

### 大多数的浮点数都遵循单精度或双精度的IEEE浮点标准。

■ 标准浮点数字长由一个符号位S,指数e和无符号(小数)的规格 化尾数m构成,如下所示。

S 指数e 无符号尾数m
--------------

#### ■ 浮点数可以用下式描述:

$$X = (-1)^{S} 1.m \cdot 2^{e-bias}$$

### 浮点数表示

#### 对于IEEE-754标准来说,还有下面的约定:

- □ 当指数e=0, 尾数m=0时, 表示0;
- □ 当指数e=255, 尾数m=0时, 表示无穷大;
- □ 当指数e=255, 尾数m≠0时, 表示NaN (Not a Number, 不是一个数)。

## 浮点数表示

### IEEE的单精度和双精度格式的参数

	单精度	双精度		
字长	32	64		
尾数	23	52		
指数	8	11		
偏置	127	1023		

### 在C语言中,提供了两种实数的表示方法,包括:

- 十进制表示法
  - 口 由数字0~9以及小数点组成,例如: 0.0、25.0、5.789、0.13、5.0、300.、-267.8230。

注: 在使用十进制表示浮点数时,必须包含小数点。

- 指数形式表示法
  - 口 由十进制数字、阶码标志(小写字母 "e"或大写字母 "E"),以及 阶码(只能为整数,可以带符号)组成。一般形式为:

a E n

□ 其中:

a和n均为十进制数; 其表示的指数为:

 $a \times 10^n$ 

■ 在标准的C语言中,将按照所能表示的数的动态范围和精度,将实数进一步的分成单精度实数、双精度实数和长双精度三种,分别用float、double和long double关键字声明这三种类型的实数。它们所分配的存储字长和表示数的范围不同。

#### 注:

- (1) 对于单片机来说, double和浮点类型相同。
- (2) 可以在具体的数值后面加后缀字母 "f" 表示该数为单精度浮点数。

#### 单精度、双精度和长双精度实数的字长及表示的范围

类型说明符	比特数 (字节数)	有效数字	数的范围	
float	32(4)	6~7	10 <sup>-37</sup> ~10 <sup>38</sup>	
double	64(8)	15~16	10 <sup>-307</sup> ~10 <sup>308</sup>	
long double	128(16)	18~19	10-4931~104932	

#### 【例】浮点数声明的例子

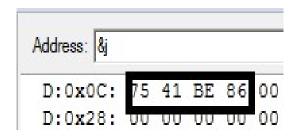
```
void main()
{
    float i=100.00,j=245.6e30;
}
```

下面对该例子进行分析,分析步骤主要包括:

- 1) 进入本书所提供资料的STC\_example\例子8-2\目录下,在Keil μVision5 集成开发环境下打开该设计。
- 2) 在集成开发环境主界面主菜单下,选择Debug->Start/Stop Debug Session。
- 3) 在调试器模式下,单步运行该程序,一直到程序的末尾。

■ 在Memory 1窗口界面内,分别输入&i和&j,可以看到将浮点变量i的值100.0分配到单片机片内数据区地址为0x08开始的位置,该变量占用四个存储字节空间,其值用十六进制数表示为0x42C80000; 类似地,可以看到将浮点变量j的值245.6e30分配到单片机内数据区地址为0x0C开始的位置,该变量占用四个存储字节空间,其值用十六进制数表示为0x7541BE86。

Memory 1					
Address: &i					
D:0x08:	42	C8	00	00	7F 80
D:0x24:	00	00	00	00	00 00
I					





■ 对于浮点数100.00来说,在计算机中存储的数0x42C80000。 对应的二进制数表示为:

#### 其中:

- 口 0: 表示符号位,表示当前是正数;
- 口 100,0010,1:表示阶数,对应的十进制数为133。在浮点标准中,这个值已经加上了偏移量127,所以实际的阶数为133-127=6,对应于2^6=64,即表示的是2的幂次方。
- 100, 1000, 0000, 0000, 0000; 表示尾数, 对应的十进制小数为0.5625。因为总是隐含1。所以,表示的小数实际值为1.5625。

因此,

$$1.5625 \times 2^6 = 1.5625 \times 64 = 100$$

### 数据类型

### --实数型

■ 对于浮点数245.6e30来说,在计算机中存储的数0x7541BE86。 对应的二进制数表示为:

#### 其中:

- 口 0: 表示符号位,表示当前是正数;
- 111,0101,0:表示阶数,对应的十进制数为234。在浮点标准中,这个值已经加上了偏移量127,所以实际的阶数为234-127=107,对应于2^107=1.6226×10^32,即表示的是2的幂次方。
- 100,0001,1010,1110,1000,0110:表示尾数,对应的十进制小数为0.51361083984375。因为总是隐含1,所以表示的小数实际值为1.51361083984375。

#### 所以:

1.  $5136108398 \ 4375 \times 2^{107} = 2.455974 \times 10^{32}$