CS 31: Introduction to Computer Systems

03: Binary Arithmetic
January 28
Clicker Question: Have you registered your clicker?

A. Yes
B. No
C. What’s a clicker?

Check your frequency:

- Iclicker2: frequency AA
- Iclicker+: green light next to selection

For new devices this should be okay,
For used you may need to reset frequency

Reset:
1. hold down power button until blue light flashes (2secs)
2. Press the frequency code: AA
vote status light will indicate success
Reading Quiz
Abstraction

User / Programmer
Wants low complexity

Applications
Specific functionality

Software library
Reusable functionality

Operating system
Manage resources

Complex devices
Compute & I/O

Compute & I/O
Operating system
Software library
Applications
User / Programmer
Abstraction
Today

• Binary Arithmetic
  – Unsigned addition
  – Subtraction
• Representation
  – Signed magnitude
  – Two’s complement
  – Signed overflow
• Bit operations
Last Class: Binary Digits: (BITS)

Most significant bit ➔ 10001111 ➙ Least significant bit

Representation: \[ 1 \times 2^7 + 0 \times 2^6 + \ldots + 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \]

\[ 10001111 = 143 \]

one byte is the smallest addressable unit - contains 8 bits
Last Class: Unsigned Integers

• Suppose we had one byte
  – Can represent $2^8$ (256) values
  – If unsigned (strictly non-negative): 0 – 255
Last Class: Unsigned Arithmetic (one byte)

- one byte
  - $2^8$ (256) values
  - unsigned: $0 - 255$

0 = 00000000
1 = 00000001
2 = 00000010
...
254 = 11111110
255 = 11111111

Number line:

Addition

$255 = 1 \times 2^7 + \ldots + 1 \times 2^0$
Last Class: Unsigned Arithmetic (one byte)

- one byte
  - $2^8$ (256) values
  - unsigned: 0 – 255

$252 = 11111100$
$253 = 11111101$
$254 = 11111110$
$255 = 11111111$

What if we add one more?
$255 + 1$ is?

Car odometer “rolls over”.

we cannot represent an infinite number of values in a finite storage space
Last Class: Arithmetic and Fixed Storage

- Fixed Storage: finite set of values
  - 1 byte: $2^8$ (256) values
  - unsigned values: $0 - 255$

- Yields **Modular Arithmetic**
  - All operations are $\%$ 256
    (eg) $255 + 4 = 259 \% 256 = 3$

Modular arithmetic: Here, all values are modulo 256.
Last Class: Unsigned Addition (4-bit)

Addition works like grade school addition:

\[
\begin{align*}
1 \\
0110 & \quad 6 \\
+ 0100 & \quad + 4 \\
\hline
1010 & \quad 10
\end{align*}
\]

Four bits give us range: 0 - 15
Last Class: Unsigned Addition (4-bit)

Addition works like grade school addition:

\[
\begin{align*}
1 & \\
0110 & + 6 & 1100 & + 12 \\
+ 0100 & + 4 & + 1010 & + 10 \\
1010 & + 10 & 1 0110 & + 6 \\
\end{align*}
\]

^carry out

Four bits give us range: 0 - 15

Overflow!
Last Class: Arithmetic and Fixed Storage

- Fixed Storage: finite set of values
  - 1 byte: $2^8$ (256) values
  - unsigned values: 0 – 255
Not Used: Signed Magnitude Representation (for 4 bit values)

One bit (usually left-most) signals:
- 0 for positive
- 1 for negative

For one byte:
1 = 00000001
-1 = 10000001

Pros: Negation (negative value of a number) is very simple!

For one byte:
0 = 00000000
-0? = 10000000

Major con: Two ways to represent zero.
Used Today: Two’s Complement Representation (for four bit values)

- Borrow nice property from number line:
  
  ![-1 0 1]

  Only one instance of zero!
  Implies: -1 and 1 on either side of it.

For an 8 bit range we can express 256 unique values:
- 128 non-negative values (0 to 127)
- 128 negative values (-1 to -128)
Two’s Complement

• Only one value for zero
• With N bits, can represent the range:
  ▪ \(-2^{N-1}\) to \(2^{N-1} - 1\)
• Most significant bit still designates
  – 0: positive
  – 1: negative

• Negating a value is slightly more complicated:
  \(1 = 00000001, \quad -1 = 11111111\)

From now on, unless we explicitly say otherwise, we’ll assume all integers are stored using two’s complement! This is the standard!
Two’s Complement

- Each two’s complement number is now:
  \[ -2^{n-1}d_{n-1} + 2^{n-2}d_{n-2} + \ldots + 2^1d_1 + 2^0d_0 \]

Note the **negative sign** on just the most significant bit. This is why first bit tells us whether the value is negative vs. positive.
If we interpret 11001 as a two’s complement number, what is the value in decimal?

Each two’s complement number is now:

\[-2^{n-1}d_{n-1} + 2^{n-2}d_{n-2} + \ldots + 2^1d_1 + 2^0d_0\]

A. -2

B. -7

C. -9

D. -25
If we interpret 11001 as a two’s complement number, what is the value in decimal?

Each two’s complement number is now:
\[-2^{n-1}\times d_{n-1} \quad + \quad 2^{n-2}\times d_{n-2} \quad + \quad ... \quad + \quad 2^1\times d_1 \quad + \quad 2^0\times d_0\]

A. -2

B. -7 \quad -16 + 8 + 1 = -7

C. -9

D. -25
“If we interpret…”

What is the decimal value of 1100?

• ...as unsigned, 4-bit value: 12 (%u)
• ...as signed (two’s comp), 4-bit value: -4 (%d)

• ...as an 8-bit value: 12. (i.e., \textbf{0000 1100})
Two’s Complement Negation

• To negate a value $x$, we want to find $y$ such that $x + y = 0$.

• For $N$ bits, $y = 2^N - x$
Negation Example (8 bits)

- For N bits, \( y = 2^N - x \)
- Negate the value (2) 00000010
- \( 2^8 - 2 = 256 - 2 = 254 \)

- Our wheel only goes to 127!
  - Put -2 where 254 would be if wheel was unsigned.
  - 254 in binary is 11111110

Given 11111110, it’s 254 if interpreted as unsigned and -2 interpreted as signed.
Negation Shortcut

• A much easier, faster way to negate:
  – Flip the bits (0’s become 1’s, 1’s become 0’s)
  – Add 1

• Negate 00101110 (46)

• Formally:
  – $2^8 - 46 = 256 - 46 = 210$
  – 210 in binary is 11010010
## Negation Two Ways

### 4 bit Examples

<table>
<thead>
<tr>
<th>x</th>
<th>-x</th>
<th>$2^4 - x$</th>
<th>Bit flip + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>0000</td>
<td>10000 − 0000 = 0000</td>
<td>1111 + 1 = 0000</td>
</tr>
<tr>
<td>0001</td>
<td>1111</td>
<td>10000 − 0001 = 1111</td>
<td>1110 + 1 = 1111</td>
</tr>
<tr>
<td>0010</td>
<td>110</td>
<td>10000 − 0010 = 1110</td>
<td>1101 + 1 = 1110</td>
</tr>
<tr>
<td>0111</td>
<td>1001</td>
<td>10000 − 0111 = 1001</td>
<td>1000 + 1 = 1001</td>
</tr>
</tbody>
</table>
Decimal to Two’s Complement with 8 bit values
(high-order bit is the sign bit)

for positive values, use same algorithm as for unsigned
• (E.g.) 6 6 – 4 = 2 \((4 : 2^2)\)
• 2 – 2 = 0 \((2 : 2^1)\): 00000110

for negative values:
• convert negation (positive) to binary
• then negate binary to get negative

E.g.: -3
• 3: 00000011
• negate: \(\text{11111100} + 1 = \text{11111101} = -3\)
Decimal to Two’s Complement with 8 bit values
(high-order bit is the sign bit)

for negative values:
• convert negation (positive) to binary
• then negate binary to get negative

Try converting -7 to Two’s Complement representation

A. 11111001
B. 00000111
C. 11111000
D. 11110011
Decimal to Two’s Complement with 8 bit values (high-order bit is the sign bit)

for negative values:
• convert negation (positive) to binary
• then negate binary to get negative

Try converting -7 to Two’s Complement representation

A. 11111001
B. 00000111
C. 11111000
D. 11110011

-7 = (1) 7: 00000111
(2) negate: 11111000 + 1 = 11111001
Addition & Subtraction for Integers

• Addition is the same as for unsigned
  – One exception: different rules for overflow
  – Can use the same hardware for both

• Subtraction is the same operation as addition
  – Just need to negate the second operand...

• $6 - 7 = 6 + (-7) = 6 + (~7 + 1)$
  – $~7$ is shorthand for “flip the bits of 7”
Subtraction Hardware

Negate and add 1 to second operand:
Can use the same circuit for add and subtract:
\[6 - 7 = 6 + \neg 7 + 1\]

Let’s call this possible +1 input: “Carry in”
(0: on add, 1: on subtract)
4-bit signed Examples:

Subtraction via Addition:

- \( a - b \) is same as \( a + \sim b + 1 \)

Subtraction: flip bits and add 1

\[
3 - 6 = 0011 \\
1001 \quad (6: 0110 \sim 6: 1001) \\
+ \quad 1 \\
1101 = -3
\]

Addition: don’t flip bits or add 1

\[
3 + -6 = 0011 \\
+ 1010 \\
1101 = -3
\]
Signed & Unsigned 4-bit Subtraction:

Unsigned subtraction: flip bits and add 1

13 - 1 =

Signed subtraction: flip bits and add 1

-3 - 1 =

A. 1100 & 1100
B. 1100 & 1010
C. 1010 & 1010
D. 1001 & 1100
Signed & Unsigned 4-bit Subtraction:

Unsigned subtraction: flip bits and add 1

$$13 - 1 = \text{1101}$$

$$\text{1110}$$ (1: 0001 ~1: 1110)

$$+ \quad 1$$

$$1 \quad \text{1100} = 12$$

Signed subtraction: flip bits and add 1

$$-3 - 1 = \text{1101}$$

$$\text{1110}$$

$$+ \quad 1$$

$$1 \quad \text{1100} = -4$$
By switching to two’s complement, have we solved this value “rolling over” (overflow) problem?

A. Yes, it’s gone.

B. Nope, it’s still there.

C. It’s even worse now.
By switching to two’s complement, have we solved this value “rolling over” (overflow) problem?

A. Yes, it’s gone.

B. Nope, it’s still there.

C. It’s even worse now.

This is an issue we need to be aware of when adding and subtracting!
Overflow, Revisited

Unsigned

192

128

64

255

0

Signed

-1

0

1

-127

127

-128

Danger Zone

Danger Zone
If we add a positive number and a negative number, will we have overflow? (Assume they are the same # of bits)

A. Always
B. Sometimes
C. Never
If we add a positive number and a negative number, will we have overflow? (Assume they are the same # of bits)

A. Always

B. Sometimes

C. Never
Signed (Two’s Complement) Overflow For Addition

- **Addition Overflow**: IFF the sign bits of operands are the same, but the sign bit of result is different.
- Not enough bits to store result!

sign of operands = sign of result

<table>
<thead>
<tr>
<th>3+4=7</th>
<th>-2+(-3)=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0011</td>
<td>1110</td>
</tr>
<tr>
<td>+0100</td>
<td>+1101</td>
</tr>
<tr>
<td>0111</td>
<td>1 1011</td>
</tr>
</tbody>
</table>

no overflow

3+4=7 -2+-3=5
0011 1110
+0100 +1101
0111 1 1011
Signed (Two’s Complement) Overflow For Addition

- **Addition Overflow**: IFF the sign bits of operands are the same, but the sign bit of result is different.
- Not enough bits to store result!

<table>
<thead>
<tr>
<th>Sign of Operands = Sign of Result</th>
<th>Sign of Operands ≠ Sign of Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Overflow</strong></td>
<td><strong>Overflow</strong></td>
</tr>
<tr>
<td>$3+4=7$</td>
<td>$4+7=11$</td>
</tr>
<tr>
<td>$0011$</td>
<td>$0100$</td>
</tr>
<tr>
<td>$+0100$</td>
<td>$+0111$</td>
</tr>
<tr>
<td>$0111$</td>
<td>$1011$</td>
</tr>
</tbody>
</table>

| $-2+3=-5$                         | $-6-8=-14$                        |
| $1110$                            | $1010$                            |
| $+1101$                           | $+1000$                           |
| $1011$                            | $10010$                           |

$0$ $011$ $1$ $110$ $0$ $100$ $1$ $010$
Subtraction Overflow Two Rules:

Rule 1:
- Positive operand - Negative operand = Positive Result: No Overflow
- Positive operand - Negative operand = Negative Result: Overflow
  • Intuition: We know a positive – negative is equivalent to a positive + positive. If this sum does not result in a positive value we have an overflow

Rule 2:
- Negative operand - Positive operand = Negative Result: No Overflow
- Negative operand - Positive operand = Positive Result: Overflow
  • Intuition: We know a negative – positive number is equivalent to a negative + negative number. If this sum does not result in a negative value we have an overflow
Signed (Two’s Complement) Overflow For Subtraction

Subtraction Overflow Rules Summarized:
• IFF the sign bits of the subtraction operands are different, and the sign bit of the Result and Subtrahend are the same as shown below:
  – Minuend - Subtrahend = Result
  – If positive – negative = negative (overflow)
  – If negative – positive = positive (overflow)
Signed (Two’s Complement) Overflow For Subtraction

Subtraction Overflow Two Rules:

- **Rule 1:**
  - Positive operand - Negative operand = Positive Result: No Overflow
  - Positive operand - Negative operand = Negative Result: Overflow

Subtrahend and Result have **different sign bits**

<table>
<thead>
<tr>
<th>Operation</th>
<th>Result</th>
<th>Sign Bits</th>
<th>Subtrahend</th>
<th>Minuend</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 - (-3) = 5</td>
<td>0101</td>
<td>+</td>
<td>0010</td>
<td>-1110</td>
</tr>
<tr>
<td>3 - (-4) = 7</td>
<td>0111</td>
<td>+</td>
<td>0011</td>
<td>-1100</td>
</tr>
</tbody>
</table>

Subtrahend and Result have the **same sign bits**

<table>
<thead>
<tr>
<th>Operation</th>
<th>Result</th>
<th>Sign Bits</th>
<th>Subtrahend</th>
<th>Minuend</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 - (-6) = 8</td>
<td>1000 (-8)</td>
<td>-</td>
<td>0010</td>
<td>-1010</td>
</tr>
<tr>
<td>3 - (-7) = 10</td>
<td>1010 (-6)</td>
<td>-</td>
<td>0011</td>
<td>-1001</td>
</tr>
</tbody>
</table>

2 - (-3) = 5
0010
-1110
0010
+0011
0101

3 - (-4) = 7
0011
-1100
0011
+0100
0111

2 - (-6) = 8
0010
-1010
0010
+0110
1000 (-8)

3 - (-7) = 10
0011
-1001
0011
+0111
1010 (-6)
Signed (Two’s Complement) Overflow For Subtraction

Subtraction Overflow Two Rules:

- **Rule 2:**
  - Negative operand - Positive operand = Negative Result: No Overflow
  - Negative operand - Positive operand = Positive Result: Overflow

Subtrahend and Result have different sign bits

Subtrahend and Result have the same sign bits

<table>
<thead>
<tr>
<th>No Overflow</th>
<th>Overflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 - (3) = -5</td>
<td>-2 - (7) = -9</td>
</tr>
<tr>
<td>1110</td>
<td>1110</td>
</tr>
<tr>
<td>-0011</td>
<td>-0111</td>
</tr>
<tr>
<td>1110</td>
<td>1101</td>
</tr>
<tr>
<td>+1101</td>
<td>+1001</td>
</tr>
<tr>
<td>1 1011 (-5)</td>
<td>1 0111 (7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overflow</th>
<th>No Overflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4 - (7) = -11</td>
<td>-2 - (3) = -5</td>
</tr>
<tr>
<td>1100</td>
<td>1100</td>
</tr>
<tr>
<td>-0111</td>
<td>-0011</td>
</tr>
<tr>
<td>1110</td>
<td>1110</td>
</tr>
<tr>
<td>+1001</td>
<td>+0111</td>
</tr>
<tr>
<td>1 0111 (7)</td>
<td>1 0011 (-6)</td>
</tr>
</tbody>
</table>
Overflow Rules

• Signed (Two’s Complement):
  – Addition:
    • The sign bits of operands are the same, but the sign bit of result is different.
  – Subtraction:
    • First compute the following: if the sign bits of the subtraction operands are different, and the sign bit of the result and subtrahend are the same. \((\text{minuend} - \text{subtrahend} - \text{result})\)
    • then, turn into an Addition operation

• Can we formalize unsigned overflow?
  – Need to include subtraction too, skipped it before.
Recall Subtraction Hardware

Negate and add 1 to second operand:
Can use the same circuit for add and subtract:
\[ 6 - 7 = 6 + \sim 7 + 1 \]

input 1 ----------------------------->
input 2 --> possible bit flipper --> ADD CIRCUIT ---> result
possible +1 input------>

Let's call this possible +1 input: “Carry in”
(0: on add, 1: on subtract)
How many of these **unsigned** operations have overflowed?

4 bit unsigned values (range 0 to 15):

<table>
<thead>
<tr>
<th>Operation</th>
<th>Inputs</th>
<th>carry-in</th>
<th>carry-out</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addition</strong> (carry-in = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$9 + 11$</td>
<td>1001 + 1011 + 0</td>
<td>1</td>
<td>0100</td>
</tr>
<tr>
<td>$9 + 6$</td>
<td>1001 + 0110 + 0</td>
<td>0</td>
<td>1111</td>
</tr>
<tr>
<td>$3 + 6$</td>
<td>0011 + 0110 + 0</td>
<td>0</td>
<td>1001</td>
</tr>
</tbody>
</table>

| **Subtraction** (carry-in = 1) |                   |          |           |
| $6 - 3$            | 0110 + 1100 + 1    | (-3)     | 0011      |
| $3 - 6$            | 0011 + 1010 + 1    | (-6)     | 1101      |

A. 1  
B. 2  
C. 3  
D. 4  
E. 5
How many of these **unsigned** operations have overflowed?

4 bit unsigned values (range 0 to 15):

<table>
<thead>
<tr>
<th>Addition (carry-in = 0)</th>
<th>carry-in</th>
<th>carry-out</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 + 11 = 1001 + 1011 + 0 = 1 0100 = 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 + 6 = 1001 + 0110 + 0 = 0 1111 = 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 + 6 = 0011 + 0110 + 0 = 0 1001 = 9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtraction (carry-in = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 3 = 0110 + 1100 + 1 = 1 0011 = 3</td>
</tr>
<tr>
<td>3 - 6 = 0011 + 1010 + 1 = 0 1101 = 13</td>
</tr>
</tbody>
</table>

A. 1  
B. 2  
C. 3  
D. 4  
E. 5  

*Pattern?
How many of these unsigned operations have overflowed?

4 bit unsigned values (range 0 to 15):

Addition (carry-in = 0)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 + 11</td>
<td>1001 + 1011 + 0 = 1 0100 = 4</td>
</tr>
<tr>
<td>9 + 6</td>
<td>1001 + 0110 + 0 = 0 1111 = 15</td>
</tr>
<tr>
<td>3 + 6</td>
<td>0011 + 0110 + 0 = 0 1001 = 9</td>
</tr>
</tbody>
</table>

Subtraction (carry-in = 1)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 3</td>
<td>0110 + 1100 + 1 = 1 0011 = 3</td>
</tr>
<tr>
<td>3 - 6</td>
<td>0011 + 1010 + 1 = 0 1101 = 13</td>
</tr>
</tbody>
</table>

A. 1  
B. 2  Pattern?  
C. 3  
D. 4  
E. 5
Overflow Rule Summary

- **Signed overflow:**
  - The sign bits of operands are the same, but the sign bit of result is different.

- **Unsigned: overflow**
  - The carry-in bit is different from the carry-out.

<table>
<thead>
<tr>
<th>$C_{in}$</th>
<th>$C_{out}$</th>
<th>$C_{in}$ XOR $C_{out}$</th>
<th>$C_{out}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

So far, all arithmetic on values that were the same size. What if they’re different?
Sign Extension

- When combining signed values of different sizes, expand the smaller to equivalent larger size:

```cpp
char y=2, x=-13;
short z = 10;

z = z + y;  // 0000000000001010 + 0000000000000101 = 0000000000001010
z = z + x;  // 0000000000001010 + 11110011 = 1111111111110011
```

Fill in high-order bits with sign-bit value to get same numeric value in larger number of bytes.
Let’s verify that this works

4-bit signed value, sign extend to 8-bits, is it the same value?

0111  --->  0000 0111  obviously still 7
1010  --->  1111 1010  is this still -6?

-128 + 64 + 32 + 16 + 8 + 0 + 2 + 0 =  -6  yes!
Operations on Bits

• For these, doesn’t matter how the bits are interpreted (signed vs. unsigned)

• Bit-wise operators (AND, OR, NOT, XOR)

• Bit shifting
### Bit-wise Operators

- bit operands, bit result (interpret as you please)

<table>
<thead>
<tr>
<th>&amp; (AND)</th>
<th></th>
<th>(OR)</th>
<th>~ (NOT)</th>
<th>^ (XOR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A &amp; B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>01010101</th>
<th>01101010</th>
<th>10101010</th>
<th>~10101111</th>
</tr>
</thead>
<tbody>
<tr>
<td>00100001</td>
<td>10111011</td>
<td>01101001</td>
<td>01010000</td>
</tr>
</tbody>
</table>

Slide 60
More Operations on Bits

- Bit-shift operators: \( \ll \) left shift, \( \gg \) right shift

- \( 01010101 \ll 2 \) is \( 01010100 \)
  - 2 high-order bits shifted out
  - 2 low-order bits filled with 0

- \( 01101010 \ll 4 \) is \( 10100000 \)

- \( 01010101 \gg 2 \) is \( 00010101 \)

- \( 01101010 \gg 4 \) is \( 00000110 \)

- \( 10101100 \gg 2 \) is \( 00101011 \) (logical shift)
  or \( 11101011 \) (arithmetic shift)

Arithmetic right shift: fills high-order bits w/sign bit

C automatically decides which to use based on type:
  - signed: arithmetic, unsigned: logical
Try out some 4-bit examples:

bit-wise operations:
• 0101 & 1101
• 0101 | 1101

Logical (unsigned) bit shift:
• 1010 << 2
• 1010 >> 2

Arithmetic (signed) bit shift:
• 1010 << 2
• 1010 >> 2
Try out some 4-bit examples:

bit-wise operations:
• \( 0101 \& 1101 = 0101 \)
• \( 0101 \mid 1101 = 1101 \)

Logical (unsigned) bit shift:
• \( 1010 << 2 = 1000 \)
• \( 1010 >> 2 = 0010 \)

Arithmetic (signed) bit shift:
• \( 1010 << 2 = 1000 \)
• \( 1010 >> 2 = 1110 \)
Additional Info: Fractional binary numbers

How do we represent fractions in binary?

-1/2

1/8

1/2

-11982  ....  -1  0  1  15  999...99
Additional Info: Representing Signed Float Values

• One option (used for floats, NOT integers)
  – Let the first bit represent the sign
  – 0 means positive
  – 1 means negative

• For example:
  – 0101   ->  5
  – 1101   -> -5

• Problem with this scheme?
Additional Info: Floating Point Representation

1 bit for sign | exponent | fraction |
8 bits for exponent
23 bits for precision

value = \((-1)^{\text{sign}} \times 1.\text{fraction} \times 2^{(\text{exponent}-127)}\)

let's just plug in some values and try it out

0x40ac49ba: 0 10000001 01011000100100110111010
  sign = 0  exp = 129  fraction = 2902458

  = 1 \times 1.2902458 \times 2^{2} = 5.16098

I don't expect you to memorize this
Up Next

• C programming