Notes on floating point

(1) Excess notation

First we need yet another signed integer representation. Idea is to use unsigned binary, but subtract off a fixed bias to get the actual number represented.

e.g. 3 bit excess 5: (the width or word size; the bias)

- excess 2: (for column)

	unsigned	excess 5	excess 2
111	6	1	4
101	5	0	3
100	4	-1	2
011	3	-2	1
010	2	-3	0
001	1	-4	-1
000	0	-5	-2

(2) Recall scientific notation for decimal numbers:

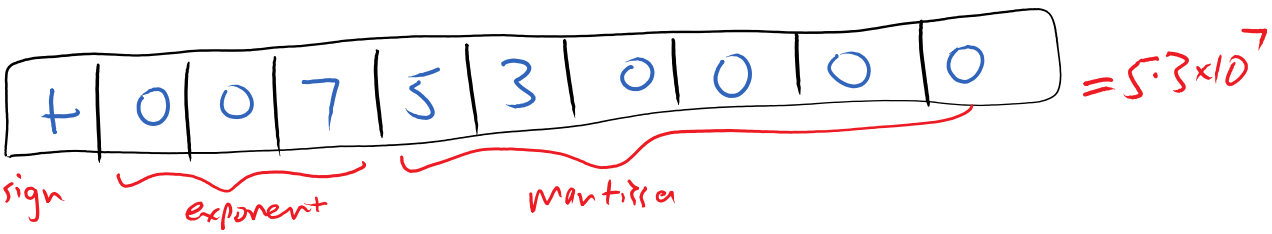
e.g. $+5.3 \times 10^{7} = 53,000,000$

$-6.743 \times 10^{-3} = -0.006743$

sign; significand or mantissa; exponent

for fixed size representation, could agree to use, e.g.

1 slot for sign, 3 slots for exponent, 6 slots for significand.



But 2 problems:

(A) 5.3×10^7 is same as 0.053×10^9. So could have used

+	0	0	9	0	0	5	3	0	0	0	0

This ambiguity is bad.

Therefore, insist numbers are normalized to have exactly one digit before decimal point. e.g. $0.053 \times 10^9 \longrightarrow 5.3 \times 10^7$.

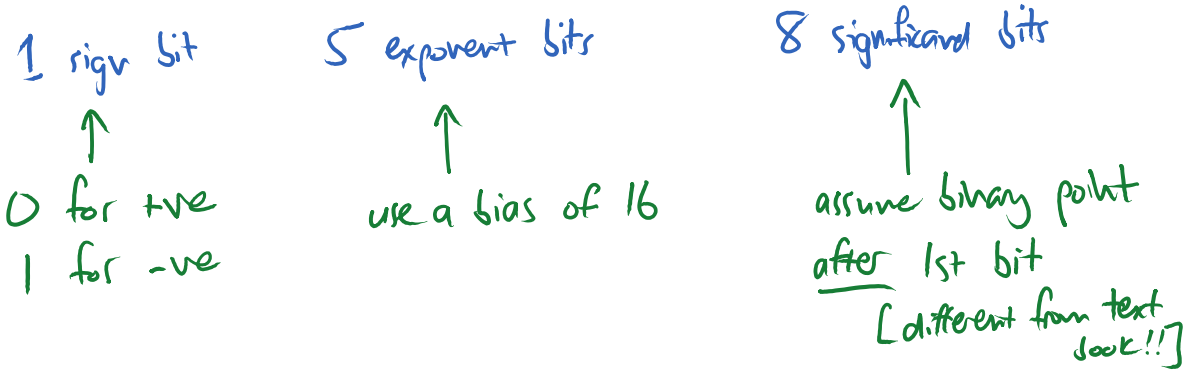
(B) No 'sign' slot for negative exponents. Designers could have chosen 2's complement or signed magnitude, but they didn't. Instead they use excess notation with a fixed bias.

e.g. with a bias of 500, $+8.945 \times 10^{-3}$ is written with exponent $-3+500 = 497$, obtaining:

+	4	9	7	8	9	4	5	0	0

In real computers we use binary but the ideas are the same.

e.g. one possible approach is:



e.g. $+17_{10} = +10001_2$

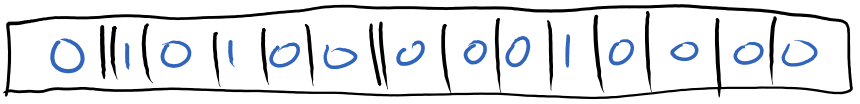
$= +1.0001_2 \times 2^{(4)}$ ← $4 + 16 = 20_{10} = 10100_2$ (bias)

→

0	1	0	1	0	0	1	0	0	0	1	0	0	0

Note: 1st bit of significand is always 1 (because of normalization).
So we can assume it is there, and not waste space by storing it.

With an implied bit, 17_{10} becomes



See book for actual IEEE standards
e.g. IEEE double: 64 bits, 11-bit exponent (bias 1023)
52-bit significand, implied bit before binary point.

Floating point arithmetic

- for addition: align binary points; add; truncate
- for multiplication: add exponents and multiply significands; truncate

[try examples of these in decimal to get the idea]

See also the suggested minilab exercises.