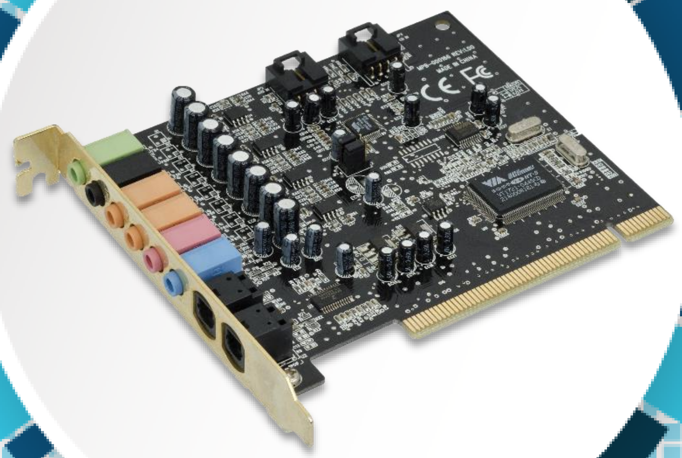

计算机组成原理

第二章 数据表示

2.2 定点与浮点数据表示

2.2 定点与浮点数据表示

1 定点数据表示

01 可表示定点小数和整数

02 表现形式：$X_0.X_1X_2X_3X_4 \ldots\ldots X_n$

定点小数　　定点整数

03 定点小数表示数的范围(补码为例)： $-1 \le x \le 1-2^{-n}$

04 定点整数表示数的范围(补码为例)：$-2^{n} \le x \le 2^{n}-1$

05 定点数据表示数的不足：数据表示范围受限

2 浮点数据表示

把数的范围和精度分别表示的一种数据表示方法。

浮点数的使用场合

当数的表示范围超出了定点数能表示的范围时。

(1)格式(一般格式)

E_S	$E_1E_2E_3\ldots\ldots E_n$	M_S	$M_1M_2M_3M_4..M_k$

$\longrightarrow$ $N=2^e \cdot m$

E: 阶码位数,决定数据的范围

M: 尾数位数,决定数的精度

2 浮点数据表示

例1

将$x = 2^{-01} \times (-0.1110)$ 表示成机器形式。假定用8位表示该数，且阶码占3位，位数占5位（均包含一位符号位）。

解：假定阶码和尾数均采用补码

E_S	$E_1E_2E_3\ldots\ldots E_n$	M_S	$M_1M_2M_3M_4..M_k$

1	1	1	1	0	0	1	0

2 浮点数据表示

一般表示格式的不足

E_S	$E_1E_2E_3……E_n$	M_S	$M_1M_2M_3M_4..M_k$

1 1 1 1 0 0 1 0

不同系统可能根据自己的浮点数格式从中提取不同位数的阶码

2 浮点数据表示

(2)IEEE 754格式

S	8位偏指数E	23位有效尾数M	单精度
S	11位偏指数E	52位有效尾数M	**双精度**

- 指数采用偏移值，其中单精度偏移值为127，双精度为1023，将浮点数的阶码值变成非负整数,便于浮点数的比较和排序。
- IEEE754尾数形式为1.XXXXXX,其中M部分保存的是XXXXXX(1被隐藏)，从而可保留更多的有效位，提高数据表示的精确度。

2 浮点数据表示

S	8位偏指数E	23位有效尾数M	单精度

与上述IEEE754格式相对应的32位浮点数的真值可表示为:

$$N = (-1)^{S} \times 2^{E-127} \times 1.M$$

随E和M的取值不同，IEEE754浮点数据表示具有不同的意义

2 浮点数据表示

$N = (-1)^S \times 2^{E-127} \times 1.M$

E=0 , M =0 ：表示机器零；

E=0 , M ≠0 ：则$N = (-1)^S \times 2^{-126} \times 0.M$,非规格化的浮点数；

$1 \le E \le 254$ ：$N = (-1)^S \times 2^{E-127} \times 1.M$ ，规格化的浮点数；

E=255 , M =0 ：无穷大的数，对应于x / 0 (其中x ≠0) ；

E=255 , M ≠0 ：N= NaN，表示一个非数值，对应于0 / 0。

http://www.cs.berkeley.edu/~wkahan/

关于浮点异常，见Kahan教授的《Lecture Notes on IEEE 754》

2 浮点数据表示

IEEE754 32位浮点数与对应真值之间的变换流程

浮点数真值对应的十进制数N

⇩

将N变换成$(-1)^s \times 2^e \times 1.M$

⇩

E = e + 01111111

⇩

保存S、E、M

⇩

单精度IEEE754

⇧

从32位二进制序列中分离出S、E、M

⇧

e = E - 01111111

⇧

将 S、e、M代入 $(-1)^s \times 2^e \times 1.M$，并按权展开

⇧

浮点数真值对应的十进制数N

2 浮点数据表示

例2

将十进制数20.59375转换成32位IEEE754格式浮点数的二进制格式。

解: 先将十进制数换成二进制数：

20.59375=10100.10011

移动小数点，使其变成1.M的形式

$10100.10011=1.010010011\times 2^{4}$

得到：

S=0, e = 4，E= 100+01111111 =10000011，M = 010010011

最后得到32位浮点数的二进制存储格式为：

0100 0001 1010 0100 1100 0000 0000 0000

= 41A4C000H