Lecture 2:
Number Systems

The Decimal System

Every day we use a number system named the *decimal system*. It consists of 10 digits, namely 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. Whenever we write a number, what we are representing is a sum of products. More specifically, each digit in a number is multipled by a power of 10 based on its position in the number, and then, the products are added together. For instance, the number 8492 means 8 thousands plus 4 hundreds plus 9 tens plus 2; or in other words,

$$8 \times 10^3 + 4 \times 10^2 + 9 \times 10^1 + 2 \times 10^0$$

In the decimal system, the number 10 is called the *base* or *radix*. For integers, the representation is

$$\sum_{i=1}^{n} d_i \times 10^{n-i}$$

where n is the total number of digits in the number and $d_1, d_2, \ldots, d_n$ are the digits of the number from left to right. And for fraction numbers, numbers after the decimal point (between 0 and 1 exclusively), the powers of 10 are negative. A fraction, its representation is

$$\sum_{i=1}^{n} d_i \times 10^{-1\times i}$$

where n is the total number of digits in the fraction and $d_1, d_2, \ldots, d_n$ are the digits from left to right. For a number, in general, given that n is the total number of digits in the integer portion and m is the number of digits, the representation is

$$\sum_{i=1}^{m} d_i \times 10^{n-i}$$

where $d_1, d_2, \ldots d_m$ are the digits of the number from left to right. The summation format of a number is called *notation form*. Moreover, in any number, the leftmost digit is called the *most significant digit* and the rightmost digit is called the *least significant digit*. For instance, in the number 345.259, 3 is the most significant digit while 9 is the least significant digit.

Positional Number Systems

The decimal system is a *positional number system*. A positional number system that represents numbers as a string of digits in which each digit is associated to a weight that is a power of r raised to the position of the digit in the number, where r is the radix, or base. That is, a number, a, of radix, r, is

$$a = \sum_{i=1}^{m} d_i \times r^{n-i}$$

where n is the total number of digits in the integer portion of a, m is the total number of digits in a and $d_1, d_2, \ldots, d_m$ are the digits of a from left to right. For a positional number system with radix r there are always r digits and the range of the digits is from 0 inclusively to r exclusively ($0 \leq d < r$). For instance, if $r = 5$, then there would be 5 digits namely 0, 1, 2, 3 and 4. Some examples of number in base 5 are 123.23, $44, 232$ and 0.23214. Last, the dot that goes between the r^0 position and the r^{-1} position is called the *radix point*.

The Binary System

The *binary system* is a base 2 positional systems, which means numbers are represented as follows

$$\sum_{i=1}^{m} b_i \times 2^{n-i}$$

where n is the total number of digits in the integer portion of the number, m is the total number of digits in the number and $b_1, b_2, \ldots, b_m$ are the digits of the number from left to right. In the binary system, its digits are 0 and 1. The binary numbers 0_2 and 1_2 are equivalent to the decimal numbers 0_{10} and 1_{10} respectively as expected. The following table shows the first five positive binary numbers and their equivalent decimal numbers.

Binary Number	**Decimal Number**
1	1
10	2
11	3
100	4
101	5

Converting Between Binary and Decimal

Converting from binary to decimal and vica versa is essential. To convert a binary number to a decimal number, one needs only evaluate the binary number in notation form. For instance,

$$\begin{aligned}
10101.101_2 &= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\
&= (1)(16) + (0)(8) + (1)(4) + (0)(2) + (1)(1) + (1)(0.5) + (0)(0.25) + (1)(0.125) \\
&= 16 + 4 + 1 + 0.5 + 0.125 \\
&= 21.625_{10}
\end{aligned}$$

Converting a decimal number to a binary number is a little bit more difficult. The decimal number must be converted in parts (integer and fractional). To convert the integer portion, you are using the *division theorem* repeatedly. The division theorem states

Given two nonnegative integers, m and n, $m = nq + r$ where q and r are integers and $0 \leq r < n$.

To derive the conversion of the integer portion of a decimal number, we let m be the decimal number and n equal 2, to get

$$m = 2q_1 + r_1$$

Afterwards, we make the quotient the new m and let 2 remain n. The new quotient will be less than the last quotient. Continue repeating this process until the quotient becomes 0.

$$\begin{aligned}
m &= 2q_0 + r_0 \\
q_0 &= 2q_1 + r_1 \\
&\vdots \\
q_{t-2} &= 2q_{t-1} + r_{t-1} \\
q_{t-1} &= 2q_t + r_t \\
q_t &= 0
\end{aligned}$$

Next, starting with the final quotient, substitute what it is equal to in the prior equation. Continue the process until all quotients have been replaced as follows

$$\begin{aligned}
q_{t-1} &= r_t \\
q_{t-2} &= 2r_t + r_{t-1} \\
q_{t-3} &= 2(2r_t + r_{t-1}) + r_{t-2} \\
&= 2^2 r_t + 2r_{t-1} + r_{t-2} \\
&\vdots \\
m &= 2^t r_t + 2^{t-1} r_{t-1} + \ldots + 2r_1 + r_0 \\
&= \sum_{i=0}^{t} 2^i r_i
\end{aligned}$$

Since n was always 2, $r_0, r_1, \ldots, r_t$ are either 0 or 1, which means, we just wrote the decimal number, m, in binary notation as required. Basically, to convert a decimal integer to a binary integer, we need to repeatly perform the integer division operation with the divisor equal to 2 until the quotient becomes 0. Afterwards, the binary number will be a concatenation of the remainders in reverse order. For instance, let us convert 52_{10} to binary.

$$\begin{aligned}
52 &= 2(26) + 0 \\
26 &= 2(13) + 0 \\
13 &= 2(6) + 1 \\
6 &= 2(3) + 0 \\
3 &= 2(1) + 1 \\
1 &= 2(0) + 1
\end{aligned}$$

Hence, $52_{10} = 110100_2$.

Now, converting the fractional portion has some similarities to the integer conversion process in that in can be expressed the fraction in notation as a sum of products of base 2.

$$m = \sum_{i=1}^{n} d_i \times 2^{-1 \times i}$$

where n is the total number and $d_1, d_2, \ldots, d_n$. However, it requires an additional step in the cycle and it may not be a finite process. In general, if the fractional portion cannot be written as a fraction whose denominator is a power of 2, the process will be infinite. This means you need to choose when you have a sufficient number of significant digits. First, if a number, m, is between 0 and 1 exclusively $(0 < n < 1)$, the product of n times 2 will be between 0 and 2 exclusively $(0 < 2n < 2)$. Hence, the integer portion of the product will be either 0 or 1. Moreover, the integer portion of the product is current digit in binary. For instance,

$$\begin{aligned}
m &= \sum_{i=1}^{n} d_i \times 2^{-1 \times i} \\
2 \times m &= 2 \times \sum_{i=1}^{n} d_i \times 2^{-1 \times i} \\
&= \sum_{i=1}^{n} d_i \times 2^{-1 \times i + 1} \\
&= d_1 + \sum_{i=2}^{n} d_i \times 2^{-1 \times i}
\end{aligned}$$

From the above example, we see that we will need to subtract the current digit to determine the remaining digits as follows

$$\begin{aligned}
2 \times m &= d_1 + \sum_{i=2}^{n} d_i \times 2^{-1 \times i} \\
2 \times m - d_1 &= d_1 + \sum_{i=2}^{n} d_i \times 2^{-1 \times i} - d_1 \\
&= \sum_{i=2}^{n} d_i \times 2^{-1 \times i}
\end{aligned}$$

Hence, by repeating the process of multiplying the fractional decimal by 2 and subtracting the integer portion of the product, we can derive the binary representation (or an approximation). Now, let us convert $0.625(\frac{5}{8})$.

$$\begin{aligned}
2 \times 0.625 &= 1.25 \\
2 \times 0.25 &= 0.5 \\
2 \times 0.5 &= 1.0
\end{aligned}$$

Therefore, $0.625_{10} = 0.101_2$.

Hexadecimal Notation

Binary is an essential number system for computers since computers function on electrical signals that are either on or off. However, for a human, dealing with binary will become cumbersome quickly because it requires a significant number of digits to represent a number. Hence, we need a means of converting it to another number system that is easier to deal with. Furthermore, the conversion process between the binary system and the system must be simple. Although dealing with the decimal system is innately easy, converting decimal to binary is not that straightforward. So the *hexadecimal system* was adopted. The hexadecimal system is a positional number system whose base is 16. It digits are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E and F where A, B, C, D, E and F represents 10, 11, 12, 13, 14 and 15 respectively.

Fortunately, binary digits grouped into sets of four, called a *nibble*, can be converted to a single hexadecimal digit. The following table provides all combinations

$0000 = 0$	$0100 = 4$	$1000 = 8$	$1100 = C$
$0001 = 1$	$0101 = 5$	$1001 = 9$	$1101 = D$
$0010 = 2$	$0110 = 6$	$1010 = A$	$1110 = E$
$0011 = 3$	$0111 = 7$	$1011 = B$	$1111 = F$

Therefore, to convert a binary number to a hexadecimal number, you need to group the digits into nibbles (for the integer portion you start with the rightmost digit and move left; whereas, for the fractional portion you start with the leftmost digit and move right) and change to their respective hexadecimal digit. And to convert a hexadecimal number to a binary number, you need to change each digit to its respective nibble. For instance,

$$\begin{aligned}
001110100101_2 &= 3A5_{16} \\
10010.1101_2 &= 12.D_{16} \\
A68_{16} &= 111001101000_2 \\
3C.2C_{16} &= 00111100.00101100_2
\end{aligned}$$