Topics of this Slideset

There are 10 kinds of people in the world: those who understand binary, and those who don’t!

- Why bits?
- Representing information as bits
  - Binary and hexadecimal
  - Byte representations: numbers, characters, strings, instructions
- Bit level manipulations
  - Boolean algebra
  - C constructs

Why Not Base 10?

Base 10 Number Representation.

- That’s why fingers are known as “digits.”
- Natural representation for financial transactions. Floating point number cannot exactly represent $1.20$.
- Even carries through in scientific notation: $1.5213 \times 10^4$

If we lived in Homer Simpson’s world, we’d all use octal!

Implementing Electronically

- 10 different values are hard to store. ENIAC (First electronic computer) used 10 vacuum tubes / digits
- They’re hard to transmit. Need high precision to encode 10 signal levels on single wire.
- Messy to implement digital logic functions: addition, multiplication, etc.
**Binary Representations**

**Base 2 Number Representation**
- Represent $15213_{10}$ as $11101101110112$
- Represent $1.20_{10}$ as $1.0011001100110011[0011]...2$
- Represent $1.5213 \times 10^4$ as $1.11011011011012 \times 2^{13}$

**Electronic Implementation**
- Easy to store with bistable elements.
- Reliably transmitted on noisy and inaccurate wires.

---

**Representing Data**

**Great Reality 6:** Whatever you plan to store on a computer ultimately has to be represented as a finite collection of bits.

That’s true whether it’s integers, reals, characters, strings, data structures, instructions, pictures, videos, etc.

---

**Some Representations**

Suppose you want to represent the finite set of natural numbers $[0 \ldots 7]$ as 3-bit strings.

<table>
<thead>
<tr>
<th>Dec.</th>
<th>Rep1</th>
<th>Rep2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>101</td>
<td>000</td>
</tr>
<tr>
<td>1</td>
<td>011</td>
<td>001</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>010</td>
</tr>
<tr>
<td>3</td>
<td>000</td>
<td>011</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>010</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>001</td>
<td>110</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>111</td>
</tr>
</tbody>
</table>

Can you see why one of these representations is “better” than the other?

To store data of type X, someone had to invent a mapping from items of type X to bit strings. That’s the *representation mapping*.

In a sense the representation is *arbitrary*. The representation is just *a mapping from the domain onto a finite set of bit strings*.

But some representations are better than others. **Why would that be?** Hint: what operations do you want to support?
You have to map X data onto bit strings (B) in a way that makes it as easy as possible to compute the operations on that data. I.e., the following diagram must *commute*.

```
X \xrightarrow{abs-op} X
\downarrow rep
B \xrightarrow{conc-op} B
\uparrow rep^{-1}
```

To carry out any operation at the C level means converting the data into bit strings, and implementing an operation on the bit strings that has the “intended effect.”

```
int x;
int y;
...
t = x + y;
```

**Important Fact:** If you are going to represent any type in \( k \) bits, you can only represent \( 2^k \) different values. *There are exactly as many ints as floats on x86.*

**Important Fact:** The same bit string can represent an integer (signed or unsigned), float, character string, list of instructions, address, etc. depending on the context.

**Conceptually, memory is a very large array of bytes.**

**Actually, it’s implemented with hierarchy of different memory types.**
- SRAM, DRAM, disk.
- Only allocate storage for regions actually used by program.

**In Unix and Windows, address space private to particular “process.”**
- Encapsulates the program being executed.
- Program can clobber its own data, but not that of others.
Byte-Oriented Memory Organization

Compiler and Run-Time System Control Allocation
- Where different program objects should be stored.
- Multiple storage mechanisms: static, stack, and heap.
- In any case, all allocation within single virtual address space.

Encoding Byte Values

** Byte = 8 bits
Which can be represented in various forms:
- Binary: 00000000_2 to 11111111_2
- Decimal: 0_{10} to 255_{10}
- Hexadecimal: 00_{16} to FF_{16}
  - Base 16 number representation
  - Use characters '0' to '9' and 'A' to 'F'
  - Write FA1D37B_{16} in C as 0xFA1D37B or 0xfA1d37b

BTW: one hexadecimal digit represents 4 bits and is called a *nybble*.

<table>
<thead>
<tr>
<th>Hex</th>
<th>Dec</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0001</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0010</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0011</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0100</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0101</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0110</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0111</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>1000</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>1001</td>
</tr>
<tr>
<td>A</td>
<td>10</td>
<td>1010</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>1011</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>1100</td>
</tr>
<tr>
<td>D</td>
<td>13</td>
<td>1101</td>
</tr>
<tr>
<td>E</td>
<td>14</td>
<td>1110</td>
</tr>
<tr>
<td>F</td>
<td>15</td>
<td>1111</td>
</tr>
</tbody>
</table>

Machine Words

Machines generally have a specific “word size.”
- It’s the nominal size of addresses on the machine.
- Most current machines run 64-bit software (8 bytes).
  - 32-bit software limits addresses to 4GB.
  - Becoming too small for memory-intensive applications.
- All x86 current hardware systems are 64 bits (8 bytes).
  - Potentially address around 1.8x10^{19} bytes.
- Machines support multiple data formats.
  - Fractions or multiples of word size.
  - Always integral number of bytes.
- X86-hardware systems operate in 16, 32, and 64 bits modes.
  - Initially starts in 286 mode, which is 16-bit.
  - Under programmer control, 32- and 64-bit modes are enabled.

Word-Oriented Memory Organization

Addresses Specify Byte Locations
- Which is the address of the first byte in word.
- Addresses of successive words differ by 4 (32-bit) or 8 (64-bit).

<table>
<thead>
<tr>
<th>32-bit words</th>
<th>64-bit words</th>
<th>bytes</th>
<th>addr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addr: 0000</td>
<td></td>
<td></td>
<td>0000</td>
</tr>
<tr>
<td>Addr: 00004</td>
<td></td>
<td></td>
<td>0001</td>
</tr>
<tr>
<td>Addr: 00008</td>
<td></td>
<td></td>
<td>0002</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0003</td>
</tr>
<tr>
<td>Addr: 0012</td>
<td></td>
<td></td>
<td>0004</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0005</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0006</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0007</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0008</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0009</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0010</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0011</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0012</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0013</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0014</td>
</tr>
<tr>
<td>Addr: 0008</td>
<td></td>
<td></td>
<td>0015</td>
</tr>
</tbody>
</table>
Data Representations

Sizes of C Objects (in Bytes)

<table>
<thead>
<tr>
<th>C Data Type</th>
<th>Alpha</th>
<th>Intel x86</th>
<th>AMD 64</th>
</tr>
</thead>
<tbody>
<tr>
<td>int</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>long int</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>char</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>short</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>float</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>double</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>long double</td>
<td>8</td>
<td>8</td>
<td>10/12</td>
</tr>
<tr>
<td>char *</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>other pointer</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

The integer data types (int, long int, short, char) can all be either signed or unsigned.

Byte Ordering

How should bytes within multi-byte word be ordered in memory?

Given 64-bit hex value 0x0001020304050607, it is common to store this in memory in one of two formats: big endian or little endian.

Big Endian: Least significant byte has highest address.

Little Endian: Least significant byte has lowest address.

Example:
- Variable x has 4-byte representation 0x01234567.
- Address given by &x is 0x100

Big Endian:

<table>
<thead>
<tr>
<th>Address</th>
<th>0x100</th>
<th>0x101</th>
<th>0x102</th>
<th>0x103</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>01</td>
<td>23</td>
<td>45</td>
<td>67</td>
</tr>
</tbody>
</table>

Little Endian:

<table>
<thead>
<tr>
<th>Address</th>
<th>0x100</th>
<th>0x101</th>
<th>0x102</th>
<th>0x103</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>67</td>
<td>45</td>
<td>23</td>
<td>01</td>
</tr>
</tbody>
</table>
Disassembly
- Yields textual representation of binary machine code.
- Generated by program that reads the machine code.

Example Fragment (IA32):

<table>
<thead>
<tr>
<th>Address</th>
<th>Instruction</th>
<th>Code</th>
<th>Assembly Rendition</th>
</tr>
</thead>
<tbody>
<tr>
<td>8048365:</td>
<td>5b</td>
<td></td>
<td>pop %ebx</td>
</tr>
<tr>
<td>8048366:</td>
<td>81 c3 ab</td>
<td>00 00 00</td>
<td>add $0x12ab,%ebx</td>
</tr>
<tr>
<td>804836c:</td>
<td>83 bb 28</td>
<td>00 00 00</td>
<td>cmpl $0x0.0x28(%ebx)</td>
</tr>
</tbody>
</table>

Deciphering Numbers: Consider the value 0x12ab in the second line of code:
- Pad to 4 bytes: 0x000012ab
- Split into bytes: 00 00 12 ab
- Make little endian: ab 12 00 00

Code to Print Byte Representations of Data
Casting a pointer to unsigned char * creates a byte array.

```c
typedef unsigned char *pointer;

void show_bytes(pointer start, int len)
{
    int i;
    for (i = 0; i < len; i++)
        printf("0%x\t0x%.2x\n", start+i, start[i]);
    printf("\n");
}
```

Printf directives:
- %p: print pointer
- %x: print hexadecimal

show_bytes Execution Example

```c
int a = 15213;
printf("int a = 15213;\n");
show_bytes((pointer)&a, sizeof(int));
```

Result (Linux):

```
int a = 15213;
0x007fff90c56c7c 0x6d
0x007fff90c56c7d 0x3b
0x007fff90c56c7e 0x00
0x007fff90c56c7f 0x00
```

Representing Integers

```c
int A = 15213;
int B = -15213;
long int C = 15213;
```

```
15213_{10} = 0011101101101101_{2} = 3B6D_{16}
```

<table>
<thead>
<tr>
<th>Linux</th>
<th>Alpha</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6D 3B 00 00</td>
<td>6D 3B 00 00</td>
</tr>
<tr>
<td>B</td>
<td>93 C4 FF FF</td>
<td>93 C4 FF FF</td>
</tr>
<tr>
<td>C</td>
<td>6D 3B 00 00</td>
<td>6D 3B 00 00</td>
</tr>
</tbody>
</table>

We’ll cover the representation of negatives shortly.
Representing Pointers

```c
int B = -15213;
int *P = &B;
```

**Linux Address:**
Hex: BFFFF8D4AFBB4CD0
In memory: D0 4C BB AF D4 F8 FF BF

**Sun Address:**
Hex: EFFFFFB2CAA2C15C0
In Memory: EF FF FB 2C AA 2C 15 C0

*Pointer values generally are not predictable. Different compilers and machines assign different locations.*

Representing Floats

All modern machines implement the IEEE Floating Point standard. This means that it is consistent across all machines.

```plaintext
IEEE 754 Floating Point Standard:

\[
\begin{array}{c|c|c}
\text{s} & \text{e-exponent} & \text{m=mantissa} \\
1\text{ bit} & 8\text{ bits} & 23\text{ bits} \\
\end{array}
\]
```

```c
float F = 15213.0;
```

Binary: 01000110011011011011010000000000
Hex: 46DB400
In Memory (Linux/Alpha): 00 B4 6D 46
In Memory (Sun): 46 6D B4 00

Note that it’s not the same as the int representation, but you can see that the int is in there, if you know where to look.

Representing Strings

Strings are represented by a sequence of characters. Each character is encoded in ASCII format.

- Standard 7-bit encoding of character set.
- Other encodings exist, but are less common.

Strings should be null-terminated. That is, the final character has ASCII code 0.

**Compatibility**

- *Byte ordering not an issue* since the data are single byte quantities.
- Text files are generally platform independent, except for different conventions of line break character(s).

Machine Level Code Representation

Encode Program as Sequence of Instructions

- Each simple operation
  - Arithmetic operation
  - Read or write memory
  - Conditional branch
- Instructions are encoded as sequences of bytes.
  - Alpha, Sun, PowerPC Mac use 4 byte instructions (Reduced Instruction Set Computer” (RISC)).
  - PC’s and Intel Mac’s use variable length instructions (Complex Instruction Set Computer (CISC)).
- Different instruction types and encodings for different machines.
- Most code is not binary compatible.

**Remember:** Programs are byte sequences too!
### Representing Instructions

```c
int sum(int x, int y) {
    return x + y;
}
```

For this example, Alpha and Sun use two 4-byte instructions. They use differing numbers of instructions in other cases.

PC uses 7 instructions with lengths 1, 2, and 3 bytes. Windows and Linux are not fully compatible.

Different machines typically use different instructions and encodings.

**Instruction sequence for sum program:**

**Alpha:** 00 00 30 42 01 80 FA 68  
**Sun:** 81 C3 E0 08 90 02 00 09  
**PC:** 55 89 E5 8B 45 OC 03 45 08 89 EC 5D C3

### Boolean Algebra

Developed by George Boole in the 19th century, Boolean algebra is the algebraic representation of logic. We encode “True” as 1 and “False” as 0.

In C, we encode “False” as 0 and “True” as any non-zero value.

**And:** $A \land B = 1$ when both $A = 1$ and $B = 1$.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>&amp;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Or:** $A \lor B = 1$ when either $A = 1$ or $B = 1$, but not both.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>\lor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Not:** $\neg A = 1$ when $A = 0$.

<table>
<thead>
<tr>
<th>A</th>
<th>\neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Xor:** $A \oplus B = 1$ when either $A = 1$ or $B = 1$, but not both.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>\oplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

These are represented in C by the operators: `&&`, `||`, `!`, `^`.
In a 1937 MIT Master’s Thesis, Claude Shannon showed that Boolean algebra would be a great way to model digital networks.

At that time, the networks were relay switches. But today, all combinational circuits can be described in terms of Boolean “gates.”

\[ \{0, 1\}, |, \&, \sim, 0, 1 \] forms a Boolean algebra.

- Or is the sum operation.
- And is the product operation.
- \( \sim \) is the “complement” operation (not additive inverse).
- 0 is the identity for sum.
- 1 is the identity for product.

Some boolean algebra properties are similar to integer arithmetic, some are not.

**Commutativity:**
- \( A \lor B = B \lor A \)
- \( A \land B = B \land A \)

**Associativity:**
- \( (A \lor B) \lor C = A \lor (B \lor C) \)
- \( (A \land B) \land C = A \land (B \land C) \)

**Product Distributes over Sum:**
- \( A \land (B \lor C) = (A \land B) \lor (A \land C) \)
- \( (A \lor B) \land (A \land C) \)

**Sum and Product Identities:**
- \( A \lor 0 = A \)
- \( A \land 1 = A \)
- \( A + 0 = A \)
- \( A \land 1 = A \)
- \( A \lor 0 = 0 \)
- \( A \land 1 = A \)
- \( \sim (\sim A) = A \)
- \( -(\neg A) = A \)

The following boolean algebra rules don’t have analogs in integer arithmetic.

**Boolean:** Sum distributes over product
- \( A \lor (B \land C) = (A \lor B) \land (A \lor C) \)
- \( A \land (B \lor C) \neq (A \land B) \lor (A \land C) \)

**Boolean:** Idempotency
- \( A \land A = A \land A \neq A \land A \)
- \( A \lor A = A \lor A \neq A \lor A \)
- \( A \land A = A \land A \neq A \land A \)
Boolean Algebra Properties

**Boolean: Absorption**
\[
A \land (A \lor B) = A \\
A \lor (A \land B) = A
\]

**Boolean: Laws of Complements**
\[
A \sim A = 1 \\
A \sim \sim A = A
\]

**Ring: Every element has additive inverse**
\[
A \sim A = 0 \\
A + \sim A = 0
\]

**Properties of \& and ^**

**Commutative sum:**
\[A \land B = B \land A\]

**Commutative product:**
\[A \lor B = B \lor A\]

**Associative sum:**
\[(A \land B) \land C = A \land (B \land C)\]

**Associative product:**
\[(A \lor B) \lor C = A \lor (B \lor C)\]

**Prod. over sum:**
\[A \land (B \lor C) = (A \land B) \lor (A \land C)\]

**0 is sum identity:**
\[A \land 0 = 0\]

**1 is prod. identity:**
\[A \lor 1 = A\]

**0 is product annihilator:**
\[A \land 0 = 0\]

**Additive inverse:**
\[A \sim A = 0\]

**Relations Between Operations**

**DeMorgan’s Laws**
Express \& and using | and vice-versa:
\[
A \land B \sim (\sim A \lor \sim B)
\]
\[
A \lor B \sim (\sim A \land \sim B)
\]

**Exclusive-Or using Inclusive Or:**
\[
A \land B = (\sim A \land B) \lor (A \land B)
\]
\[
A \lor B = (A \lor B) \land \sim (A \land B)
\]

**Generalized Boolean Algebra**
We can also operate on bit vectors (bitwise). All of the properties of Boolean algebra apply:

\[
\begin{array}{cccc}
01101001 & 01101001 & 01101001 \\
\lor & 01010101 & \land & 01111101 & 01010101 & \land & 01010101 \\
01000001 & 01111101 & 00111100 & 10101010
\end{array}
\]
Bit Level Operations in C

The operations &, |, ∼, ^ are all available in C.
- Apply to any integral data type: long, int, short, char.
- View the arguments as bit vectors.
- Operations are applied bit-wise to the argument(s).

**Examples:** (char data type)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>∼ 0x41</td>
<td></td>
<td></td>
<td>0xBE</td>
</tr>
<tr>
<td>∼ 01000001₂</td>
<td></td>
<td></td>
<td>10111111₂</td>
</tr>
<tr>
<td>∼ 0x00</td>
<td></td>
<td></td>
<td>0xFF</td>
</tr>
<tr>
<td>∼ 00000000₂</td>
<td></td>
<td></td>
<td>11111111₂</td>
</tr>
<tr>
<td>0x69 &amp; 0x55</td>
<td></td>
<td></td>
<td>0x41</td>
</tr>
<tr>
<td>01101001₂ &amp; 01010101₂</td>
<td></td>
<td></td>
<td>0100001₂</td>
</tr>
<tr>
<td>0x69</td>
<td>0x55</td>
<td></td>
<td>0x7D</td>
</tr>
<tr>
<td>01101001₂</td>
<td>01010101₂</td>
<td></td>
<td>01111111₂</td>
</tr>
</tbody>
</table>

Contrast to Logical Operators in C

Remember the operators: &&, ||, !.
- View 0 as “False.”
- View anything nonzero as “True.”
- Always return 0 or 1.
- Always do short-circuit evaluation (early termination)
- There isn’t a “logical" xor, but != works if you know the inputs are boolean.

**Examples:**

<table>
<thead>
<tr>
<th>Boolean op</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>!0x41</td>
<td></td>
<td></td>
<td>0x00</td>
</tr>
<tr>
<td>!0x00</td>
<td></td>
<td></td>
<td>0x01</td>
</tr>
<tr>
<td>!!0x41</td>
<td></td>
<td></td>
<td>0x01</td>
</tr>
<tr>
<td>!!0x00</td>
<td></td>
<td></td>
<td>0x01</td>
</tr>
<tr>
<td>!!0x69 &amp;&amp; 0x55</td>
<td></td>
<td></td>
<td>0x01</td>
</tr>
<tr>
<td>!!0x69</td>
<td></td>
<td>0x55</td>
<td></td>
</tr>
</tbody>
</table>

Representing Sets with Masks

A bit vector \( a \) may represent subsets \( S \) of some “reference set” (actually list) \( L \): \( a_j = 1 \) iff \( L[j] \in S \)

Bit vector A: 01101001 represents \( \{0, 3, 5, 6\} \)

Bit vector B: 01010101 represents \( \{0, 2, 4, 6\} \)

Operations:
Given the two sets above, perform these bitwise ops to obtain:

<table>
<thead>
<tr>
<th>Set operation</th>
<th>Boolean op</th>
<th>Result</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intersection</td>
<td>A &amp; B</td>
<td>01000001</td>
<td>( {0, 6} )</td>
</tr>
<tr>
<td>Union</td>
<td>A</td>
<td>B</td>
<td>01111101</td>
</tr>
<tr>
<td>Symmetric difference</td>
<td>A ^ B</td>
<td>00111100</td>
<td>( {2, 3, 4, 5} )</td>
</tr>
<tr>
<td>Complement</td>
<td>∼A</td>
<td>10010110</td>
<td>( {1, 2, 4, 7} )</td>
</tr>
</tbody>
</table>

What bit operations on these set representations correspond to: intersection, union, complement?
Shift Operations

Left Shift: \( x << y \)
Shift bit vector \( x \) left by \( y \) positions

- Throw away extra bits on the left.
- Fill with 0's on the right.

Right Shift: \( x >> y \)
Shift bit vector \( x \) right by \( y \) positions.

- Throw away extra bits on the right.
- **Logical shift**: Fill with 0’s on the left.
- **Arithmetic shift**: Replicate with most significant bit on the left.

Arithmetic shift is useful with two’s complement integer representation.

<table>
<thead>
<tr>
<th>Argument ( x )</th>
<th>01100010</th>
</tr>
</thead>
<tbody>
<tr>
<td>( &lt;&lt; 3 )</td>
<td>00010000</td>
</tr>
<tr>
<td>Log. ( &gt;&gt; 2 )</td>
<td>00011000</td>
</tr>
<tr>
<td>Arith. ( &gt;&gt; 2 )</td>
<td>00110000</td>
</tr>
</tbody>
</table>

Cool Stuff with XOR

Bitwise XOR is a form of addition, with the extra property that each value is its own additive inverse: \( A \oplus A = 0 \).

```c
void funny_swap(int *x, int *y)
{
    *x = *x ^ *y; /* #1 */
    *y = *x ^ *y; /* #2 */
    *x = *x ^ *y; /* #3 */
}
```

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A ( \oplus ) B</td>
</tr>
<tr>
<td>2</td>
<td>A ( \oplus ) B</td>
</tr>
<tr>
<td>3</td>
<td>A ( \oplus ) B</td>
</tr>
<tr>
<td>End</td>
<td>B</td>
</tr>
</tbody>
</table>

Is there ever a case where this code fails?

Main Points

It’s all about bits and bytes.
- Numbers
- Programs
- Text

Different machines follow different conventions.
- Word size
- Byte ordering
- Representations

Boolean algebra is the mathematical basis.
- Basic form encodes “False” as 0 and “True” as 1.
- General form is like bit-level operations in C; good for representing and manipulating sets.