

HARDWARE (COMPILERS, OS, RUNTIME) AND C, C++, TO CUDA

ERNEST YEUNG ERNESTYALUMNI@GMAIL.COM

CONTENTS

1.	Introduction; why I'm writing this	3
2.	80x86 Assembly; 80x86 Architecture	3
2.1.	Basic 80x86 Architecture	3
2.2.	Registers (for 80x86)	4
3.	Compilers; Compiler Operation and Code Generation	4
4.	gdp; good debugging processes (in C/C++/CUDA)	4
5.	Pointers in C; Pointers in C categorified (interpreted in Category Theory) and its relation to actual, physical, computer memory and (memory) addresses ((address) bus; pointers, structs, arrays in C	4
5.1.	Structs in C	6
Part 1.	C, Stack, Heap Memory Management in C	6
6.	C, Stack and Heap Memory Management, Heap and Stack Memory Allocation	6
7.	Data segments; Towards Segmentation Fault	6
7.1.	Stack	7
7.2.	Stack overflow	7
7.3.	Heap	8
7.4.	More Segmentation Faults	8
Part 2.	C++	8
8.	static	8
9.	Free Store	8
10.	Copy vs. Move	8
10.1.	Copy constructor	8
10.2.	Move Constructor	9
11.	vtable; virtual table	9
11.1.	How are virtual functions and vtable implemented?	10
11.2.	pImpl, shallow copy, deep copy	10
Part 3.	Integers, numeric encodings, number representation, binary representation, hexadecimal representation	11
12.	Introduction: $\text{int} \neq \mathbb{Z}$, $\text{float} \neq \mathbb{R}$	11
13.	Bits and Bytes	11
13.1.	Machine Words	11
13.2.	Word-oriented Memory Organization	11
13.3.	Byte-ordering: Big-Endian, Little-Endian	11
13.4.	Representing Strings	12
13.5.	Machine-Level Code Representation	12
13.6.	Boolean Algebra	12

Date: 1 Nov 2017.

Key words and phrases. C, C++, CUDA, CUDA C/C++, Compilers, OS, runtime, classes, Object Oriented Programming, Segmentation Fault, memory, memory addresses.

13.7.	Representing and Manipulating Sets; bit vectors as sets	12
14.	Unsigned integers	12
15.	Two’s complement	13
15.1.	Numeric Ranges	13
15.2.	Negating with complement	13
15.3.	Power of 2 Multiply with shift (left shift bitwise operation)	13
15.4.	Unsigned power-of-2 Divide, with shift (right shift bitwise operator)	13
16.	Endianness	14
16.1.	Big-endian	14
16.2.	Little-endian	14

Part 4.	Floating Point Numbers	14
17.	Floating point, IEEE Floating	14
17.1.	Fractional Binary Numbers	14

Part 5.	Complexity, Data Structures, Algorithms	16
18.	Complexity, Big- O	16
18.1.	Multi-Part Algorithms: Add vs. Multiply, e.g. $O(A + B)$ vs. $O(A * B)$	16
18.2.	Amoritized Time (e.g. dynamically resizing array ”ArrayList” or maybe <code>std::vector</code>)	16
18.3.	$\log N$ runtimes ($O(\log N)$)	16
18.4.	Recursive run times $O(2^N)$	16
19.	Data Structures	17
19.1.	Lists (arrays), Python Lists	17
19.2.	Linked Lists, $O(1)$ insertion	17
19.3.	Stack, $O(1)$ pop, $O(1)$ push	18
19.4.	Queue, $O(1)$ insert, $O(1)$ delete, $O(N)$ search, $O(N)$ space complexities	18
19.5.	STL Containers	18
19.6.	vector	19
19.7.	Trees	19
19.8.	Graphs	19
20.	Search	20
20.1.	Binary Search $O(\log N)$	20
20.2.	Breadth-first search (BFS)	20
20.3.	Depth-first search	20
20.4.	Binary Search Tree (BST), $O(\log N)$ height of the tree, run-time complexity, worst case $O(N)$	20
20.5.	Heaps	20
20.6.	Self-balancing tree	21
20.7.	Red-Black Tree; search $O(\log N)$, space $O(N)$, insert $O(\log N)$, delete $O(\log N)$	21
21.	Recursion	21
22.	Sorting	21
22.1.	Bubble Sort, $O(N^2)$ in time, $O(1)$ in space	21
22.2.	Merge Sort, $O(N \log N)$ in time, $O(N)$ in space	21
22.3.	Quick Sort, $O(N \log N)$ average time complexity, $O(N^2)$ worst time complexity, $O(1)$ in space (if inplace)	21
22.4.	Time Complexities of all Sorting Algorithms	22
22.5.	Hashing, $O(1)$ lookup, or $O(n)$ bucket lookup	22
23.	References for Data Structures and Algorithms	22

Part 6.	Linux System Programming	22
24.	File Descriptors	22

	HARDWARE (COMPILERS, OS, RUNTIME) AND C, C++, TO CUDA	3
24.1.	Creation and Release of Descriptors	22
25.	Unix execution model	23
25.1.	Processes and Programs	23
25.2.	File entry	23
25.3.	Readiness of fds	23
25.4.	Level Triggered	23
25.5.	Edge Triggered	23
26.	Multiplexing I/O on fds	24
26.1.	Multiplexing I/O with Non-blocking I/O	24
26.2.	Multiplexing I/O via signal driven I/O	24
26.3.	Multiplexing I/O via polling I/O	24
27.	epoll Event poll	24
27.1.	epoll Syntax	24
28.	References and Resources for Linux System Programming	24
Part 7.	C++ Template Metaprogramming	24
Part 8.	Functional Programming and Category Theory	24
29.	Actors, Concurrent Systems	24
Part 9.	Technical Interviews	24
Part 10.	Embedded Systems	25
	References	26

ABSTRACT. I review what, how, and why Segmentation Fault is and occurs in the specific context of C and hopefully to CUDA, virtual memory addresses in C++, and to CUDA.

1. INTRODUCTION; WHY I’M WRITING THIS

I didn’t realize the importance of having a profound understanding of C/C++ and its relation to hardware, in particular memory (addresses), until I learned more specifically about the work done at NVIDIA from its news and job postings. Coming from a theoretical and mathematical physics background, I wanted to get myself up to speed as fast as possible, provide good textbook and online references, and directly relate these topics, which seem to be classic topics in computer science (at hardware level) and electrical engineering, to CUDA, to specifically GPU parallel programming with direct CUDA examples.

Note that there is a version of this LaTeX/pdf file in the usual ”letter” format that is exactly the same as the content here, `HrdwCCppCUDA.pdf`. Meyer (1997) [1]

”When programmers first began giving up assembly language in favor of using HLLs, they generally understood the low-level ramifications of the HLL statements they were using and could choose their HLL statements appropriately.Unfortunately, the generation of computer programmers that followed them did not have the benefit of mastering assembly language. As such, they were not in a position to wisely choose statements and data structures that HLLs could efficiently translate into machine code.”

pp. 2, Ch. 1 of Hyde (2006) [3], where HLL stands for high-level language. I was part of that generation and knew of nothing of assembly.

2. 80x86 ASSEMBLY; 80x86 ARCHITECTURE

cf. Ch. 3 of Hyde (2006) [3]

Main difference between complex instruction set computer (CISC) architectures such as 80x86 and reduced instruction set computer (RISC) architectures like PowerPC is the way they use memory. RISC provde relatively clumsy memory access, and applications avoid accessing memory. 80x86 access memory in many different ways and applications take advantage of these facilities.

2.1. **Basic 80x86 Architecture.** Intel CPU generally classified as a *Von Neumann machine*, containing 3 main building blocks:

- (1) CPU
- (2) memory
- (3) input/output (I/O) devices

These 3 components are connected together using the *system bus*. System bus consists of

- address bus
- data bus
- control bus

CPU communicates with memory and I/O devices by placing a numeric value on the address bus to select 1 of the memory locations or I/O device port locations, each of which has a unique binary numeric address. Then the CPU, I/O, and memory devices pass data among themselves by placing data on the data bus. Control bus contains signals that determine the direction of data transfer (to or from memory, and to or from an I/O device).

So

$$\begin{array}{c} \mathbf{Memory} \\ \text{Obj}(\mathbf{Memory}) = \text{memory locations} \\ \mathbf{I/O} \\ \text{Obj}(\mathbf{I/O}) = \text{I/O device port locations} \end{array}$$

and

$$\begin{array}{c} \mathbf{address\ bus} \\ \text{Obj}(\mathbf{address\ bus}) = \text{unique binary numeric addresses} \end{array}$$

Likewise,

$$\begin{array}{c} \mathbf{data\ bus} \\ \text{Obj}(\mathbf{data\ bus}) = \text{data} \end{array}$$

And so

$$\begin{array}{l} \text{CPU} : \mathbf{Memory} \times \mathbf{address\ bus} \rightarrow \text{Hom}(\mathbf{Memory}, \mathbf{address\ bus}) \\ \text{CPU} : (\text{memory location, unique binary numeric address}) \mapsto \& \end{array}$$

in the language of category theory.

2.2. **Registers (for 80x86).** cf. 3.3 Basic 80x86 Architecture, from pp. 23 of Hyde (2006) [3].

Example: to add 2 variables, x, y and store $x + y$ into z , you must load 1 of the variables into a register, add the 2nd. operand to the register, and then store register’s value into destination variable. ”Registers are middlemen in almost every calculation.” Hyde (2006) [3].

There are 4 categories of 80x86 CPU registers:

- general-purpose
- special-purpose application-accessible
- segment
- special-purpose kernel-mode

Segment registers not used very much in modern 32-bit operating systems (e.g. Windows, Linux; what about 64-bit?) and special-purpose kernel-mode registers are intended for writing operating systems, debuggers, and other system-level tools.

2.2.1. *80x86 General-Purpose Registers.* 80x86 CPUs provide several general-purpose registers, including 8 32-bit registers having the following names: (8 bits in 1 byte, 32-bit is 4 bytes)

$$EAX, EBX, ECX, EDX, ESI, EDI, EBP, ESP$$

where E is *extended*; 32-bit registers distinguished from 8 16-bit registers with following names

$$AX, BX, CX, DX, SI, DI, BP, SP$$

and finally

$$AL, AH, BL, BH, CL, CH, DL, DH$$

Hyde says that it’s important to note about the general-purpose registers is they’re not independent. 80x86 doesn’t provide 24 separate registers, but overlaps the 32-bit registers with 16-bit registers, and overlaps 16-bit registers with 8-bit registers. e.g.

$$\begin{array}{l} \&AL \neq \&AH, \text{ but} \\ \&AL, \&AH \in \&AX \text{ and } \&AX \in \&EAX, \text{ so that} \\ \&AL, \&AH, \&AX \in \&EAX \end{array}$$

3. COMPILERS; COMPILER OPERATION AND CODE GENERATION

cf. Ch. 5 Compiler Operation and Code Generation, from pp. 62 and pp. 72- of Hyde (2006) [3]

4. GDP; GOOD DEBUGGING PROCESSES (IN C/C++/CUDA)

Learn C the hard way Lecture 4, Using Debugger (GDB)

5. POINTERS IN C; POINTERS IN C CATEGORIFIED (INTERPRETED IN CATEGORY THEORY) AND ITS RELATION TO ACTUAL, PHYSICAL, COMPUTER MEMORY AND (MEMORY) ADDRESSES ((ADDRESS) BUS; POINTERS, STRUCTS, ARRAYS IN C

From Shaw (2015) [4], Exercise 15,

e.g. `ages[i]`, You’re indexing into array `ages`, and you’re using the number that’s held in i to do it:

$$\begin{array}{ll} a : \mathbb{Z} \rightarrow \text{Type} \in \mathbf{Type} & a : \mathbb{Z} \rightarrow \mathbb{R} \text{ or } \mathbb{Z} \\ a : i \mapsto a[i] & \text{e.g. } a : i \mapsto a[i] \end{array}$$

Index $i \in \mathbb{Z}$ is a location *inside* `ages` or a , which can also be called *address*. Thus, $a[i]$.

Indeed, from [cppreference for Member access operators](#),

Built-in *subscript* operator provides access to an object pointed-to by pointer or array operand. And so `E1[E2]` is exactly identical to `*(E1+E2)`.

To C, e.g. `ages`, or a , is a location in computer’s memory where, e.g., all these integers (of `ages`) start, i.e. where a starts.

$$\begin{array}{ll} \mathbf{Memory}, & \text{Obj}(\mathbf{Memory}) \ni \text{memory location. Also, to specify CPU,} \\ \mathbf{Memory}_{CPU}, & \text{Obj}(\mathbf{Memory}_{CPU}) \ni \text{computer memory location} \end{array}$$

It’s *also* an address, and C compiler will replace e.g. `ages` or array a , anywhere you type it, with address of very 1st integer (or 1st element) in, e.g. `ages`, or array a .

$$\begin{array}{l} \mathbf{Arrays} \stackrel{\cong}{\longleftrightarrow} \mathbf{address} \\ \text{Obj}(\mathbf{Arrays}) \stackrel{\cong}{\longleftrightarrow} \text{Obj}(\mathbf{address}) \\ a \stackrel{\cong}{\longleftrightarrow} 0\text{x}17 \end{array}$$

”But here’s the trick”: e.g. ”`ages` is an address inside the *entire computer*.” (Shaw (2015) [4]).

It’s not like i that’s just an address inside `ages`. `ages` array name is actually an address in the computer.

”This leads to a certain realization: C thinks your whole computer is 1 massive array of bytes.”

”What C does is layer on top of this massive array of bytes the concept of types and sizes of those types.” (Shaw (2015) [4]).

Let

$$\begin{array}{ll} \mathbf{Type} & \text{Obj}(\mathbf{Type}) \ni \text{int, char, float} \\ \mathbf{Memory}_{CPU} := 1 \text{ massive array of bytes} & \text{Obj}(\mathbf{Type}) \xrightarrow{\text{sizeof}} \mathbb{Z}^+ \\ \text{Obj}(\mathbf{Memory}_{CPU}) & T \xrightarrow{\text{sizeof}} \text{sizeof}(T) \\ & \text{float} \xrightarrow{\text{sizeof}} \text{sizeof(float)} \end{array}$$

How C is doing the following with your arrays:

- *Create* a block of memory inside your computer:

$$\text{Obj}(\mathbf{Memory}_{CPU}) \supset \text{Memory block}$$

Let $\text{Obj}(\mathbf{Memory}_{CPU})$ be an ordered set. Clearly, then memory can be indexed. Let $b \in \mathbb{Z}^+$ be this index. Then $\text{Memory block}(0) = \text{Obj}(\mathbf{Memory}_{CPU})(b)$.

- *Pointing* the name `ages`, or a , to beginning of that (memory) block.
Entertain, possibly, a category of pointers, $\mathbf{Pointers} \equiv \mathbf{ptrs}$.

$$\begin{array}{l} \mathbf{ptrs} \\ \text{Obj}(\mathbf{ptrs}) \ni a, \text{ e.g. } \mathbf{ages} \end{array}$$

$$a \mapsto \text{Memory block}(0)$$

$$\text{Obj}(\mathbf{ptrs}) \rightarrow \text{Obj}(\mathbf{Memory}_{CPU})$$

- *indexing* into the block, by taking the base address of `ages`, or *a*

$$a \xrightarrow{\cong} \text{base address } 0x17$$

$$\text{Obj}((T)\mathbf{array}) \xrightarrow{\cong} \text{Obj}(\mathbf{addresses})$$

$$a[i] \equiv a + i \xrightarrow{\cong} \text{base address} + i * \mathbf{sizeof}(T) \mapsto a[i] \in T \text{ where } T, \text{ e.g. } T = \mathbb{Z}, \mathbb{R}$$

$$\text{Obj}((T)\mathbf{array}) \xrightarrow{\cong} \text{Obj}(\mathbf{addresses}) \rightarrow T$$

”A pointer is simply an address pointing somewhere inside computer’s memory with a type specifier.” Shaw (2015) [4]
C knows where pointers are pointing, data type they point at, size of those types, and how to get the data for you.

5.0.1. *Practical Pointer Usage.*

- Ask OS for memory block (chunk of memory) and use pointer to work with it. This includes strings and **structs**.
- Pass by reference - pass large memory blocks (like large structs) to functions with a pointer, so you don’t have to pass the entire thing to the function.
- Take the address of a function, for dynamic callback. (function of functions)

”You should go with arrays whenever you can, and then only use pointers as performance optimization if you absolutely have to.” Shaw (2015) [4]

5.0.2. *Pointers are not Arrays.* No matter what, pointers and arrays are not the same thing, even though C lets you work with them in many of the same ways.

From Eli Bendersky’s website, [Are pointers and arrays equivalent in C?](#)

He also emphasizes that

5.0.3. *Variable names in C are just labels.* ”A variable in C is just a convenient, alphanumeric pseudonym of a memory location.” (Bendersky, [5]). What a compiler does is, create label in some memory location, and then access this label instead of always hardcoding the memory value.

”Well, actually the address is not hard-coded in an absolute way because of loading and relocation issues, but for the sake of this discussion we don’t have to get into these details.” (Bendersky, [5]) (EY : 20171109 so it’s on the address bus?)

Compiler assigns label *at compile time*. Thus, the great difference between arrays and pointers in C.

5.0.4. *Arrays passed to functions are converted to pointers.* cf. Bendersky, [5].

Arrays passed into functions are always converted into pointers. The argument declaration `char arr_place[]` in

```
void foo(char arr_arg [], char* ptr_arg)
{
    char a = arr_arg[7];
    char b = ptr_arg[7];
}
```

is just syntactic sugar to stand for `char* arr_place`.

From Kernighan and Ritchie (1988) [6], pp. 99 of Sec. 5.3, Ch. 5 Pointers and Arrays,

When an array name is passed to a function, what is passed is the location of the initial element. Within the called function, this argument is a local variable, and so an array name parameter is a pointer, that is, a variable containing an address.

Why?

The C compiler has no choice here, since,

array name is a label the C compiler replaces *at compile time* with the address it represents (which for arrays is the address of the 1st element of the array).

But function isn’t called at compile time; it’s called *at run time*.

At run time, (where) something should be placed on the stack to be considered as an argument.

Compiler cannot treat array references inside a function as labels and replace them with addresses, because it has no idea what actual array will be passed in at run time.

EY : 20171109 It can’t anticipate the exact arguments that’ll it be given *at run-time*; at the very least, my guess is, it’s given instructions.

Bendersky [5] concludes by saying the difference between arrays and pointers does affect you. One way is how arrays can’t be manipulated the way pointers can. Pointer arithmetic isn’t allowed for arrays and assignment to an array of a pointer isn’t allowed. cf. van der Linden (1994) [7]. Ch. 4, 9, 10.

Bendersky [5] has this one difference example, ”actually a common C gotcha”:

”Suppose one file contains a global array:”

```
char my_Arr[256]
```

Programmer wants to use it in another file, *mistakingly* declares as

```
extern char* my_arr;
```

When he tries to access some element of the array using this pointer, he’ll most likely get a segmentation fault or a fatal exception (nomenclature is OS-dependent).

To understand why, Bendersky [5] gave this hint: look at the assembly listing

```
char a = array_place[7];
```

```
0041137E  mov  al,byte ptr [_array_place+7 (417007h)]
00411383  mov  byte ptr [a],al
```

```
char b = ptr_place[7];
```

```
00411386  mov  eax,dword ptr [_ptr_place (417064h)]
0041138B  mov  cl,byte ptr [eax+7]
0041138E  mov  byte ptr [b],cl
```

or my own, generated from `gdb` on Fedora 25 Workstation Linux:

```
0x00000000004004b1 <main+11>:  movzbl 0x200b8f(%rip),%eax          # 0x601047 <array_place+7>
0x00000000004004b8 <main+18>:  mov     %al,-0x1(%rbp)
```

```
0x00000000004004bb <main+21>:  mov     0x200be6(%rip),%rax          # 0x6010a8 <ptr_place>
0x00000000004004c2 <main+28>:  movzbl 0x7(%rax),%eax
0x00000000004004c6 <main+32>:  mov     %al,-0x2(%rbp)
```

”How will the element be accessed via the pointer? What’s going to happen if it’s not actually a pointer but an array?” Bendersky [5]

EY : 20171106. Instruction-level, the pointer has to

- `mov 0x200be6(%rip),%rax` - 1st., copy value of the pointer (which holds an address), into `%rax` register.
- `movzbl 0x7(%rax),%eax` - off that address in register ‘
- `mov %al,-0x2(%rbp)` - mov contents `-0x2(%rbp)` into register `%al`

If it’s not actually a pointer, but an array, the value is copied into the `%rax` register is an actual `char` (or `float`, some type). *Not* an address that the registers may have been expecting!

5.1. **Structs in C.** From Shaw (2015) [4], Exercise 16, `struct` in C is a collection of other data types (variables) that are stored in 1 block of memory. You can access each variable independently by name.

- The `struct` you make, i.e.g `struct Person` is now a *compound data type*, meaning you can refer to `struct Person` using the same kinds of expressions you would for other (data) types.
- This lets you pass the whole `struct` to other functions
- You can access individual members of `struct` by their names using `x->y` if dealing with a ptr.

If you didn’t have `struct`, you’d have to figure out the size, packing, and location of memory of the contents. In C, you’d let it handle the memory structure and structuring of these compound data types, `structs`. (Shaw (2015) [4])

Part 1. C, Stack, Heap Memory Management in C

6. C, STACK AND HEAP MEMORY MANAGEMENT, HEAP AND STACK MEMORY ALLOCATION

cf. Ex. 17 of Shaw (2015) [4]

Consider chunk of RAM called stack, another chunk of RAM called heap. Difference between heap and stack depends on where you get the storage.

Heap is all the remaining memory on computer. Access it with `malloc` to get more.

Each time you call `malloc`, the OS uses internal functions (EY : 20171110 address bus or overall system bus?) to register that piece of memory to you, then returns ptr to it.

When done, use `free` to return it to OS so OS can use it for other programs. Failing to do so will cause program to *leak* memory. (EY: 20171110, meaning this memory is unavailable to the OS?)

Stack, on a special region of memory, stores temporary variables, which each function creates as locals to that function. How stack works is that each argument to function is *pushed* onto stack and then used inside the function. Stack is really a stack data structure, LIFO (last in, first out). This also happens with all local variables in `main`, such as `char action`, `int id`. The advantage of using stack is when function exits, *C compiler* pops these variables off of stack to clean up.

Shaw’s mantra: If you didn’t get it from `malloc`, or a function that got it from `malloc`, then it’s on the stack.

3 primary problems with stacks and heaps:

- If you get a memory block from `malloc`, and have that ptr on the stack, then when function exits, ptr will get popped off and lost.
- If you put too much data on the stack (like large structs and arrays), then you can cause a *stack overflow* and program will abort. Use the heap with `malloc`.
- If you take a ptr, to something on stack, and then pass or return it from your function, then the function receiving it will *segmentation fault*, because actual data will get popped off and disappear. You’ll be pointing at dead space.

cf. Ex. 17 of Shaw (2015) [4]

7. DATA SEGMENTS; TOWARDS SEGMENTATION FAULT

cf. Ferres (2010) [8]

When program is loaded into memory, it’s organized into 3 *segments* (3 areas of memory): let executable program generated by a compiler (e.g. `gcc`) be organized in memory over a range of addresses (EY : 20171111 assigned to physical RAM memory by address bus?), ordered, from low address to high address.

- *text* segment or code segment - where compiled code of program resides (from lowest address); code segment contains code executable or, i.e. code binary.
As a memory region, text segment may be placed below heap, or stack, in order to prevent heaps and stack overflows from overwriting it.
Usually, text segment is sharable so only a single copy needs to be in memory for frequently executed programs, such as text editors, C compiler, shells, etc. Also, text segment is often read-only, to prevent program from accidentally modifying its instructions. cf. [Memory Layout of C Programs; GeeksforGeeks](#)
- Data segment: data segment subdivided into 2 parts:

- initialized data segment - all global, static, constant data stored in data segment. Ferres (2010) [8]. Data segment is a portion of virtual address of a program.
Note that, data segment not read-only, since values of variables can be altered at run time.
This segment can also be further classified into initialized read-only area and initialized read-write area.
e.g. `char s[] = "hello world"` and `int debut = 1` *outside the main (i.e. global)* stored in initialized read-write area.
`const char *string = "hello world"` in global C statement makes string literal ”hello world” stored in initialized read-only area. Character pointer variable `string` in initialized read-write area. cf. [Memory Layout of C Programs](#)
- uninitialized data stored in BSS. Data in this segment is initialized by kernel (OS?) to arithmetic 0 before program starts executing.
Uninitialized data starts at end of data segment (”largest” address for data segment) and contains all global and static variables initialized to 0 or don’t have explicit initialization in source code.
e.g. `static int i;` in BSS segment.
e.g. `int j;` global variable in BSS segment.
cf. [Memory Layout of C Programs](#)
- Heap - ”grows upward” (in (larger) address value, begins at end of BSS segment), allocated with `calloc`, `malloc`, ”dynamic memory allocation”.
Heap area shared by all shared libraries and dynamically loaded modules in a process.
Heap grows when memory allocator invokes `brk()` or `sbrk()` system call, mapping more pages of physical memory into process’s virtual address space.
- Stack - store local variables, used for passing arguments to functions along with return address of the instruction which is to be executed after function call is over.
When a new stack frame needs to be added (resulting from a *newly called function*), stack ”grows downward.” (Ferres (2010) [8])
Stack grows automatically when accessed, up to size set by kernel (OS?) (which can be adjusted with `setrlimit(RLIMIT_STACK,...)`).

7.0.1. *Mathematical description of Program Memory (data segments), with **Memory, Addresses**.* Let **Address**, with $\text{Obj}(\mathbf{Address}) \cong \mathbb{Z}^+$ be an ordered set.
Memory block $\subset \text{Obj}(\mathbf{Memory})$, s.t.

$$\text{Memory block} \stackrel{\cong}{\mapsto} \{ \text{low address} , \text{low address} + \mathbf{sizeof}(T), \dots \text{high address} \} \equiv \text{addresses}_{\text{Memory block}} \subset \mathbb{Z}^+$$

where $T \in \text{Obj}(\mathbf{Types})$ and \cong assigned by address bus, or the virtual memory table, and $\text{addresses}_{\text{Memory block}} \subset \text{Obj}(\mathbf{Addresses})$.

Now,
text segment, (initialized) data segment, (uninitialized) data segment, heap, stack, command-line arguments and environmental variables $\subset \text{addresses}_{\text{Memory block}}$, that these so-called data segments are discrete subsets of the set of all addresses assigned for the memory block assigned for the program.

Now, $\forall i \in \text{text segment} , \forall j \in (\text{initialized}) \text{ data segment} , i < j$ and $\forall j \in (\text{initialized}) \text{ data segment} , \forall k \in (\text{uninitiaized}) \text{ data segment} , j < k$, and so on. Let’s describe this with the following notation:

$$(1) \qquad \text{text segment} < (\text{initialized}) \text{ data segment} < (\text{uninitialized}) \text{ data segment} < \\ < \text{heap} < \text{stack} < \text{command-line arguments and environmental variables}$$

Consider stack of variable length $n_{\text{stack}} \in \mathbb{Z}^+$. Index the stack by $i_{\text{stack}} = 0, 1, \dots n_{\text{stack}} - 1$. ”Top of the stack” is towards ”decreasing” or ”low” (memory) address, so that the relation between ”top of stack” to beginning of the stack and high address to low address is *reversed*:

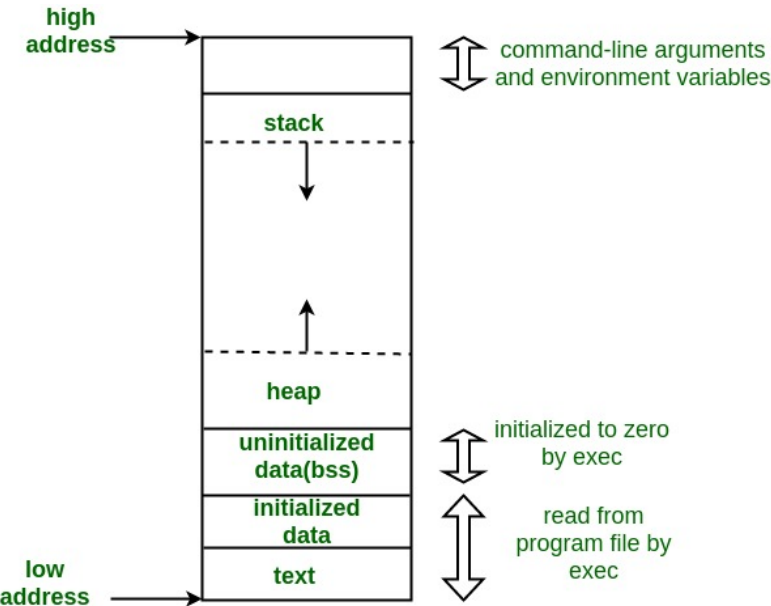
$$i_{\text{stack}} \mapsto \text{high address} - i_{\text{stack}}$$

Call stack is composed of stack frames (i.e. "activation records"), with each stack frame corresponding to a subroutine call that's not yet terminated with a routine (at any time).

The *frame pointer* FP points to location where stack pointer was.

Stack pointer usually is a register that contains the "top of the stack", i.e. stack's "low address" currently, cf. [Understanding the stack](#), i.e.

(2)
$$\text{eval}(RSP) = \text{high address} - i_{\text{stack}}$$



7.0.2. *Mathematical description of strategy for stack buffer overflow exploitation.* Let $n_{\text{stack}} = n_{\text{stack}}(t)$. Index the stack with i_{stack} (from "bottom" of the stack to the "top" of the stack):

$$0 < i_{\text{stack}} < n_{\text{stack}} - 1$$

Recall that $i_{\text{stack}} \in \mathbb{Z}^+$ and

$$i_{\text{stack}} \mapsto \text{high address} - i_{\text{stack}} \equiv x = x(i_{\text{stack}}) \in \text{Addresses}_{\text{Memory block}} \subset \text{Obj}(\mathbf{Address})$$

Let an array of length L (e.g. `char` array) `buf`, with $\&\text{buf} = \&\text{buf}(0) \in \text{Obj}(\mathbf{Address})$, be s.t. $\&\text{buf} = x(n_{\text{stack}} - 1)$ (starts at "top of the stack and "lowest" address of stack at time t , s.t.

$$\&\text{buf}(j) = \&\text{buf}(0) + j\text{sizeof}(T)$$

with $T \in \mathbf{Types}$).

Suppose return address of a function (such as `main`), $\text{eval}(\text{RIP})$ be

$$\text{eval}(\text{RIP}) = \&\text{buf} + L \text{ or at least } \text{eval}(\text{RIP}) \geq \&\text{buf} + L$$

If we write to `buf` more values than L , we can write over $\text{eval}(\text{RIP})$, making $\text{eval}(\text{RIP})$ a different value than before.

7.1. **Stack.** cf. <https://www.geeksforgeeks.org/memory-layout-of-c-program/>

Traditionally, stack area adjoined heap area and they grow in opposite direction.

When stack ptr meets heap ptr, free memory exhausted.

(with modern large address spaces, virtual memory techniques, they may be placed almost anywhere, but they still typically grow opposite direction)

stack area contains program stack, UFO structure, typically located in higher parts of memory.

On standard x86, grows toward address 0.

"stack ptr" register tracks top of stack, adjusted to point to top each time a value is "pushed" onto stack.

set of values pushed for 1 function call is "stack frame", stack frame consists at minimum of return address.

Newly called function allocates room on stack for its automatic and temporary variables.

\implies recursion: each time recursive function calls itself, new stack frame is used, so 1 set of variables (in recursion) doesn't interfere with variables from another instance of the function.

cf. [Where are static variables in C/C++? Tutorialspoint](#)

static variables remain in memory; lifetime is entire program, stored in data segment of memory (near heap); data segment is part of virtual address space of a program.

asked here <https://www.codingame.com/work/cpp-interview-questions/>, what is a static variable

cf. Ferres (2010) [8]

Stack and functions: When a function executes, it may add some of its state data to top of the stack (EY : 20171111, stack grows downward, so "top" is smallest address?); when function exits, stack is responsible for removing that data from stack.

In most modern computer systems, each thread has a reserved region of memory, stack. Thread's stack is used to store location of function calls in order to allow return statements to return to the correct location.

- OS allocates stack for each system-level thread when thread is created.
- Stack is attached to thread, so when thread exits, that stack is reclaimed, vs. heap typically allocated by application at runtime, and is reclaimed when application exits.
- When thread is created, stack size is set.
- Each byte in stack tends to be reused frequently, meaning it tends to be mapped to the processor's cache, making it very fast.
- Stored in computer RAM, *just like* the heap.
- Implemented with an actual stack data structure.
- stores local data, return addresses, used for parameter passing
- Stack overflow, when too much stack is used (mostly from infinite (or too much) recursion, and very large allocation)
- Data created on stack can be used without pointers.

Also note, for **physical location in memory**, because of [Virtual Memory](#), makes your program think that you have access to certain addresses where physical data is somewhere else (even on hard disc!). Addresses you get for stack are in increasing order as your call tree gets deeper.

[memory management - What and where are the stack and heap? Stack Overflow, Tom Leys' answer](#)

cf. <https://stackoverflow.com/questions/79923/what-and-where-are-the-stack-and-heap/80113#80113>

Stack is memory set aside for a thread of execution. Each thread gets a stack, while there's typically only 1 heap for the application (although it isn't uncommon to have multiple heaps for different types of allocation).

To what extent are they controlled by the OS or language runtime?

The OS allocates the stack for each system-level thread when the thread is created. Typically the OS is called by the language runtime to allocate the heap for the application.

What is their scope?

The stack is attached to a thread, so when the thread exits the stack is reclaimed. The heap is typically allocated at application startup by the runtime, and is reclaimed when the application (technically process) exits.

What determines the size of each of them?

The size of the stack is set when a thread is created. The size of the heap is set on application startup, but can grow as space is needed (the allocator requests more memory from the operating system).

7.2. **Stack overflow.** If you use heap memory, and you overstep the bounds of your allocated block, you have a decent chance of triggering a segmentation fault (not 100

On stack, since variables created on stack are always contiguous with each other; writing out of bounds can change the value of another variable. e.g. buffer overflow.

7.3. **Heap.** Heap contains a linked list of used and free blocks. New allocations on the heap (by `new` or `malloc`) are satisfied by creating suitable blocks from free blocks.

This requires updating list of blocks on the heap. This meta information about the blocks on the heap is stored *on the heap* often in a small area in front of every block.

- Heap size set on application startup, but can grow as space is needed (allocator requests more memory from OS)
- heap, stored in computer RAM, like stack.

7.3.1. *Memory leaks.* Memory leaks occurs when computer program consumes memory, but memory isn’t released back to operating system.

”Typically, a memory leak occurs because dynamically allocated memory becomes unreachable.” (Ferres (2010) [8]).

Programs `./Cmemory/heapstack/Memleak.c` deliberately leaks memory by losing the pointer to allocated memory.

Note, generally, the OS delays real memory allocation until something is written into it, so program ends when virtual addresses run out of bounds (per process limits).

7.4. **More Segmentation Faults.** The operating system (OS) is running the program (its instructions). Only from the hardware, with **memory protection**, with the OS be signaled to a memory access violation, such as writing to read-only memory or writing outside of allotted-to-the-program memory, i.e. data segments. On `x86_64` computers, this **general protection fault** is initiated by protection mechanisms from the hardware (processor). From there, OS can signal the fault to the (running) process, and stop it (abnormal termination) and sometimes core dump.

For **virtual memory**, the memory addresses are mapped by program called *virtual addresses* into *physical addresses* and the OS manages virtual addresses space, hardware in the CPU called memory management unit (*MMU*) translates virtual addresses to physical addresses, and kernel manages memory hierarchy (eliminating possible overlays). In this case, it’s the *hardware* that detects an attempt to refer to a non-existent segment, or location outside the bounds of a segment, or to refer to location not allowed by permissions for that segment (e.g. write on read-only memory).

7.4.1. *Dereferencing a ptr to a NULL ptr (in C) at OS, hardware level.* The problem, whether it’s for dereferencing a pointer that is a null pointer, or uninitialized pointer, appears to (see the `./Cmemory/` subfolder) be at this instruction at the register level:

```
x0000000000040056c <+38>:      movzbl (%rax),%eax
```

```
// or
```

```
0x000000000004004be <+24>:      movss  %xmm0,(%rax)
```

involving the register RAX, a temporary register and to return a value, upon assignment. And in either case, register RAX has trying to access virtual (memory) address `0x0` (to find this out in `gdb`, do `i r` or `info register`).

Modern OS’s run user-level code in a mode, such as *protected mode*, that uses ”paging” (using secondary memory source than main memory) to convert virtual addresses into physical addresses.

For each process (thread?), the OS keeps a *page table* dictating how addresses are mapped. Page table is stored in memory (and protected, so user-level code can’t modify it). For every memory access, given (memory) address, CPU translates address according to the page table.

When address translation fails, as in the case that *not all addresses are valid*, and so if a memory access generates an invalid address, the processor (hardware!) raises a *page fault exception*. ”This triggers a transition from *user mode* (aka *current privilege level (CPL) 3* on `x86/x86-64`) into *kernel mode* (aka *CPL 0*) to a specific location in the kernel’s code, as defined by the *interrupt descriptor table* (IDT).” cf. [What happens in OS when we dereference a NULL pointer in C?](https://stackoverflow.com/questions/12645647/what-happens-in-os-when-we-dereference-a-null-pointer-in-c)

Kernel regains control and send signal (EY : 20171115 to the OS, I believe).

In modern OS’s, page tables are usually set up to make the address `0` an invalid virtual address.

cf. [What happens in OS when we dereference a NULL pointer in C?](https://stackoverflow.com/questions/12645647/what-happens-in-os-when-we-dereference-a-null-pointer-in-c)

Part 2. C++

8. STATIC

9. FREE STORE

cf. **GotW #9, Memory Management - Part I**

Free store is 1 of the 2 dynamic memory areas, allocated/freed by `new/delete` (free store and heap). Object lifetime can be less than the time the storage is allocated.

Let $t_{0,obj}, t_{f,obj}, t_{f,obj} \geq t_{0,obj},$

$t_{0,alloc}, t_{f,alloc}, t_{f,alloc} \geq t_{0,alloc},$

with the condition that

$$t_{0,alloc} \leq t_{0,obj} \leq t_{f,obj} \leq t_{f,alloc}$$

For $t \in (t_{0,alloc}, t_{0,obj})$, storage may be accessed and manipulated through a `void*` but non of the proto-object’s *nonstatic* members or member functions may be accessed, have their addresses taken, or be otherwise manipulated.

cf. 11.2 Introduction of Ch. 11 Select Operations [9].

10. COPY VS. MOVE

cf. 17.1 Introduction of Ch. 17 Construction, Cleanup, Copy, and Move of Stroustrup [9].

Difference between *move* and *copy*: after a copy, 2 objects must have same value; whereas after a move, the source of the move isn’t required to have its original value. So moves can be used when source object won’t be used again.

Refs.: Sec. 3.2.1.2, Sec. 5.2, notion of moving a resource, Sec. 13.2-Sec.13.3, object lifetime and errors explored further in Stroustrup [9]

5 situations in which an object is copied or moved:

- as source of an *assignment*
- as object initializer
- as function argument
- as function return value
- as an exception

10.1. **Copy constructor.** cf. **Copy constructors, cppreference.com**

Copy constructor of class T is non-template constructor whose 1st parameter is `T&`, `const T&`, `volatile T&`, or `const volatile T&`.

```
class_name (  const  class_name & )
class_name (  const  class_name & ) = default ;
class_name (  const  class_name & ) = delete ;
```

10.1.2. *Explanation.*

- (1) Typical declaration of a copy constructor.
- (2) Forcing copy constructor to be generated by the compiler.
- (3) Avoiding implicit generation of copy constructor.

Copy constructor called whenever an object is **initialized** (by **direct-initialization** or **copy-initialization**) from another object of same type (unless **overload resolution** selects better match or call is **elided** (???)), which includes

- initialization `T a = b;` or `T a(b);`, where b is of type T;

- function argument passing: `f(a);`, where `a` is of type `T` and `f` is `void f(T t);`
- function return: `return a;` inside function such as `T f()`, where `a` is of type `T`, which has no **move constructor**.

```
struct A
{
    int n;
    A(int n = 1) : n(n) { }
    A(const A& a) : n(a.n) { } // user-defined copy ctor
};

struct B : A
{
    // implicit default ctor B::B()
    // implicit copy ctor B::B(const B&)
};

int main()
{
    A a1(7);
    A a2(a1); // calls the copy ctor
    B b;
    B b2 = b;
    A a3 = b; // conversion to A& and copy ctor
}
```

i.e. cf. [Copy Constructor in C++](#)

Definition 1. *Copy constructor* is a member function which initializes an object using another object of the same class.

10.1.4. *When is copy constructor called?*

- (1) When object of class returned by value
- (2) When object of class is passed (to a function) by value as an **argument**.
- (3) When object is constructed based on another object of same class (or overloaded)
- (4) When compiler generates temporary object

However, it's not guaranteed copy constructor will be called in all cases, because C++ standard allows compiler to optimize the copy away in certain cases.

10.1.5. *When is used defined copy constructor needed? shallow copy, deep copy.* If we don't define our own copy constructor, C++ compiler creates default copy constructor which does member-wise copy between objects.

We need to define our own copy constructor only if an object has pointers or any run-time allocation of resource like file handle, network connection, etc.

10.1.6. *Default constructor does only shallow copy.*

10.1.7. *Deep copy is possible only with user-defined copy constructor.* We thus make sure pointers (or references) of copied object point to new memory locations.

```
MyClass t1, t2;
MyClass t3 = t1;           // ———> (1)
t2 = t1;                   // ———> (2)
```

Copy constructor called when new object created from an existing object, as copy of existing object, in (1). Assignment operator called when already initialized object is assigned a new value from another existing object, as assignment operator is called in (2).

10.1.9. *Why argument to a copy constructor should be const?* cf. [Why copy constructor argument should be const in C++?, geeksforgeeks.org](#)

- (1) Use **const** in C++ whenever possible so objects aren't accidentally modified.
- (2) e.g.

```
#include <iostream>

class Test
{
    /* Class data members */
public:
    Test(Test &t)    { /* Copy data members from t */ }
    Test()           { /* Initialize data members */ }
};

Test fun()
{
    Test t;
    return t;
};

int main()
{
    Test t1;
    Test t2 = fun();  error: invalid initialization of non-const reference of type Test& from a
}
```

`fun()` returns by value, so compiler creates temporary object which is copied to `t2` using copy constructor (because this temporary object is passed as argument to copy constructor since compiler generates temp. object). Compiler error is because **compiler-created temporary objects cannot be bound to non-const references**.

10.2. **Move Constructor.** For a class, to control what happens when we move, or move and assign object of this class type, use special member function *move constructor*, *move-assignment operator*, and define these operations. Move constructor and move-assignment operator take a (usually nonconst) rvalue reference, to its type. Typically, move constructor moves data from its parameter into the newly created object. After move, it must be safe to run the destructor on the given argument. cf. Ch. 13 of Lippman, Lajole, and Moo (2012) [11]

11. VTABLE; VIRTUAL TABLE

I was given this answer to a question I posed to a 20 year C++ veteran and it was such an important answer (as I did not know a virtual table existed, at all before), that I will copy this, repeat this and explore this extensively:

"The keyword you're looking for is virtual table: " [How are virtual functions and vtable implemented?, stackoverflow](#)


```
    Impl *pImpl;    // and pointer to it
};
```

Because `Widget` no longer mentions types `std::string`, `std::vector`, and `Gadget`, `Widget` clients no longer need to `#include` headers for these types. That speeds compilation.

incomplete type is a type that has been declared, but not defined, e.g. `Widget::Impl`. There are very few things you can do with an incomplete type, but declaring a pointer to it is 1 of them.

`std::unique_ptr`s is advertised as supporting incomplete types. But, when `Widget w`;, `w`, is destroyed (e.g. goes out of scope), destructor is called and if in class definition using `std::unique_ptr`, we didn't declare destructor, compiler generates destructor, and so compiler inserts code to call destructor for `Widget`'s data member `m_Impl` (or `pImpl`).

`m_Impl` (or `pImpl`) is a `std::unique_ptr<Widget::Impl>`, i.e., a `std::unique_ptr` using default deleter. The default deleter is a function that uses `delete` on raw pointer inside the `std::unique_ptr`. Prior to using `delete`, however, implementations typically have default deleter employ C++11's `static_assert` to ensure that raw pointer doesn't point to an incomplete type. When compiler generates code for the destruction of the `Widget w`, then, it generally encounters a `static_assert` that fails, and that's usually what leads to the error message.

To fix the problem, you need to make sure that at point where code to destroy `std::unique_ptr<Widget::Impl>` is generated, `Widget::Impl` is a complete type. The type becomes complete when its definition has been seen, and `Widget::Impl` is defined inside `widget.cpp`. For successful compilation, have compiler see body of `Widget`'s destructor (i.e. place where compiler will generate code to destroy the `std::unique_ptr` data member) only inside `widget.cpp` after `Widget::Impl` has been defined.

For compiler-generated move assignment operator, move assignment operator needs to destroy object pointed to by `m_Impl` (or `pImpl`) before reassigning it, but in the `Widget` header file, `m_Impl` (or `pImpl`) points to an incomplete type. Situation is different for move constructor. Problem there is that compilers typically generate code to destroy `pImpl` in the event that an exception arises inside the move constructor, and destroying `pImpl` requires `Impl` be complete.

Because problem is same as before, so is the fix - *move definition of move operations into the implementation file*.

For copying data members, support copy operations by writing these functions ourselves, because (1) compilers won't generate copy operations for classes with move-only types like `std::unique_ptr` and (2) even if they did, generated functions would copy only the `std::unique_ptr` (i.e. perform a *shallow copy*), and we want to copy what the pointer points to (i.e., perform a *deep copy*).

If we use `std::shared_ptr`, there'd be no need to declare destructor in `Widget`.

Difference stems from differing ways smart pointers support custom deleters. For `std::unique_ptr`, type of deleter is part of type of smart pointer, and this makes it possible for compilers to generate smaller runtime data structures and faster runtime code. A consequence of this greater efficiency is that pointed-to types must be complete when compiler-generated special functions (e.g. destructors or move operations) are used. For `std::shared_ptr`, type of deleter is not part of the type of smart pointer. This necessitates larger runtime data structures and somewhat slower code, but pointed-to types need not be complete when compiler-generated special functions are employed.

Part 3. Integers, numeric encodings, number representation, binary representation, hexadecimal representation

12. INTRODUCTION: `int` $\neq \mathbb{Z}$, `float` $\neq \mathbb{R}$

cf. [Systems 1, Introduction to Computer Systems, Introduction lecture at University of Texas, CS429; Don Fussell](#).

`int` $\neq \mathbb{Z}$, `float` $\neq \mathbb{R}$

e.g. $x^2 \geq 0$?

`float`: Yes!

`int`: $40000 \times 40000 \rightarrow 1600000000$

$50000 \times 50000 \rightarrow ??$

Is $(x + y) + z = x + (y + z)$?

`unsigned` and `signed int`: Yes!

`float`'s

$$(1e20 + -1e20) + 3.14 \rightarrow 3.14$$

$$1e20 + (-1e20 + 3.14) \rightarrow ??$$

Can't assume "usual" properties due to finiteness of representations, so that integer operations satisfy "ring" properties - commutativity, associativity, distributivity
floating point operations satisfy "ordering" properties - monotonicity, values of signs

13. BITS AND BYTES

cf. [Lecture 2, Bits and Bytes, CS429](#).

Consider a (number) system with radix or base b ($b > 1$),
string of digits d_1, \dots, d_n ,

$$(3) \quad c = \sum_{i=1}^n d_i b^{n-i}, \quad 0 \leq d_i < b \quad \forall i = 1 \dots n$$

with notation $c \equiv c_b$ to help denote the base or "radix", e.g. 15213_{10} , 00000000_2 .

Byte $\equiv B$, 1 Byte = 8 bits. Note that $2^8 = 256$

e.g. Binary 00000000_2 to 11111111_2

Decimal: 0_{10} to 255_{10}

Hexadecimal: 00_{16} to FF_{16}

13.1. **Machine Words.** Machine has "Word size".

Nominal size of integer-valued data, including addresses.

13.2. **Word-oriented Memory Organization.** Addresses specify Byte locations.

Addresses of successive words differ by 4 (32-bit) or 8 (64-bit).

32-bit words:

Addr = 0000, Addr = 0004

64-bit words:

Addr = 0000, Addr = 0008

13.3. **Byte-ordering: Big-Endian, Little-Endian.** $b = 8$

$$(4) \quad c = \sum_{i=0}^{n-1} d_i 8^{n-1-i}$$

Big-Endian: Least significant byte has highest address.

For address a ,

$$a \mapsto d_0$$

$$a + 1 \mapsto d_1$$

$$\vdots$$

$$a + n - 1 \mapsto d_{n-1}$$

Little-Endian: Least significant byte has lowest address

$$a \mapsto d_{n-1}$$

$$a + 1 \mapsto d_{n-2}$$

$$\vdots$$

$$a + n - 1 \mapsto d_0$$

13.4. **Representing Strings.** C Strings

- represented by array of characters, each character encoded in ASCII format - string should by null-terminated meaning final character = 0

13.5. **Machine-Level Code Representation.** Encode program as sequence of Instructions.

Each simple operation - arithmetic operation, read or write memory, conditional branch. - instructions encoded as bytes, i.e. programs are byte sequences too!

13.6. **Boolean Algebra.** Boole’s Algebraic representation of logic.

And:
 $A \& B = 1$ when both $A = 1$ and $B = 1$
Or:
 $A | B = 1$ when either $A = 1$ or $B = 1$
Not:
 $\sim A = 1$ when $A = 0$,
 $\sim A = 0$ when $A = 1$,

Exclusive-Or (Xor) $A \wedge B = 1$ when either $A = 1$ or $B = 1$, but not both

\wedge	0	1
0	0	1
1	1	0

13.6.1. *Application of Boolean Algebra.* Claude Shannon 1937 MIT Master’s Thesis. Encode closed switch as 1, open switch as 0. Connection when $A \& \sim B | \sim A \& B = A \wedge B$

13.6.2. *Integer Algebra.* $\langle \mathbb{Z}, +, *, -, 0, 1 \rangle$ forms a ring

13.6.3. *Boolean Algebra.* $\langle \{0, 1\}, |, \&, \sim, 0, 1 \rangle$ forms a ”Boolean algebra”

Or (|) is ”sum” operation.
And (&) is ”product” operation.
 \sim is ”complement” operation.
0 is additive identity.
1 is multiplicative identity.

cf. https://en.wikipedia.org/wiki/Bitwise_operation

Bitwise OR may be used to set to 1 selected bits of register. e.g. fourth bit of 0010 may be set by performing bitwise OR with pattern with only 4th bit set.

Bitwise AND is equivalent to multiplying corresponding bits. Thus, if both bits in compared position are 1, bit in resulting binary representation is ($1 \times 1 = 1$); otherwise result is 0 ($1 \times 0 = 0$ and $0 \times 0 = 0$).

Operation may be used to determine whether particular bit is set (1) or clear (0). e.g. given bit pattern 0011 (decimal 3), to determine whether 2nd bit is set, use bitwise AND with bit pattern containing 1 only in 2nd. bit. Because result 0010 is non-zero, we know 2nd. bit in original pattern is set. This is often called *bit masking* (by analogy, use of masking tape covers, or *masks*, portions that are not of interest)

Bitwise AND may be used to clear selected bits (or flags) of a register in which each bit represents an individual Boolean state.

- This technique is an efficient way to store a number of Boolean values using as little memory as possible.
- Also, easy to check parity (even or odd?) of binary number by checking value of lowest valued bit.

Bitwise XOR may be used to invert selected bits in a register (also called toggle or flip). Any bit may be toggled by XORing with 1. e.g. given bit pattern 0010, 2nd and 4th bits may be toggled by bitwise XOR with bit pattern containing 1 in 2nd. and 4th. positions.

Consider Boolean Ring:

$$\langle \{0, 1\}, \wedge, \&, I, 0, 1 \rangle \simeq \mathbb{Z}_2 \equiv \text{integers} \mod 2$$

where I is identity operation $I(A) = A$,
and $A \wedge A = 0$ is the additive inverse (existence) property.

13.6.4. *De Morgan’s Laws.*

$$A \& B = \sim (\sim A | \sim B)$$

$$A | B = \sim (\sim A \& \sim B)$$

13.6.5. *Exclusive-Or using Inclusive Or.*

$$A \wedge B = (\sim A \& B) | (A \& \sim B)$$

$$A \wedge B = (A | B) \& \sim (A \& B)$$

All properties of Boolean Algebra apply to bit vectors.

13.7. **Representing and Manipulating Sets; bit vectors as sets.** Width w bit vector represents subsets of $\{0, \dots, w - 1\}$, s.t. for $a \in \{0, 1\}^w$, $a_j \in \{0, 1\}$, $\forall j = 0, 1, \dots w - 1$.

If $a_j = 1, j \in A$; otherwise,
If $a_j = 0, j \notin A$

13.7.1. *Operations.*

$\& \rightarrow$ intersection
 $| \rightarrow$ union
 $\wedge \rightarrow$ symmetric difference
 $\sim \rightarrow$ complement

Logic operations in C: $\&\&$, $||$, $!$

Shift operations

cf. [Systems 1, Integers lecture at University of Texas, CS429](#) is a **very good lecture**; both *mathematically rigorous* and full of useful, *clear* examples.

14. UNSIGNED INTEGERS

Consider $\mathbb{Z}_{2^w}^+ \rightarrow \mathbb{Z}^+$, where $\mathbb{Z}_{2^w}^+$ represent ”unsigned” integers.

$$(5) \quad \mathbb{Z}_{2^w}^+ \rightarrow \mathbb{Z}^+ \\ (x_0, x_1, \dots, x_{w-1}) \mapsto \sum_{i=0}^{w-1} x_i \cdot 2^i$$

To recap using CS429’s notation, for an unsigned integer x ,
 $B2U \equiv$ base-2-unsigned

$$B2U(x) = \sum_{i=0}^{w-1} x_i \cdot 2^i$$

where $B2U : \text{”}\mathbb{Z}^+ \text{”} \rightarrow \mathbb{Z}_{2^w}^+$
So w is the total number of ”bits” to represent x .
For hexadecimal (-based) numbers, observe this relationship:

$$(6) \quad 2^w = 16^v = 2^{4v} \text{ so } w = 4v$$

So for a hexadecimal number, v hexadecimal numbers represent $w = 4v$ bits.
Observe that

$$(7) \quad \mathbb{Z}_{2^w}^+ = \{0, \dots 2^w - 1\}$$

Also, **modular addition** $(\mathbb{Z}_{2^w}^+, +)$ forms an **abelian group**.

15. TWO’S COMPLEMENT

cf. [Systems 1 Integers](#)

Denote the notation for so-called ”two’s complement” numbers as \mathbb{Z}_{2^w} , which is a representation for integers in such a manner:

$$\begin{aligned} \mathbb{Z}_{2^w} &\rightarrow \mathbb{Z} \\ (8) \qquad (x_{w-1}, x_{w-2}, \dots x_0) &\mapsto -x_{w-1} \cdot 2^{w-1} + \sum_{i=0}^{w-2} x_i \cdot 2^i \end{aligned}$$

with

$$(9) \qquad \mathbb{Z}_{2^w} = \{-2^{w-1} \dots 2^{w-1} - 1\}$$

To recap using CS429’s notation, for integer $x \in \mathbb{Z}$,
 $B2T \equiv$ base-2-Two’s complement

$$B2T(x) = -x_{w-1} \cdot 2^{w-1} + \sum_{i=0}^{w-2} x_i \cdot 2^i$$

where $B2T : \text{”}\mathbb{Z}\text{”} \rightarrow \mathbb{Z}_{2^w} = \{-2^{w-1} \dots 2^{w-1} - 1\}$

Observe that

$$\begin{aligned} (10) \qquad \max(\mathbb{Z}_{2^w}^+) &= 2 \max \mathbb{Z}_{2^w} + 1 \\ |\min \mathbb{Z}_{2^w}| &= \max \mathbb{Z}_{2^w} + 1 \end{aligned}$$

To recap and to clean up and unify the notation, recall the following:
radix (i.e. base) b
string of digits $d_1, \dots d_n$,

$$c = \sum_{i=1}^n d_i b^{n-i}, \quad 0 \leq d_i < b \quad \forall i = 1 \dots n$$

cf. [wikipedia](#), **”Radix”** i.e.

$$\begin{aligned} \{0, 1, \dots b - 1\}^n &\rightarrow \mathbb{Z} \\ (d_1, \dots d_n) &\mapsto c = \sum_{i=1}^n d_i b^{n-i} \end{aligned}$$

or (alternative notation),

$$\begin{aligned} \{01, , \dots b - 1\}^w &\rightarrow \mathbb{Z} \\ (d_{w-1}, d_{w-2}, \dots, d_0) &\mapsto c = \sum_{i=0}^{w-1} d_i b^i \end{aligned}$$

In Fussell’s notation, Fussell (2011) [\[13\]](#),

$$\begin{aligned} \{0, 1, \dots b - 1\}^w &\rightarrow \mathbb{Z} \text{ or } \mathbb{Z}_{2^w}^+ \rightarrow \mathbb{Z}^+ \\ (x_{w-1}, x_{w-2}, \dots x_0) &\mapsto c = \sum_{i=0}^{w-1} x_i 2^i, \ x_i < 2 \end{aligned} \qquad B2U(X) = B2U(x_{w-1}, \dots x_0) = \sum_{i=0}^{w-1} x_i 2^i$$

I will use this notation:

$$B2U(x_w, x_{w-1}, \dots x_1) = \sum_{i=1}^w x_i 2^{i-1}$$

For the *Two’s Complement*,

$$B2T(X) = B2T(-x_{w-1}, x_{w-2}, \dots x_0) = -x_{w-1} 2^{w-1} + \sum_{i=0}^{w-2} x_i 2^i$$

$$\{0, 1, \dots b - 1\}^w \rightarrow \mathbb{Z}$$

or with the notation I’ll use,

$$B2T(-x_w, x_{w-1}, \dots x_1) = -x_w 2^{w-1} + \sum_{i=1}^{w-1} x_i 2^{i-1}$$

Note that for hexadecimal,

$$c = \sum_{i=1}^w x_i 2^{i-1} = \sum_{i=1}^v y_i (16)^{i-1} = \sum_{i=1}^v y_i 2^{4i-4}$$

v hexadecimal digits $\leftrightarrow 4v = w$ bits.

15.1. **Numeric Ranges.** For unsigned values,

$$\begin{aligned} \text{UMin} &= 0 \\ 000 \dots 0 \\ \text{UMax} &= 2^w - 1 \end{aligned}$$

For why $\text{UMax} = 2^w - 1$, consider the geometric progression:

$$\begin{aligned} &\sum_{k=1}^n ar^{k-1} \\ (1-r) \sum_{k=1}^n ar^{k-1} &= a - ar^n \text{ or } \sum_{k=1}^n ar^{k-1} = \frac{a(1-r^n)}{1-r} \\ \implies \sum_{i=1}^w 2^{i-1} &= \frac{1-2^w}{1-2} = 2^w - 1 \end{aligned}$$

$$\begin{aligned} \text{TMin} &= -2^{w-1} \\ 100 \dots 0 \\ \text{TMax} &= 2^{w-1} - 1011 \dots 1 \end{aligned}$$

$$-1 \leftrightarrow 111 \dots 1$$

15.2. **Negating with complement.**

$$\begin{aligned} (11) \qquad \sim: \mathbb{Z}_{2^w} &\rightarrow \mathbb{Z} \text{ s.t.} \\ \sim x + 1 &= -x \quad \forall x \in \mathbb{Z}_{2^w} \\ \sim: \mathbb{Z}_{2^w}^+ &\rightarrow \mathbb{Z} \text{ s.t.} \\ \sim x &= 2^w - 1 - x \end{aligned}$$

15.3. **Power of 2 Multiply with shift (left shift bitwise operation).** For both $\mathbb{Z}_{2^w}^+, \mathbb{Z}_{2^w}$,

$$(12) \qquad u << k = u \cdot 2^k \qquad \forall u \in \mathbb{Z}_{2^w}^+ \text{ or } \mathbb{Z}_{2^w} \text{ and } k \in \mathbb{Z}$$

15.4. **Unsigned power-of-2 Divide, with shift (right shift bitwise operator).**

$$(13) \qquad u >> k = \lfloor u/2^k \rfloor \qquad \forall u \in \mathbb{Z}_{2^w}^+, \quad k \in \mathbb{Z}$$

16. ENDIANNESS

Consider $address \in \mathbb{Z}_{2^3}^+$. Consider a $value \in \mathbb{Z}_{2^3}^+$ or \mathbb{Z}_{2^3} .
 $\forall (x_0, x_1, \dots x_{w-1}) \in \mathbb{Z}_{2^3}^+$ or \mathbb{Z}_{2^3}
Given address a ,

16.1. **Big-endian.**

Definition 2 (Big-endian). *Most-significant byte value is at the lowest address, i.e.*

(14)

$$\begin{aligned} a &\mapsto x_{w-1} \\ a+1 &\mapsto x_{w-2} \\ a+2 &\mapsto x_{w-3} \\ &\vdots \\ a+w-1 &\mapsto x_0 \end{aligned}$$

16.2. **Little-endian.**

Definition 3 (Little-endian). *Least-significant byte value is at the lowest address, i.e.*

(15)

$$\begin{aligned} a &\mapsto x_0 \\ a+1 &\mapsto x_1 \\ &\vdots \\ a+w-1 &\mapsto x_{w-1} \end{aligned}$$

For $Little\text{-}endian \rightarrow \mathbb{Z}$,
($addresses$) $\in \mathbb{Z}_{2^3}^+ \mapsto \mathbb{Z}$,
 $a \mapsto (x_0, x_1, \dots x_{w-1})$
For $\mathbb{Z} \rightarrow Big\text{-}endian$,
 $n \mapsto (x_{w-1}, x_{w-2}, \dots x_0)$.

Instead, to tackle the confusing problem of byte ordering, think about the integer as itself first, and then consider mapping multibyte binary values to memory.

(16)

$$\begin{aligned} \mathbb{Z} &\rightarrow (\mathbb{Z}_{2^w} \longleftarrow \mathbb{Z}_{(2^3)^w} = \textbf{addresses}) \\ \mathbb{Z} \ni (x_{w-1}, x_{w-2}, \dots x_1, x_0) &\xrightarrow{Little\text{-}Endian} ((x_0, x_1, \dots x_{w-2}, x_{w-1}) \longleftarrow (a, a+1, \dots a+w-2, a+w-1)) \\ \mathbb{Z} \ni (x_{w-1}, x_{w-2}, \dots x_1, x_0) &\xrightarrow{Big\text{-}Endian} ((x_{w-1}, x_{w-2}, \dots x_1, x_0) \longleftarrow (a, a+1, \dots a+w-2, a+w-1)) \end{aligned}$$

Part 4. Floating Point Numbers

17. FLOATING POINT, IEEE FLOATING

IEEE Standard 754
Recall our previous notation for the sum representation of a number c :

(17)

$$c = \sum_{i=1}^n d_i b^{n-i}; \quad 0 \leq d_i < b, \quad \forall i = 1 \dots n$$

e.g.

$$c = 42_{10} = 4 \cdot 10^{2-1} + 2 \cdot 10^{2-2} = d_1 b^{n-1} + d_2 b^{n-2}$$

By writing the exponent to be $n-i$, then we can write the digit representing the largest value in the summation ”starting from the left” as such:

$$c = (d_1, d_2, \dots d_{n-1}, d_n)$$

e.g. $(4, 2) = 42$.
However, the [CS429h](#) lectures uses the following notation for the `unsigned int`:

$$B2U(X) = \sum_{i=0}^{w-1} x_i 2^i \text{ or } (x_{w-1}, x_{w-2}, \dots x_1, x_0) \mapsto \sum_{i=0}^{w-1} x_i 2^i$$

e.g.

$$(x_3, x_2, x_1, x_0) = (1, 0, 1, 1) \mapsto x_0 2^0 + x_1 2^1 + x_2 2^2 + x_3 2^3$$

17.1. **Fractional Binary Numbers.**

17.1.1. *Representation.* Bits to right of ”binary point” represent fractional powers of 2.
Binary number is used to represent a rational number, which in turn is a floating-point *representation*.

(18)

$$\sum_{k=-j}^i b_k \cdot 2^k \equiv \sum_{k=-M}^N b_k 2^k$$

(19)

$$b \equiv (b_N, b_{N-1}, \dots b_1, b_0, b_{-1}, \dots b_{-M}) \mapsto \sum_{k=-M}^n b_k 2^k \equiv c$$

i.e.

$$\sum_{k=-M}^N b_k 2^k = b_N \cdot 2^N + b_{N-1} 2^{N-1} + \dots + b_1 \cdot 2 + b_0 \cdot 1 + b_{-1} \cdot \frac{1}{2} + \dots + b_{-M} \cdot 2^{-M}$$

e.g. $101.11_2 \mapsto 5\frac{3}{4}$
Divide by 2 by shifting right:

(20)

$$\begin{aligned} \frac{c}{2} &= \sum_{k=-M}^N b_k 2^k / 2 = \sum_{k=-M-1}^{N-1} b_{k+1} 2^k \\ (b_N, b_{N-1}, \dots b_{-M}) &\xrightarrow{\cdot \frac{1}{2}} (0, b_N, b_{N-1}, \dots b_{-M-1}, b_{-M}) \end{aligned}$$

Multiply by 2 by shifting left:

(21)

$$\begin{aligned} 2 \cdot c &= \sum_{k=-M}^N b_k \cdot 2^k \cdot 2 = \sum_{k=-M+1}^{N+1} b_{k-1} 2^k \\ (b_N, b_{N-1}, \dots b_{-M}) &\xrightarrow{\cdot 2} (b_N, b_{N-1}, \dots b_{-M}, 0) \end{aligned}$$

This implies that we first choose the ”most significant bit”, the digit b_k representing the largest value in the sum representation to start from the *left*.
Numbers of form $0.111111\dots_2$ just below 1.0 (remember those are bits, 0 or 1).

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^M} + \dots \rightarrow 1.0$$

Use $1.0 - \epsilon$ notation.

17.1.2. *Limitation of representable numbers.* Can only exactly represent numbers of form $\frac{x}{2^k}$
The other numbers have repeating bit representations.

Value	Representation
$\frac{1}{3}$	0.0101010101[01]... ₂
$\frac{1}{5}$	0.001100110011[0011]... ₂
$\frac{1}{10}$	0.0001100110011[0011]... ₂

17.1.3. *Floating Point Representation. Numerical Form:*

(22)
$$-1^s M 2^E$$

where

sign bit s , +1 or -1
significand (mantissa, or coefficient) M , normally a fractional value in range [1.0, 2.0)
Exponent E weights value by power of 2.

From [wikipedia](#),

(23)
$$\frac{s}{b^{p-1}} \cdot b^e$$

where s significand (ignoring an implied decimal point), p precision (number of digits in significand), base b .
e.g.

$$1.528535047 \times 10^5 = \frac{1528535047}{10^9} \times 10^5$$

Another example:

$$\begin{aligned} &\left(\sum_{n=0}^{p-1} c_n 2^{-n}\right) \times 2^e = \\ &(1 \times 2^{-0} + 1.2^{-1} + 0 \cdot 2^{-2} + 0 \cdot 2^{-3} + 1 \cdot 2^{-4} + \dots + 1 \cdot 2^{-23}) \times 2^1 \\ &\approx 1.5707964 \times 2 \approx 3.1415928 \end{aligned}$$

For the encoding of this numerical form,

Most significant bit (MSB) is sign bit.

exp field encodes E
frac field encodes M

s exp frac

Sizes for the encoding:
Single precision: 8 exp bits, 23 frac bits; 31 bits + 1 s bit = 32 bit
and so
 2^{8-1}

$$\sum_{i=0}^{M-1} c_i 2^{-i} \xrightarrow{\max} 2^{M-1} = 2^{23-1}$$

Double precision: 11 exp bits, 52 frac bits; 63 bits + 1 s bit = 64 bits
 $2^{\text{exp}-1} = 2^{11-1} = 1024$. This is the max value of e in decimal.
 $2^{M-1} = 2^{52-1}$

17.1.4. *"Normalized" Numeric Values.* Condition:

$$\text{exp} \neq 000 \dots 0 \text{ and } \text{exp} \neq 111 \dots 1$$

Exponent coded as biased value:

(24)
$$E = \text{Exp} - \text{Bias}$$

Exp: unsigned value determined by exp
Bias: Bias value

Note that E is the actual value desired, e.g. $2^{13} = 2^E$, whereas Exp is the actual binary representation.

- Single precision: 127 (Exp: 1...254, E: $-126 \dots 127$)
- Double precision: 1023 (Exp: 1...2046, E: $-1022 \dots 1023$)

In general

(25)
$$\text{Bias} = 2^{e-1} - 1$$

where e is number of exponent bits.
[wikipedia](#) says this: the exponent is "biased" in the engineering sense of the word - value stored is offset from actual value by the **exponent bias**. Biasing is done because exponents have to be signed value in order to be able to represent both tiny and huge values. It's not represented in two's complement, usual representation for signed values; it'd make comparison harder.
Exponent is stored as unsigned value which is suitable for comparison, and when interpreted, it's converted into an exponent within a *signed* range by subtracting the bias.
[wikipedia](#) uses this notation:

$$2^{k-1} - 1 = \text{bias for floating point number}$$

where k is number of bits in exponent.
Number of normalized floating-point numbers in system (B, P, L, U) where

- base B
- precision of system to P numbers
- L smallest exponent representable in system
- U largest exponent used in system

(26)
$$2(B-1)(B^{P-1})(U-L+1)+1$$

Smallest positive normalized floating-point number.
Underflow level = UFL = B^L
TODO: Understand all of this in https://en.wikipedia.org/wiki/Floating-point_arithmetic
Significand coded with implied leading 1:
Get extra leading bit for "free".
 $xxx \dots x$: bits of frac
Minimum when $000 \dots 0$ ($M = 1.0$) Maximum when $111 \dots 1$ ($M = 2.0 - \epsilon$)

17.1.5. *Denormalized Values and other Special Values.* Consider ”denormalized” values, the special case with the following condition:

$$\text{exp} = 000 \dots 0$$

The Value is this:
Exponent value $E = -\text{Bias} + 1$ (compare this to Eq. 24).
Significand value $M = 0.x_1x_2 \dots x_n$, i.e. $x_1x_2 \dots x_n$: bits of frac.
Consider these special cases:

- $\text{exp} = 000 \dots 0$, $\text{frac} = 000 \dots 0$, which represents value 0
Note that there are distinct values for +0 and −0
- $\text{exp} = 000 \dots 0$, $\text{frac} \neq 000 \dots 0$
These represent numbers that are very close to 0.0. They lose precision as they get smaller. There’s ”gradual underflow”. TODO: understand what it means for ”gradual underflow”, cf. https://www.cs.utexas.edu/users/fussell/courses/cs429h/lectures/Lecture_4-429h.pdf, pp. 10

Consider special values with the condition that

$$\text{exp} = 111 \dots 1$$

- They include the following cases:
- $\text{exp} = 111 \dots 1$, $\text{frac} = 000 \dots 0$. This represents the value at infinity, ∞ . Operation that overflows. There are positive infinity and negative infinity.
e.g. $1.0/0.0 = -1.0/-0.0 = +\infty$, $1.0/-0.0 = -\infty$
 - $\text{exp} = 111 \dots 1$, $\text{frac} \neq 000 \dots 0$ This is Not-a-Number (NaN). It represents case when no numeric value can be determined. e.g. `sqrt(-1)`, $\infty - \infty$.

Part 5. Complexity, Data Structures, Algorithms

18. COMPLEXITY, BIG-*O*

Also note, let N = number of elements in a collection.

$$\log_b N = n \leftrightarrow b^n = N$$

$N \log_b N = n \leftrightarrow b^n = N^N$ but rather, use Stirling’s approximation:

$$N \ln N \cong \log_b N! = \sum_{i=1}^N \log_b i = n$$

$$O(N \log_b N) = O(\log_b N!) \text{ via Stirling's approximation: } \ln N! = N \ln N - N + O(\ln N)$$

18.1. **Multi-Part Algorithms: Add vs. Multiply, e.g. $O(A + B)$ vs. $O(A * B)$.** If your algorithm is ”do this (A), then, when completed, do that (B)” then add runtimes: $O(A + B)$.
If your algorithm is ”do this for each time you do that”, then multiply runtimes: $O(A * B)$.

18.2. **Amoritized Time (e.g. dynamically resizing array ”ArrayList” or maybe `std::vector`).** Dynamically resize array x : if there are N elements, s.t. $N = \text{max. capacity at } N$, $|x| = N \mapsto |x| = 2N$ and there are N copies to be made ($T = N$).

Let X = number of elements to be inserted.
Suppose when $x = 1, 2, 4, 8, 16, \dots 2^j, \dots X$, $x = 2^j$ copies are made.
total number of copies:

$$\sum_{j=0}^{\lfloor \log_2 X \rfloor} 2^j = \sum_{j=0}^{\lfloor \log_2 X \rfloor} 2^{\lfloor \log_2 X \rfloor - j} = \sum_{j=0}^{\lfloor \log_2 X \rfloor} \frac{X}{2^j} = X \left(\frac{1 - 2^{-(\lfloor \log_2 X \rfloor + 1)}}{1 - 1/2} \right) = 2X (1 - 2/X) \cong 2X \text{ if } X \text{ large}$$

Note that $X = 2^y$
 $\log_2 X = y$
 X insertions take $O(2X)$. ”Amoritized” time for each insertion is $O(1)$.

18.3. $\log N$ **runtimes** ($O(\log N)$). e.g. binary search. find element (example) c in N -element *sorted* array. $x = x_i$, $i = 0, 1, \dots N - 1 \mapsto i' = 1, 2 \dots N$.
First, compare c to midpoint of array.
If $c = x_{\lceil \frac{N}{2} \rceil}$, done.
So let $n_1 = N$.

if $c < x_{\lceil \frac{N}{2} \rceil}$, consider $x_1, x_2, \dots x_{\lfloor \frac{N}{2} \rfloor}$, if $c > x_{\lceil \frac{N}{2} \rceil}$, consider $x_{\lceil \frac{N}{2} \rceil + 1}, \dots x_N$

Let $n_2 = \lfloor \frac{N}{2} \rfloor$.
By induction, $n_j = \lfloor \frac{N}{2^{j-1}} \rfloor$
We stop when we either find the value c or we’re down to just 1 element.
 \implies total runtime is the a matter of how many steps (dividing N by 2 each time).
 $J = ?$ s.t. $n_J = 1 = \lfloor \frac{N}{2^{J-1}} \rfloor \implies 0 = \log_2 N + (-J + 1)$ or

$$J = \log_2 N$$

When you see a problem where number of elements $n_0 = N, \dots n_j$ gets halved each time, it’ll likely be $O(\log_2 N)$ runtime.
Finding an element in a **balanced binary search tree**: $O(\log N)$; with each comparison, we go either left or right.

18.4. **Recursive run times** $O(2^N)$.

Let $n_1 = N$	2^1	$f(n_1 - 1) = f(N - 1)$ calls
$n_2 = N - 1$	$4 = 2^2$	$f(n_2 - 1)$ calls
\dots	\mapsto	\dots
$n_j = N - j + 1$	2^j	$f(n_j - 1)$ calls
until $j = N$	2^N	

$$\sum_{j=1}^N 2^j = \sum_{j=0}^{N-1} 2^{N-j} = 2^N \sum_{j=0}^{N-1} 2^{-j} = 2^N \left(\frac{1 - 2^{-N}}{1 - 1/2} \right) = 2^{N+1}(1 - 2^{-N}) \cong 2^N \quad (N \text{ large}) \implies 2^{N+1} - 1 \text{ nodes}$$

Try to remember this pattern. When you have a recursive function that makes multiple calls, runtime will often (not always) look like

(27) $O(\text{branches}^{\text{depth}})$

branches = number of times each recursive call branches.
space complexity is $O(N)$, only $O(N)$ nodes exist at any time.
cf. VI Big O, pp. 46, McDowell, Example 1

```
for (int i=0; i < N; i++)
{ sum += array[i]; }
```

$O(N)$.
cf. VI Big O, pp. 46, McDowell, Example 2

```
for (int i = 0; i < N; i++)
{
for (int j = 0; j < N; j++)
{
std::cout << i << j;
}
}

O(N * N) = O(N^2) Or see it as printing O(N^2) total number of pairs.
cf. VI Big O, pp. 46, McDowell, Example 3
```

```
for (int i = 0; i < N; i++)
{
for (int j = i + 1; j < N; j++)
{
std::cout << i << j;
}
}
```

$i = 1, 2 \dots N$
 $j = i + 1, \dots N$

$$\sum_{j=i+1}^N 1 = N - i$$
$$\sum_{i=1}^N (N - i) = N \cdot N - \sum_{i=1}^N i = N^2 - \frac{N(N + 1)}{2} = N^2 - \frac{N^2}{2} - \frac{N}{2} = \frac{N^2 - N}{2} \cong N^2 \quad (\text{large})$$

cf. VI Big O, pp. 46, McDowell, Example 4

```
for (int i = 0; i < NA; i++)
{
for (int j = i + 1; j < NB; j++)
{
if (a[i] < b[j])
{
std::cout << a[i] << b[j];
}
}
}
```

if statement within j ’s for loop is $O(1)$ time since it’s just a sequence of constant time statements.
 $\implies O(NA \cdot NB)$
cf. VI Big O, pp. 47, McDowell, Example 5

```
for (int i = 0; i < NA; i++)
{
for (int j = i + 1; j < NB; j++)
{
for (int k = 0; k < 1000000; k++)
{
//...
}
}
```

```
}

 $\implies O(NA \cdot NB \cdot 1000000) = O(100000(NA)(NB)) \cong O(NA \cdot NB)$ 
100,000 units of work is still constant, so run time is  $O(NA \cdot NB)$ .
cf. VI Big O, pp. 48, McDowell, Example 6
```

```
void reverse(int array[], const int N)
{
for (int i = 0; i < N / 2; i++)
{
int other = N - i - 1;
int temp = array[i];
array[i] = array[other];
array[other] = temp;
}
}
```

$O(N)$ time. The fact that it only goes through half of the array (in terms of iterations) doesn’t impact big O time.
cf. VI Big O, pp. 47, McDowell, Example 8
Let the longest string be of length L .

sort each string $\rightarrow L \log L$ (merge sort or best ”worst” case for a sort)
 N_a = number of strings in the array of strings.
 $N_a L \log L$ = total number of sorts.

Sort the full array.
You should also take into account that you need to compare the strings.
Each string comparison takes $O(L)$ time (compare each string element at each position).
 $O(N_a \log N_a)$ comparisons (sort the full array of strings)

$$\implies \boxed{O(N_a L \log L) + O(L N_a \log N_a) = O(N_a L (\log L N_a))}$$

19. DATA STRUCTURES

19.1. **Lists (arrays), Python Lists.** cf. <https://wiki.python.org/moin/TimeComplexity>, from 4. Quiz: Python Lists, Lesson 2: List-Based Collections, Udacity, [Data Structures and Algorithms in Python](#)

Operation	Average	Amoritized Worst Case
Copy	$O(N)$	$O(N)$
Append(1)	$O(1)$	$O(1)$
pop last	$O(1)$	$O(1)$
insert	$O(N)$	$O(N)$
Get item	$O(1)$	$O(1)$
Set item	$O(1)$	$O(1)$
Delete item	$O(N)$	$O(N)$
Get slice	$O(k)$	$O(k)$
<i>Get length</i>	$O(1)$	$O(1)$

Insert and delete is $O(N)$. Need to move every element with index greater than k , and reindex each element.
Get array length is $O(1)$. cf. <https://stackoverflow.com/questions/21614298/what-is-the-runtime-of-array-length>

19.2. **Linked Lists, $O(1)$ insertion.**

19.2.1. *Linked Lists vs. arrays.* main difference: Linked lists and arrays store difference information in each element.

Both cases: each element stores a value.
Both cases: stores 1 more information.

Array: index as a number.
Linked List: reference to next element, e.g. store addresses of next element, e.g. x_{0123}
e.g.

x_{0122}	x_{0122}
value : 8	value : 8
next : x_{0123}	next : nullptr

Trick to remember: Given elements at addresses 1, 2, 3. If you delete the next reference for 1 (to 2), and replace it with a new object, you’ll lose your reference to 3.
Assign your next reference to object at 2 before assigning next reference to 1.
Insertion takes constant time since you’re just shifting around (constant, finite number of) pointers, instead of iterating every element of the list.

Doubly linked list: Also have pointers to previous element, e.g.

1	2	3
value:8	value:2	value:6
next:2	next:3	next:4
prev:0	prev:1	prev:2

cf. https://en.wikipedia.org/wiki/Linked_list

Disadvantages of Linked Lists:

- Use more memory than arrays because storage used by their pointers
- Nodes in linked list must be read in order from beginning: inherently sequential access
- nodes stored in continguously (possibly)
- difficult to reverse traverse, doubly linked list helps, but memory consumed in allocating space for back-ptr

Linked lists vs. dynamic arrays:
List data structures comparison

	Linked list	Array	Dynamic Array	Balanced Tree	Hashed array tree
indexing	$O(N)$	$O(1)$	$O(1)$	$O(\log N)$	$O(1)$
insert/delete at beginning	$O(1)$		$O(N)$	$O(\log N)$	$O(N)$
insert/delete at end	$O(1)$ when last element is known $O(N)$ when last element is unknown		$O(1)$ amoritized	$O(\log N)$	$O(N)$
insert/delete in middle	search time $+O(1)$		$O(N)$	$O(\log N)$	$O(N)$
wasted space (average)	$O(N)$	0	$O(N)$	$O(\log N)$	$O(\sqrt{N})$

head of list is 1st. node.
tail of 1st. last node in 1st. (or rest of list)

19.3. **Stack**, $O(1)$ **pop**, $O(1)$ **push**. [”Stack Details”, Udacity, Data Structures and Algorithms](#)
top element \mapsto Linked List head
push top element, pop top element. Last in, first out.
Last element you put in (new head), is the First out (when you pop).

19.4. **Queue**, $O(1)$ **insert**, $O(1)$ **delete**, $O(N)$ **search**, $O(N)$ **space complexities**. [https://en.wikipedia.org/wiki/Queue_\(abstract_data_type\)](https://en.wikipedia.org/wiki/Queue_(abstract_data_type)) [Lesson 2: List-Based Collections, 11. Queues](#)
Oldest element (tail) comes out first.
First In, First Out.
first element you put in (tail) is the first out (when you pop).

19.4.1. *Queue as Linked List.* head. ”oldest element in the queue”, ”first”
tail. ”newest element in queue”, ”last”,
(I guess it grows from the tail)
add element to tail, enqueue.
Dequeue, remove head.
peek - peek at head.

Save references to head and tail.

19.4.2. *Deque. double-ended queue.* One can enqueue and dequeue from either end.

19.4.3. *Priority Queue.* . Assign each element in queue with a priority.
Remove oldest and highest priority element first.

19.5. **STL Containers.** cf. Ch. 31, STL Containers, Bjarne Stroustrup (2013) [9]
pp. 886 Stroustrup (2013) [9] `std::vector<T, A>` contiguous allocated sequence of Ts;
cf. pp. 888 Sec. 31.2.1 ”Container Representation”, [9]
vector element data structure is most likely an array:
vector: `rep` \leftarrow elements — free space
list : likely represented by sequence of links pointing to elements and number of elements; doubly-linked list of T; use when you need to insert and delete elements without moving existing elements.
map : likely implemented as (balanced) tree of nodes pointing to (key, value) pairs:
unordered_map likely implemented as hash table
unordered_map `rep` \leftarrow hash table `\leftarrow , \leftarrow , ... (k, v), (k, v) , ...`
string for short strings, characters are stored in the string handle itself, for longer strings elements are stored continguously on free-store (like **vector** elements). Like **vector**, **string** can grow into ”free space” allocated to avoid repeated reallocations.
string: `rep` \leftarrow characters – free space
Like a built-in array, **array** is simply sequence of elements, with no handle, **array**: elements
cf. pp. 894, Sec. 31.3, Operations Overview, Stroustrup (2013) [9]
Standard Container Operation Complexity

	[] Sec. 31.2.2	List Sec. 31.3.7	Front Sec. 31.4.2	Back Sec. 31.3.6	Iterators Sec. 33.1.2
vector	const	$O(n)+$		const +	Ran
list		const	const	const	Bi
forward_list		const	const		For
deque	const	$O(N)$	const	const	Ran
stack				const	
queue			const	const	
priority_queue			$O(\log (n))$	$O(\log (n))$	
map	$O(\log (n))$	$O(\log (n))+$			Bi
set		$O(\log (n))+$			Bi
unordered_map	const+	const+			For
unordered_set		const+			For
string	const	$O(n)+$	$O(n)+$	const+	Ran
array	const				Ran

Ran - random-access iterator, "For" - "forward iterator", "Bi" - "bidirectional iterator"

19.6. **vector.**

19.6.1. *vector and growth.* Layout of **vector** : elem \leftarrow "front" of elements,
space \leftarrow "front" of extra space (meets end of elements)
last \leftarrow end of extra space
alloc
Use of both size (number of elements), and capacity (number of available slots for elements without reallocation) makes growth through **push_back()** reasonably efficient: there's not an allocation operation each time we add an element, only every time we exceed capacity (Sec. 13.6)
adding half the size is common, standard doesn't specify by how much capacity is increased

19.6.2. *vector and Nesting.* **vector** (and similarly contiguously allocated data structures) has 3 major advantages:

- elements of **vector** compactly stored; no per-element memory overhead (contiguous on memory address)
amount of memory consumed by **vec** of type **vector<X>** roughly
`sizeof(vector<X>)+vec.size()*sizeof(X); sizeof(vector<X>)` is about 12 bytes, insignificant for large vectors
- fast traversal, consecutive access, to get to next element, code doesn't have to indirect through a pointer
- simple and efficient random access, makes sort and binary search efficient

vs. doubly-linked list, **list**, incurs 4-words-per-element memory overhead (2 links plus free-store allocation header)
be careful don't unintentional compromise efficiency of access, e.g. 2-dim. matrix
don't do **vector<vector<double>>**,
do **vector<double>** and compute locations from the indices

19.6.3. *vector vs. array.* **vector** resource handle, i.e. allows it to be resized and enable efficient move semantics,
disadvantage to arrays that don't rely on storing elements separately from handle; keeps sequence of elements on stack or in another object.

19.7. **Trees.** value, left pointer, right pointer,
levels - how many connections it takes to reach the root +1 child can only have 1 parent height - number of edges between it and farthest leaf
depth - number of edges to root.
height, depth are inverse

19.8. **Graphs.** edges can store data too.
Directed graph edges have a sense of direction.
 A = set of ordered pairs of vertices.
undirected graph.
acyclic (no cycles)
DAG Directed graph with no cycles.

19.8.1. *Connectivity.* connected graph has only 1 connected component. \exists path \forall pair of vertices.
minimum number of elements (edges) to remove, to disconnect a component.

weakly connected directed graph if replacing all its directed edges with undirected edges produces a connected (undirected) graph.

19.8.2. *Graph Representation.* vertex object:
list of edges.

Edge Object
vertices.
Edge List: $= E$
Adjacency list $l = l(i)$ s.t. $\forall i \in V, l(i) \in \mathbf{Set}$ s.t. $l(i)$ = set of all adjacent vertices.

Adjacent Matrix.
Let V = set of all vertices.
 $\forall v \in V$, label them: $v = v(i), i \in \mathbb{N}$

$\forall v_i \in V$, so $\forall i = 0, 1, \dots N - 1$,
Consider $\forall w = w_j \in V$, so $\forall j = 0, 1, \dots N - 1$.

If v_i, w_j are adjacent, let $a(i, j) = 1$, otherwise $w(i, j) = 0$. (adjacent means \exists edge s.t. $E = \{v_i, w_j\}$)
 $\implies a$ is an adjacency matrix.
Adjacency list.
 $\forall v = v_i \in V, i = 0, 1, \dots N - 1$
 $E = \{(v_o, v_t)|v_o, v_t \in V\}$
if $\exists e \in E$ s.t. $v_i = e(0) = v_o$, then $e(1) = v_t$ is adjacent to it.
Adjacency list useful for counting number of edges that a node has or number of adjacencies.

19.8.3. *Graph Traversal.* Depth-first search (DFS)
Breadth-first search (BFS)

DFS.
implementation: stack.
keep 2 structures: 1. Seen list, 2. stack.
If seen before, pop stack and go back.
 $O(|E| + |V|)$ visit every edge twice $O(2|E| + |V|) = O(|E| + |V|)$.
 $O(|V|)$ time to look up a vertex.
More on procedure:
begin with any node v_1 . Put $v(v_1)$ into seen. Put v_1 on stack.
if $\exists \{e\} \subset E$ s.t. for $e = (v_o, v_t)$, $v_o = v_1$, then, put $v(v_t)$ into seen. Put v_t on stack.
Then consider v_t .
If $v(v_t) \in$ seen, consider another edge.
If you run out of edges with new node, pop stack.
Eulerian path $O(|E|)$

Hamiltonian path

20. SEARCH

20.1. **Binary Search** $O(\log(N))$. <https://classroom.udacity.com/courses/ud513/lessons/7123524086/concepts/71154040750923>

$O(\log(N) + 1) = O(\log(N))$ Binary search efficiency.
Given array $a = a_i, i = 0, 1, \dots N - 1, N$ elements in an array, s.t. $a(i) \leq a(j)$ if $i \leq j, i, j \in 0, 1 \dots N - 1$ (i.e. sorted array)
Given x to search for,
Consider m_j s.t.

For $j = 1$, given $N, m_1 := \begin{cases} \frac{N}{2} & \text{if } N \text{ odd} \\ \frac{N}{2} - 1 & \text{if } N \text{ even} \end{cases}$

 m_1 = midpoint to compare against.

if $x = a(m_1)$ done.
if $x < a(m_1)$, consider a_i s.t. $i = 0, \dots m_1 - 1$
if $x > a(m_1)$, consider a_i s.t. $i = m_1 + 1 \dots N - 1$

For j , given $a_i, i = l, \dots r, l \leq r, (l, r$ are included in the range),
Let $L_j := r - l + 1$

$$m_j := \begin{cases} \frac{L_j}{2} + l & \text{if } L_j \text{ odd} \\ \frac{L_j}{2} - 1 + l & \text{if } L_j \text{ even} \end{cases}$$

If $x = a(m_j)$ done,
if $x < a(m_j)$, consider a_i s.t. $i = 0, \dots m_j - 1$
If $x > a(m_j)$, consider a_i s.t. $i = m_j + 1, \dots N - 1$
Stop when $L_j = 0$.

20.2. **Breadth-first search (BFS)**. https://en.wikipedia.org/wiki/Breadth-first_search

starts at tree root, or some arbitrary node, and explores all neighbor nodes at present depth prior to moving on to nodes at next depth level.
Worst-cast performance $O(|V| + |E|) = O(b^d)$, Worst-case space complexity $O(|V|) = O(b^d)$
<https://www.quora.com/What-are-the-advantages-of-using-BFS-over-DFS-or-using-DFS-over-BFS-What-are-the-applications-and-downsides-of-each>
Pro:
1. Solution definitely found out by BFS, 2. never get trapped in blind alley, unwanted nodes, 3. if there are more than 1 solution, will find solution with shortest steps
Con:
1. Memory constraints: as it stores all nodes of present level to go for next level 2. if solution far away, consumes time
Application of BFS:
1. Find shortest Path, 2. Check graph with bipertiteness

20.3. **Depth-first search**. Pre-order - check off nodes as soon as you see it, before seeing children.
root check it off, pick first left child, ...

In-order.
check off node once left child is seen.
went left most to right, went through all nodes in order.

Post-order.
Check off leaf, don't check off parent, check right child.

Delete $O(N)$. insert.
When do we want to use these structures? pros and cons.

20.4. **Binary Search Tree (BST)**, $O(\log N)$ **height of the tree, run-time complexity, worst case $O(N)$** . Given node x , functions l, r, v s.t. $l(x), r(x) \in \mathbf{Nodes} \cup \emptyset, v(x)$ is some ordered value, so $v(x) \in \mathbf{OrderedSet}$.
Binary Search tree has the condition that $\forall x$,

$$v(l(x)) \leq v(x) \leq v(r(x))$$

Given search value y ,
if $v(x) = y$, done,
if $v(x) < y, x_1 = r(x)$
if $v(x) > y, x_1 = l(x)$

height of tree is $O(\log N)$, run-time complexity.
insert $O(\log(N))$
delete $O(\log(N))$

Unbalanced, distribution of nodes skewed, worse case $O(N)$ search, insert, delete e.g. $5 \rightarrow 10 \rightarrow 15 \rightarrow 20$

20.5. **Heaps**. max-heap: parent must always have bigger value than its child.
min-heap: parent must always have smaller value than its children.
Unlike binarytree, heap can have any number of children.

Pro:
1. heap finds max, min (root) 2. heap data structure efficiently use graph algorithm, e.g. Dijkstra
Con:
1. takes more time to compare and execute.
heap = max. efficient implementation of priority queue.

Heap	find min	delete min	insert	decrease-value
Binary Heap	$O(1)$	$O(\log N)$	$O(\log N)$	$O(\log N)$

e.g. max binary heap.
"complete" all levels except last are completely full.
if not, continue adding values from left to right.

Search: $O(N)$ no guarantee child is \leq . end up searching entire tree.
improved search: if $x > \text{root}$, quit search (for max-heap)
In general, if $v(\text{node}) = \text{node value} < x$, no need to check children of node.

Search: worse cast: $O(N)$
Average case: $O(N/2) = O(N)$

20.5.1. *Heapify (Worst case $O(\log N)$)*. stick new element in open spot of the tree.
heapify - reorder tree based on heap property: \forall given node C , if parent node P of $C, v(P) \geq v(C)$ (max heap)

Swap P and C when $v(C) > v(P)$
Extract; remove root. replace root with right most element. Then swap when necessary.

Heapify: worst case $O(\log N)$.
As many operations as height of the tree.

20.5.2. *Heap implementation (array)*. max-binary heap: since we know how many children (2), \forall parent, Use some math to find next node. If level l , nodes per level $l = 2^{l-1}$
Sorted array into tree.
Let sorted array be $a = a_i$ s.t.

$$\begin{aligned} a_0 &\mapsto l = 1 \\ a_1, a_2 &\mapsto l = 2 \\ a_3, a_4, a_5, a_6 &\mapsto l = 3 \end{aligned}$$

$\left(\sum_{m=1}^{l-1} 2^{m-1}\right) - 1 = k$ index to start from.
Tree vs. Array
if tree uses nodes, need pointers: Arrays save space.

20.6. **Self-balancing tree**. Balanced minimize number of levels to use

20.7. **Red-Black Tree; search** $O(\log N)$, **space** $O(N)$, **insert** $O(\log N)$, **delete** $O(\log N)$. root black,
level 1 black
level 2 red
level 3 black

null leaf nodes must be colored black (i.e. all leaves are black)
Rule 4 (optional) root must be black.
Rule 5 every node; every path descending down must contain same number of black nodes. i.e. \forall path from given node, to any of its descendant NIL nodes (leaf) must contain same number of black nodes.
red-black tree = self balancing binary search tree (\forall node x , $v(l(x)) \leq v(x) \leq v(r(x))$)

20.7.1. *Red-Black Trees, insertion*. Insert Red Nodes Only.
runtime for insertion worst case $O(\log N)$. Binary search tree worse case $O(N)$ because BST could be unbalanced.

21. RECURSION

cf. <https://en.wikipedia.org/wiki/Recursion>, Udacity
Recursion

- must call itself at some pt.
- base case
- alter input parameter

Theorem 1 (Recursion (from set theory)). *Given* $X \in \mathbf{Set}$, $a \in X$, $f : X \rightarrow X$
 $\exists ! F : \mathbb{N} \rightarrow X$ s.t.
 $F(0) = a$
 $F(n + 1) = f(F(n))$

e.g. $Fib(0) = 0$
 $Fib(1) = 1$
 $\forall n > 1, n \in \mathbb{Z}, Fib(n) := Fib(n - 1) + Fib(n - 2) = f(Fib(n - 1), Fib(n - 2))$

22. SORTING

22.1. **Bubble Sort, $O(N^2)$ in time, $O(1)$ in space**. cf. [Efficiency of Bubble Sort, Lesson 3: Searching and Sorting, Data Structures and Algorithms, Python](#)
Time complexity of bubble sort:
 \forall iteration, there were $N - 1$ comparisons. Worst case is $N - 1$ iterations to order all N elements.
 $\implies O(N^2)$ worse case
 $O(N^2)$ average case
 $O(N)$ best case (it was already sorted!)

Space complexity $O(1)$ no extra arrays needed. Sort was "in-place."
From wikipedia, "Bubble Sort" (Optimizing Bubble Sort), cf. https://en.wikipedia.org/wiki/Bubble_sort, Observe that for an array of elements a indexed by $i = 1, 2 \dots N$,
 $j = 1 \mapsto a(N)$ largest (otherwise it would not have been swapped; contradiction)

 $j = 2 \mapsto a(N - 1)$ 2nd. largest. Iterated on $2, \dots N - 1$ $N - 2$ total (don't include the first). \vdots
 $j \mapsto a(N - j + 1)$ j th largest, iterated on $2, \dots N - j + 1$, $N - j$ total.

22.2. **Merge Sort, $O(N \log N)$ in time, $O(N)$ in space**. cf. [Efficiency of Merge Sort, Lesson 3: Searching and Sorting, Data Structures and Algorithms, Python](#)
Use approximation (approximate to the *worse case*!) to count the number of comparisons (operations) in a "pass" (iteration) multiply each iteration with number of comparisons per iteration.
Number of iterations $\cong \log N$ since $2^J - 1 =$ array size.
 N comparisons for $\log(N)$ steps $\implies O(N \log N)$
Space complexity. Auxilliary space = $O(N)$ at each step we need total arraying size N to copy into.

22.3. **Quick Sort, $O(N \log N)$ average time complexity, $O(N^2)$ worst time complexity, $O(1)$ in space (if inplace)**.

22.3.1. *First element pivot Implementation*. cf. <https://stackoverflow.com/questions/22504837/how-to-implement-quick-sort->
Consider array $a = a_i$, $i = 0, 1, \dots N - 1$; $|a| = N$ (array of length N).
Given $l, r \in 0, 1, \dots N - 1$,
Let $p := a(l)$ (pivot value)
 $i = i_i$; $i_1 = l$
Let $j = j_j$ s.t. $j_1 = l + 1, \dots, j_{r-l-1} = r - 1$ (range over $r - 1 - (l + 1) + 1 = r - l - 1$)
if $a(j_j) \leq p$
 $i_{j+1} = i_j + 1$
 $a(i_{j+1}) = a(i_j + 1) \leftrightarrow a(j_j)$.

 $a(i_{r-l-1}) \leftrightarrow a(l)$
 $\mapsto i_{r-l-1}$.
Let's work out a table for the first few steps:

iteration	i	j	swaps	condition	example
1	l	$l + 1$			
2	$l + 1$	$l + 1$	$a(l + 1) \leftrightarrow a(l + 1)$	$p \geq a(l + 1)$	5 3 4
	l	$l + 2$			
	$l + 1$	$l + 2$			
	$l + 1$	$l + 2$	$a(l + 1) \leftrightarrow a(l + 2)$	$p \geq a(l + 2)$	5 6 3
	$l + 2$	$l + 2$	$a(l + 2) \leftrightarrow a(l + 2)$	$p \geq a(l + 2)$	5 3 4
3	$l + 1$	$l + 3$	$a(l + 1) \leftrightarrow a(l + 3)$	$p \geq a(l + 3)$	5 8 6 3 \mapsto 5 3 6 8

25. UNIX EXECUTION MODEL

cf. <https://stackoverflow.com/questions/4204915/please-explain-the-exec-function-and-its-family/37558902>

25.1. **Processes and Programs.** A process is something in which a program executes. $pid \in \mathbb{N} \xrightarrow{\text{surj}}$ process

These 2 operations that you can do is the entire UNIX execution model:

fork - creates new process containing duplicate of current program, including its state. There are differences between the processes so to be able to distinguish which is parent, which is child:

$$pid_P \xrightarrow{\text{surj}} pid_C > 0$$

$$\text{prog}_i \text{ s.t. } \text{prog}_i \in pid_P \mapsto \text{prog}_i \in pid_C$$

exec - replaces program in current process with brand new program:

$$\text{prog}_i \in pid_i \mapsto \text{prog}_{i'} \in pid_i$$

$$\text{s.t. } \text{prog}_i = \emptyset, \text{prog}_i \neq \text{prog}_{i'}$$

More details:

fork() makes a near duplicate of current process, identical in almost every way; 1 process calls **fork()** while 2 processes return from it:

$$pid_P \xrightarrow{\text{fork}()} pid_P, pid_C > 0$$

Examples:

- When **fork**, **exec** used in sequence: In shells, **find**:

$$pid_{\text{shell}} \xrightarrow{\text{fork}()} pid_{\text{shell}}, pid_C \quad \text{prog}_j, pid_C \xrightarrow{\text{exec}()} \text{prog}_{j'} \in pid_C$$

Shell forks, then child loads **find** program into memory, setting up all command line arguments, standard I/O, etc.

- **fork()** only. Daemons simply listening on TCP port and fork copy of themselves to process specific request, while parent goes back to listening.

Notice that nominally, parent waits for child to exit.

close-on-exec flag, if set, will have fd automatically be closed during a successful **execve** (or **exec**) (cf. <http://man7.org/linux/man-pages/man2/fcntl.2.html>)

25.2. **File entry.** Every fd points to a data structure called *file entry* in the kernel.

$$fd \mapsto \text{file entry}$$

i.e. $\forall fd, \exists$ file entry and pointer to that file entry.

open file description - an entry in system-wide table of open files.

open file description records file offset and file status flags.

$n_{\text{offset}} \equiv$ file offset, per fd, from the beginning of file entry.

fork() results in both parent and child processes using same fd and reference to same offset n_{offset} in file entry.

If process was last to reference file entry, kernel then deallocates that file entry.

Each file entry contains

- the type
- array of function pointers, that translates generic operations on fds to file-type specific implementations.

open call implementation greatly varies for different file types, even if higher level API exposed is the same.

e.g. sockets - Same **read**, **write** used for byte-stream type connections, but different system calls handle addressed messages like network datagrams.

For a process pid, consider fd table of fds. \forall fd, $fd \mapsto$ open file description \in open file table (system-wide).

\forall open file description \in open file table,

open file description = n_{offset} , status flags, and

open file description $\xrightarrow{\text{inode ptr}}$ inode table.

A blocking system call is 1 that suspends or puts calling process on wait until an event occurs ¹

25.3. **Readiness of fds.** cf. <https://medium.com/@copyconstruct/nonblocking-i-o-99948ad7c957>

A fd is considered *ready* if process can perform I/O operation on fd without blocking (i.e. without process having to wait or be suspended). For fd to be considered "ready", it doesn't matter if **the operation actually transfer any data** - all that matters is that the I/O operation can be performed without blocking (i.e. the calling process won't wait or be suspended).

A fd changes into a *ready* state when an I/O *event* happens, such as the arrival of new input or completion of a socket connection or when space is available on previously full socket send buffer after TCP transmits queued data to the socket peer.

There are 2 ways to find out about the readiness status of a descriptor - edge triggered and level-triggered.

25.4. **Level Triggered.** cf. <https://medium.com/@copyconstruct/nonblocking-i-o-99948ad7c957>

To determine if fd is ready, the process tries to perform a non-blocking I/O operation. Also seen as "pull" or "poll" model. (User "pull" on the fd).

At time t_0 , process pid tries I/O operation on non-blocking fd. If I/O operation blocks, system call returns error **err**.

At time $t_1 > t_0$, pid tries I/O on fd (again). Say call blocks again, and returns **err**.

At time $t_2 > t_1$, pid tries I/O on fd (again). Assume call blocks again, and returns **err**.

At time $t_3 > t_2$, pid polls status of fd, and fd is ready. pid can then choose to actually perform entire I/O operation (e.g. read all data available on socket).

At time $t_4 > t_3$, assume pid polls status of fd, and fd isn't ready; call blocks (again), and I/O operation returns **err**.

At time $t_5 > t_4$, pid polls for status of fd, and fd is ready. pid can then choose to only perform a partial I/O operation (e.g., reading only half of all data available)

At time $t_6 > t_5$, pid polls status of fd, and fd is ready. This time, pid can choose to perform no I/O.

25.5. **Edge Triggered.** Process *pid* receives a notification only when fd is "ready" (usually when there's any new activity on fd since it was last monitored). This can be seen as the "push" model, in that a notification is pushed to the process about readiness of a fd.

process pid only notified that fd is ready for I/O, but not provided additional information like for instance how many bytes arrived on socket buffer.

Thus, pid only armed with incomplete data as it tries to perform any subsequent I/O operation e.g. at $t_2 > t_0$, pid gets notification about fd being ready.

byte stream $(b_n)_{n \in \mathbb{N}}$ available for I/O stored in buffer.

Assume $N_b = 1024$ bytes available for reading, when pid gets notification at $t = t_2$.

$$\implies (b_n)_{n < N_b} \subset (b_n)_{n \in \mathbb{N}}$$

Assume pid only reads $N'_b = 500$ bytes $< N_b$

That means for $t = t_3, t_4, \dots t_{i_2}$, there are still $N_b - N'_b = 524$ bytes available in $(b_n)_{n < N_b}$ that pid can read without blocking.

But since pid can only perform I/O once it gets next notification, these $N_b - N'_b$ bytes remain sitting in buffer for that duration.

Assume pid gets next notification at $t_{i_2} = t_6$, when additional N_b bytes have arrived in buffer.

Then total amount of data available on $(b_n)_{n \in \mathbb{N}} = 2N_b - N'_b = 1548 = 524 + 1024$.

Assume pid reads in $N_b = 1024$ bytes.

This means that at end of 2nd. I/O operation, $N_b - N'_b = 524$ bytes still remain in $(b_n)_{n \in \mathbb{N}}$ that pid won't read before next notification arrives.

While it might be tempting to perform all I/O immediately once notification arrives, doing so has consequences.

- large I/O operation on single fd has potential to starve other fd's
- Furthermore, even with case of level-triggered notifications, an extremely large Write or send call has potential to block.

¹<https://stackoverflow.com/questions/19309136/what-is-meant-by-blocking-system-call>, <https://www.quora.com/What-is-the-difference-between-a-blocking-system-call-and-an-non-blocking-system-call>

26. MULTIPLEXING I/O ON FDS

In above, we only described how pid handles I/O on a single fd.
Often, pid wants to handle I/O on more than 1 fd.
e.g. program prog needs to log to `stdout`, `stderr`, while accept socket connections, and make outgoing RPC connections to other services.

- There are several ways of multiplexing I/O on fds.
- nonblocking I/O (fd itself is marked as non-blocking; operations may finish partially)
 - signal-drive I/O (pid owning fd is notified when I/O state of fd changes)
 - polling I/O (e.g. `select`, `poll`, both provide level-triggered notifications about readiness of fd)

26.1. **Multiplexing I/O with Non-blocking I/O.** We could put all fds in non-blocking mode.
pid can try to perform I/O operations on fds to check if any I/O operations result in error.
kernel performs I/O operation on fd and returns err, or partial output, or result of I/O operation if it succeeds.
Cons:

- Frequent checks
- Infrequent checks. If such operations are conducted infrequently, then it might take pid unacceptably long time to respond to I/O event that’s available.

When does this approach make sense?
Operations on output fds (e.g. writes) don’t generally block.
- In such cases, it might help to try to perform I/O operation first, and revert back to polling when operation returns err.
It might also make sense to use non-blocking approach with edge-triggered notifications, where fds can be put in nonblocking mode, and once pid gets notified of I/O event, it can repeatedly try I/O operations until system calls would block with an EAGAIN or EWOULDBLOCK.

26.2. **Multiplexing I/O via signal driven I/O.** .
Kernel instructed to send the pid a signal when I/O can be performed on any of the fds.
pid will wait for signals to be delivered when an of the fds is ready for I/O operation.
Cons: Signals are expensive to catch, rendering signal drive I/O impractical for cases where large amount of I/O is performed.
Typically used for ”exceptional conditions.”

26.3. **Multiplexing I/O via polling I/O.** .
fds placed in non-blocking mode.
pid uses **level triggered mechanism** to ask kernel by means of system call (`select` or `poll`) which fds are capable of performing I/O.

26.3.1. *poll, and select vs. poll.* With `select`, we pass in 3 sets of fds we want to monitor for reads, writes, and exceptional cases.
With `poll`, we pass in a set of fds, **each marked with the events it specifically needs to track.**

26.3.2. *What happens in the kernel (with select and poll)?* Both `select` and `poll` are *stateless*, meaning, every time a `select` or `poll` system call is made, the kernel checks *every fd*, in the input array passed as 1st. argument, for the occurrence of an event and return the result to the pid.
⇒ this means that the cost of `select/poll` is O(N), where N is the number of fds monitored.
The implementation of `select` and `poll` comprises 2 tiers:

- Specific top tier which decodes incoming request, as well as several device or socket specific bottom layers
- bottom layers comprise of *kernel poll functions* used both by `select`, `poll`

27. EPOLL EVENT POLL

cf. [”The method to epoll’s madness” by Cindy Sridharan](#)^[14]
`epoll`, a Linux specific construct, allows for a pid to monitor multiple fds and get notifications when I/O is possible on them.

27.1. **epoll Syntax.** *epoll*, unlike *poll*, is not a system call. It’s a kernel data structure that allows a pid to multiplex I/O on multiple fds.
This data structure can be created, modified, deleted by 3 system calls: `epoll_create`, `epoll_ctl`, `epoll_wait`.

28. REFERENCES AND RESOURCES FOR LINUX SYSTEM PROGRAMMING

<http://igm.univ-mlv.fr/~yahya/progsys/linux.pdf>
http://www.cs.fsu.edu/~baker/opsys/examples/forkexec/fork_wait.c
<http://cs241.cs.illinois.edu/coursebook/> <https://www.csd.uoc.gr/~hy556/material/tutorials/cs556-3rd-tutorial.pdf> <https://www.cs.rutgers.edu/~pxk/417/notes/content/02r-sockets-programming-slides.pdf> <https://www.cs.cmu.edu/~srini/15-441/S10/lectures/r01-sockets.pdf>

Part 7. C++ Template Metaprogramming

From this question on [stack overflow](#), comes this [tutorial](#).
cf. Vandevoorde, Josuttis, and Gregor (2017) ^[15].

Part 8. Functional Programming and Category Theory

Best discussion of the relation to monads in category theory to monads in ”computer science”: <https://ncatlab.org/nlab/show/monad+%28in+computer+science%29>
Best article on practical implementation (real daily programming, real programming experiences) of functional programming without pulling punches on Category Theory:
<https://nalaginrut.com/archives/2019/10/31/8%20essential%20patterns%20you%20should%20know%20about%20functional%20programming%20in%20c%2B%2B14>
More links:
<https://nbviewer.jupyter.org/github/ivanmurashko/articles/blob/master/cattheory/cattheory.pdf> https://wiki.ifs.hsr.ch/SemProgAnTr/files/mof_categories_final.pdf

29. ACTORS, CONCURRENT SYSTEMS

<https://sourceforge.net/projects/subjectizer/> <https://github.com/Stiffstream/subjectizer#helloworld-example>

Part 9. Technical Interviews

[”Interview Introduction”](#), [Data Structures and Algorithms in Python](#), [Udacity](#)

- Clarifying the Question
- Generating Inputs and Outputs
- Generating Test Cases
- Brainstorming
- Runtime Analysis
- Coding
- Debugging

Part 10. Embedded Systems

Lee and Seshia (2016) [16].

cf. Ch. 3 Discrete Dynamics, Lee and Seshia (2016) [16].

showed how state machines can be used to model discrete dynamics.

cf. Sec. 3.1., Lee and Seshia (2016) [16].

pp. 43, Ex. 3.1., Lee and Seshia (2016) [16]. e.g. Consider system that counts number of cars that enter and leave parking garage in order to keep track of how many cars are in garage at any time.

Ignore for now how to design sensors that detect entry or departure of cars.

Assume **ArrivedDetector** actor produces an event when car arrives, and **DepartureDetector** actor produces an event when car departs.

Counter actor keeps running count, starting from initial value i .

Each time count changes, it produces an output event that updates display.

Each entry or departure modeled as discrete event. Discrete event **occurs at** an instant of time rather than over time.

Signal u into **up** input port (from **ArrivalDetector**, upon arrival) is function of form

$$u : \mathbb{R} \rightarrow \{\text{absent}, \text{present}\}, \text{ i.e.}$$

Definition 4 (Signal). *Signal is a function of the form $e : \mathbb{R} \rightarrow \{\text{absent} \cup X, X \in \text{ObjSet}\}$ s.t.*

\exists 1-to-1 $f : T \rightarrow \mathbb{N}$, s.t. f order preserving, i.e. $\forall t_1, t_2 \in T$, if $t_1 \leq t_2$, then $f(t_1) \leq f(t_2)$, where $T = T_{\text{present}} = \{t \in \mathbb{R} | e(t) \neq \text{absent}\}$.

REFERENCES

[1] Bertrand Meyer. **Object-Oriented Software Construction** (Book/CD-ROM) (2nd Edition) 2nd Edition. Prentice Hall; 2 edition (April 13, 1997). ISBN-13: 978-0136291558

[2] Randall Hyde. **Write Great Code: Volume 1: Understanding the Machine** October 25, 2004. No Starch Press; 1st edition (October 25, 2004). ISBN-13: 978-1593270032

[3] Randall Hyde. **Write Great Code, Volume 2: Thinking Low-Level, Writing High-Level**. 1st Edition. No Starch Press; 1 edition (March 18, 2006). ISBN-13: 978-1593270650

[4] Zed A. Shaw. **Learn C the Hard Way: Practical Exercises on the Computational Subjects You Keep Avoiding (Like C)** (Zed Shaw’s Hard Way Series) 1st Edition. Addison-Wesley Professional; 1 edition (September 14, 2015) ISBN-13: 978-0321884923.

[5] Eli Bendersky. [Are pointers and arrays equivalent in C?](#)

[6] Brian W. Kernighan, Dennis M. Ritchie. **C Programming Language**, 2nd Ed. 1988.

[7] Peter van der Linden. **Expert C Programming: Deep C Secrets** 1st Edition. Prentice Hall; 1st edition (June 24, 1994) ISBN-13: 978-0131774292

[8] Leo Ferres. "Memory management in C: The heap and the stack". [Memory management in C: The heap and the stack](#)

[9] Bjarne Stroustrup. **The C++ Programming Language**, 4th Edition. Addison-Wesley Professional; 4 edition (May 19, 2013). ISBN-13: 978-0321563842

[10] Carlo Ghezzi, Mehdi Jazayeri. **Programming Language Concepts** 3rd Edition. Wiley; 3 edition (June 23, 1997). ISBN-13: 978-0471104261 [https://vowi.fsinf.at/images/7/72/TU_Wien-Programmiersprachen_VL_\(Puntigam\)_-_E-Book_SS08.pdf](https://vowi.fsinf.at/images/7/72/TU_Wien-Programmiersprachen_VL_(Puntigam)_-_E-Book_SS08.pdf)

[11] Stanley B. Lippman, Josée Lajoie, Barbara E. Moo. **C++ Primer** (5th Edition). Addison-Wesley Professional; 5 edition (August 16, 2012) ISBN-13: 978-0321714114

[12] Scott Meyers. **Effective Modern C++: 42 Specific Ways to Improve Your Use of C++11 and C++14**. 1st Edition. O’Reilly Media; 1 edition (December 5, 2014) ISBN-13: 978-1491903995

[13] Prof. Donald S. Fussell (Instructor). CS429H (378H) - Systems I (Honors) (Computer Organization, Architecture, and Programming). Spring 2011. University of Texas. <https://www.cs.utexas.edu/users/fussell/courses/cs429h/>

[14] Cindy Sridharan. "The method to epoll’s madness." *Medium*. Oct. 29, 2017. <https://medium.com/@copyconstruct/the-method-to-epolls-madness-d9d2d6378642>

[15] David Vandevoorde, Nicolai M. Josuttis, Douglas Gregor. **C++ Templates: The Complete Guide** (2nd Edition). Addison-Wesley Professional; 2 edition (September 18, 2017). ISBN-13: 978-0321714121

[16] Edward Ashford Lee, Sanjit Arunkumar Seshia. **Introduction to Embedded Systems: A Cyber-Physical Systems Approach** (The MIT Press) Second Edition. The MIT Press; Second edition (December 30, 2016). ISBN-10: 0262533812. ISBN-13: 978-0262533812