FLOATING POINT FORMAT STANDARDS

There is an internationally agreed standard for mini and microcomputers produced by the Institute of Electrical & Electronics Engineers, **IEEE - 754-1985**.

This standard is of particular interest in the PC world as it is used by Intel in their processors.

The standard defines three formats - Single, Double and Quad.

The principal features are given in the following table.

FORMAT	SINGLE	DOUBLE	QUAD
Sign Bit	1	1	1
Exponent	8	11	15
Significand	23	52	112
Total Bits	32	64	128

Sign bit

Gives the sign of the number 0 represents positive, and 1 represents negative.

Exponent

This is in a format known as biased form, which is an alternative way of representing negative numbers, although the general method would also work with two's complement.

Significand

This is stored in a variation on normalised form. Instead of the number being normalised so that it takes the form:

`0.1bbbbbbb`

it is normalised so that it takes the form

`1.bbbbbbbbb`

and the most significant **`1.`** is not stored. It will be added by the circuitry when the number is manipulated.

When stored in memory, a single precision number would appear as :-

S	Exponent	Significand
1	8 bits	23 bits