like?
- What is a *global variable*?
- How do they differ from local variables? Where are they each located in memory?
- What is a string literal? Where are they located in memory?

Tute Questions

3. What is wrong with the following code?

```
#include <stdio.h>

int *get_num_ptr(void);

int main(void) {
    int *num = get_num_ptr();
    printf("%d\n", *num);
}

int *get_num_ptr(void) {
    int x = 42;
    return &x;
}
```

Assuming we still want `get_num_ptr` to return a pointer, how can we fix this code?

How does fixing this code affect each variable's location in memory?

Tute Questions

4. Consider the following C program:

```c
#include <stdio.h>

int main(void) {
    char str[10];
    str[0] = 'H';
    str[1] = 'i';
    printf("%s", str);
    return 0;
}
```

What will happen when the above program is compiled and executed?

In particular, what does this look like **in memory**?

How could we fix this program?

Compiling

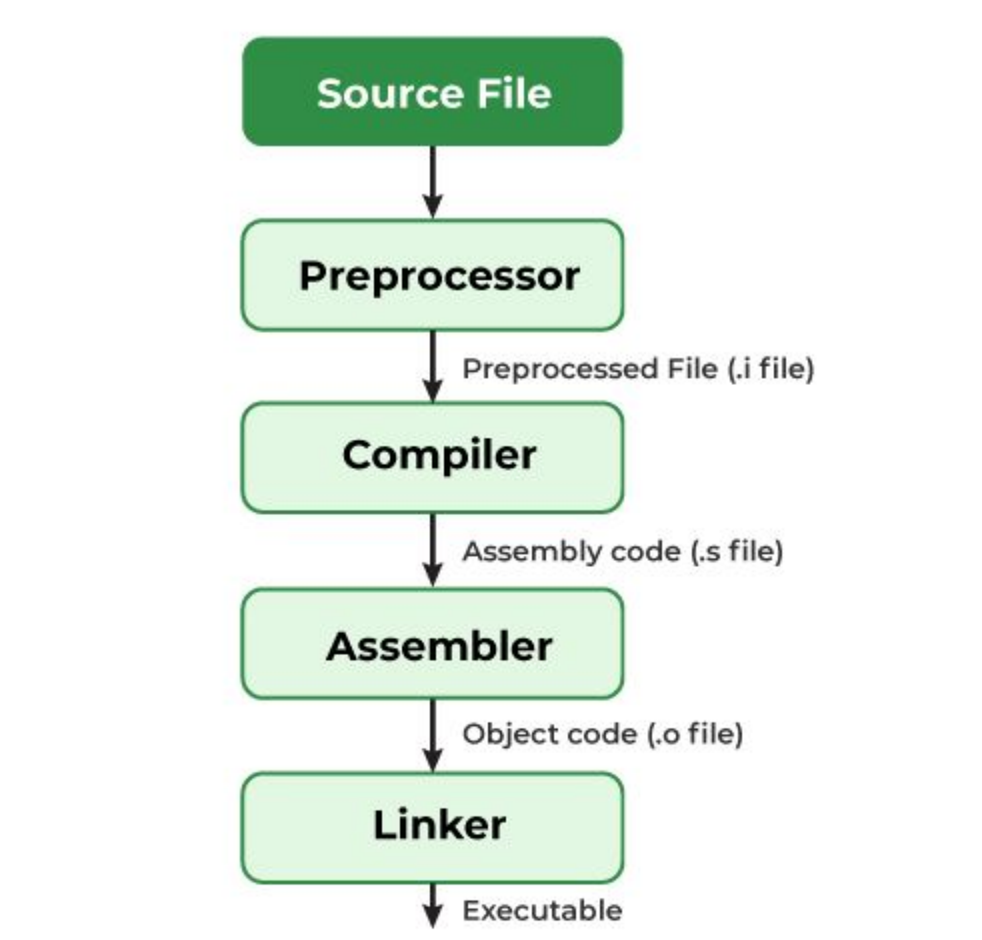
Steps of a compiler

Preprocessing

Compiling (no way)

Assembling

Linking (dark arts)



Preprocessing

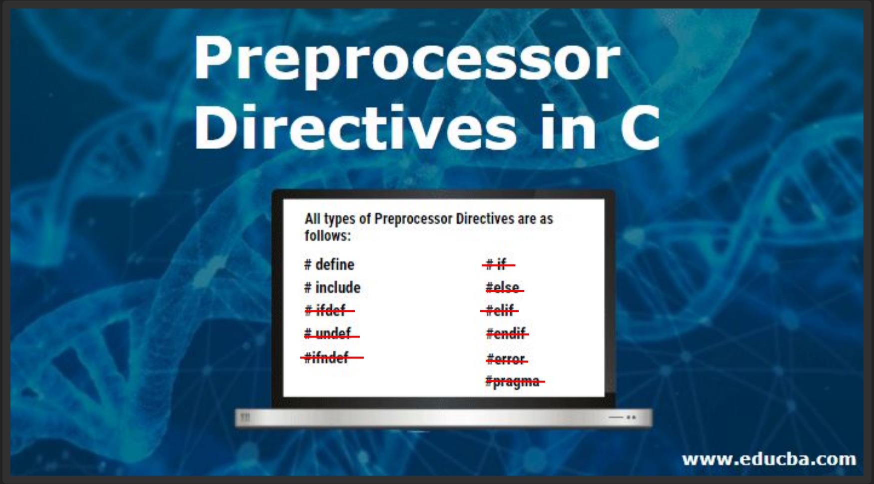
Take care of #defines.

Take of #includes (literally just copy paste the entire included file into this file)

Remove comments (goodbye style)

Other directives which start with '#'

(you don't need to know weird C preprocessor directives for this course)



Compiling

Turns C code into assembly code which matches your CPU

(for example, produce x86 assembly code for a windows PC, ARM assembly code for a macbook, or MIPS assembly code if you are doing COMP1521)

("Assembly" is just a weird programming language that maps nicely to the hardware inside a CPU.)

C code:

```
f = (g + h) - (i + j);
```

Compiled MIPS code:

```
add t0, g, h   # temp t0 = g + h
add t1, i, j   # temp t1 = i + j
sub f, t0, t1  # f = t0 - t1
```

Assembling

Turns the compiled assembly code into the binary code we mentioned earlier (for .text section).

It doesn’t work yet though, because we don’t have a function definition for library functions like “printf”.

We need to include the definition of library functions like “printf” in our binary file somehow.

Linking

Link definitions of library functions to their calls in the binary code.

This is super complicated so we won't spend time going through exactly how it works.

To generalise, we either link the functions to definitions which exist elsewhere in RAM, or we copy-paste in the definitions themselves into this executable.

Tute Questions

10. For each of the following commands, describe what kind of output would be produced:

    a. `clang -E x.c`
    b. `clang -S x.c`
    c. `clang -c x.c`
    d. `clang x.c`

    In particular, how do these commands relate to what we will be studying in COMP1521?

    You can use the following simple C code as an example:

```c
#include <stdio.h>
#define N 10

int main(void) {
    char str[N] = { 'H', 'i', '\0' };
    printf("%s\n", str);
    return 0;
}
```

C revision/new stuff

For loops

```c
int main(void) {
    /* initialize */
    int x = 0;
    /* loop condition */
    while (x < 100) {
        /* body */
        printf("hi\n");
        /* loop step */
        x++;
    }

    for (int x = 0; x < 100; x++) {
        printf("hi\n");
    }

}
```

Tute Questions

6. Consider the following while loop:

```c
#include <stdio.h>

int main(void) {
  int i = 0;
  while (i < 10) {
    printf("%d\n", i);
    i++;
  }
  return 0;
}
```

How could we rewrite the above program using a for loop? What subtle difference would there be between the two programs?

Argc and Argv

Argc = number of command line arguments

Argv = array containing command line arguments (array of strings)

Note that both include the name of the program as the first argument!!!!!!!!!!

```c
#include <stdio.h>

int main(int argc, char *argv[]) {
  return 0;
}
```

Tute Questions

8. In the following program, what are argc and argv? The following program prints number of command-line arguments and each command-line argument on a separate line.

```c
#include <stdio.h>

int main(int argc, char *argv[]) {
    printf("argc=%d\n", argc);
    for (int i = 0; i < argc; i++) {
        printf("argv[%d]=%s\n", i, argv[i]);
    }
    return 0;
}
```

What will be the output of the following commands?

```
$ dcc -o print_arguments print_arguments.c
$ ./print_arguments I love MIPS
```

Tute Questions

9. The following program sums up command-line arguments.
   Why do we need the function atoi in the following program?

   The program assumes that command-line arguments are integers. What if they are not integer values?

```
#include <stdio.h>
#include <stdlib.h>

int main(int argc, char *argv[]) {
    int sum = 0;
    for (int i = 0; i < argc; i++) {
        sum += atoi(argv[i]);
    }
    printf("sum of command-line arguments = %d\n", sum);
    return 0;
}
```

Recursion

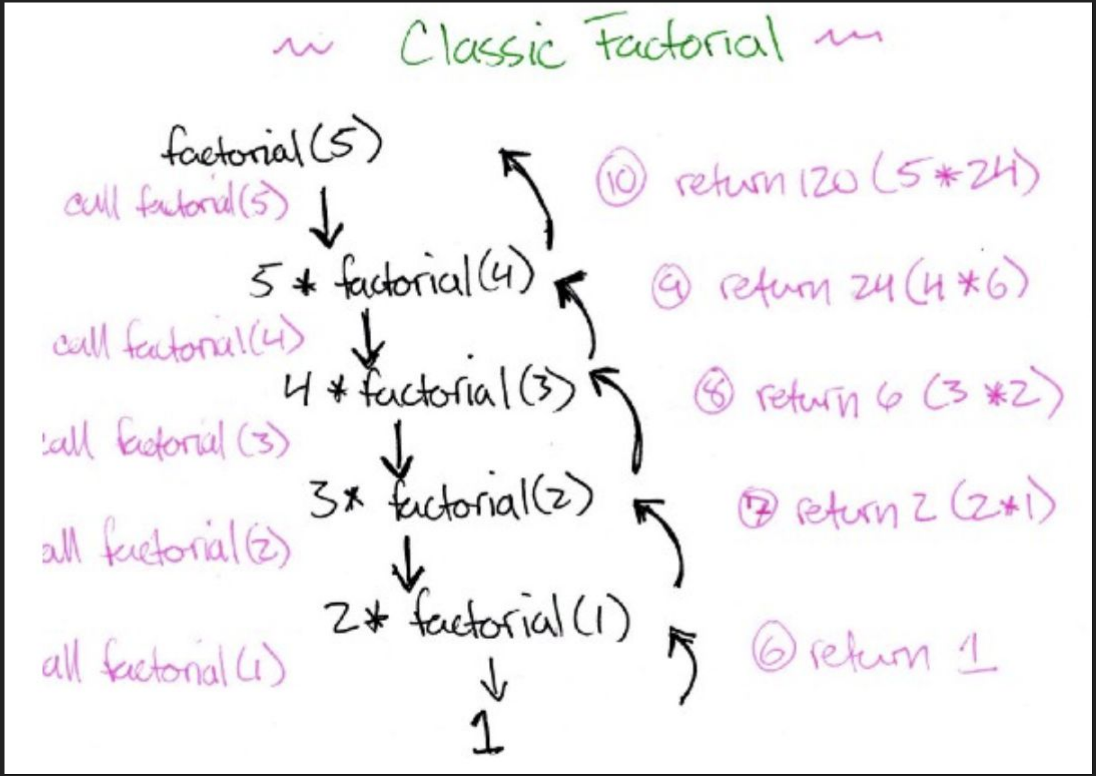
Recursion allows programmers to write more simple or concise code to solve certain types of problems.

Sometimes this code can be really confusing to read, compared to an iterative approach (like using a loop). Regardless, in a lot of cases recursion can massively simplify the logic for a program.

Recursion

Intuition:

We “walk” down to the smallest case, and figure out the solution on the way up.



Recursion

How to write a recursive solution.

1: Find one or more “base cases”. This is where your program will stop recursion and typically return a very simple answer.

2: Find a way to solve the other cases. This is typically done by recursively calling your function on a smaller input (or an input closer to the base case).

Eventually, the input will become small enough where you reach the base case.