

05 – Floating Point Numbers

CS1022 – Introduction to Computing II

Dr Jonathan Dukes / jdukes@tcd.ie School of Computer Science and Statistics D.A.Patterson, J.L.Hennessy, "Computer Organisation and Design: ARM Edition", Morgan-Kaufmann, 2016.

Section 3.5: Floating Point

(Available in the Library, doesn't have to be the ARM Edition!!) 32-bits ... 2³² unique values that we can use to represent different things

e.g. unsigned integers

```
0 ... 2<sup>32</sup>-1 or
```

e.g. signed integers using 2's complement

$$-2^{31} \dots 0 \dots +2^{31} -1$$

How do we represent **real** numbers like 2½ or 3.14159265...?

How do we represent values with really large or really small magnitudes?

The values 2.2 x 10¹¹ and 1.3 x 10⁻⁸ are examples of (normalized) scientific notation in decimal form

$$f \times 10^e$$

Values expressed in normalised scientific notation satisfy the condition:

$$1 \le |f| < 10$$

Normalized scientific notation give us one *canonical form* in which to express a value using scientific notation and allows quick, visual comparison of magnitude

Generally, as computer scientists, we avoid expressing the same thing in different ways (a==b?)

Convert the following binary numbers to decimal numbers with fractions

```
10010101
1.1
101000.01
```

Convert the following decimal numbers to binary floating point numbers

10½

51/4

7.75

2.1

Like decimal values, we can express binary values using scientific notation (again, in normalized form)

e.g.

```
1010.1 = 1.0101 \times 2^{3}

0.00101 = 1.01 \times 2^{-3}
```

The general form is again:

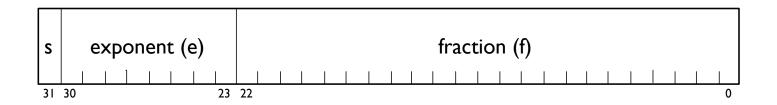
$$f \times 2^e$$

and in normalised form, f satisfies the following condition:

$$1_2 \le |f| < 10_2$$

The normalized form of a binary number expressed using scientific notation forms the basis for its representation in a computer

Use a different interpretation of a 32-bit value to represent floating point numbers, e.g. IEEE 754



$$(-1)^s \times f \times 2^e$$

How can we represent ...

... positive and negative values?

... values with positive and negative exponents?

Where is the binary (radix) point?

Sign bit?

0 ⇒ positive floating-point number

 $1 \Rightarrow$ negative floating-point number

Positive and negative exponents?

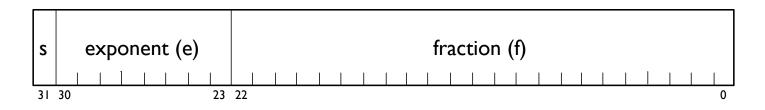
Option 1: 2's Complement exponents



Option 2: Biased exponents

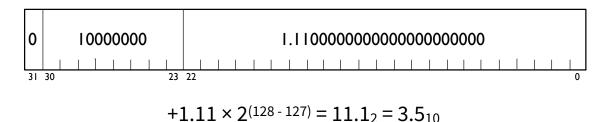


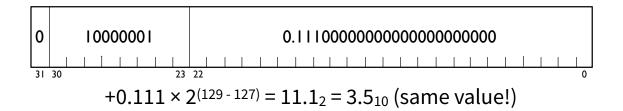
Subtract a constant bias (b = 127) from stored exponent to obtain signed exponent



$$(-1)^s \times f \times 2^{e-b}$$

The following two representations are of the same value (3.5_{10})





We don't want multiple representations of the same value!

if
$$(a == b)$$
.

Storing floating-point numbers in normalized form avoids this problem:

$$1 \le |f| < 2$$
, so f is in the form $1.ddddd...$

With normalisation

 0.0101×2^{-4}

... becomes ...

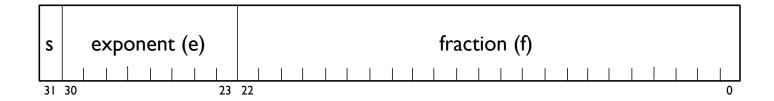
1.0100 x 2⁻⁶

adjust fraction so there is a single 1 to left of radix point compensate by adjusting exponent accordingly

If there is always going to be a 1 to the left of the radix point, we don't need to store it!

Increases precision (by one bit) – similar to not storing the 2 LSBs of a branch target offset!

Final IEEE 754 Floating-Point Representation

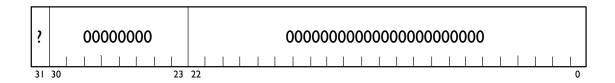


$$(-1)^s \times 1.f \times 2^{(e-b)}$$

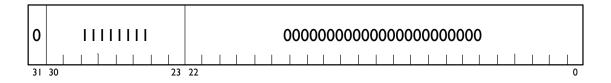
Examples?

Special bit patterns, e.g.

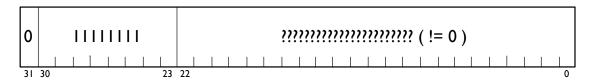
Zero (±)



Infinity (±)

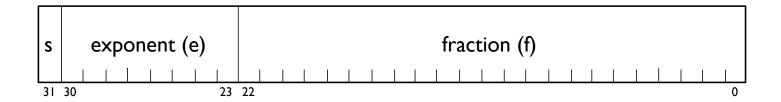


Not a Number (NaN)

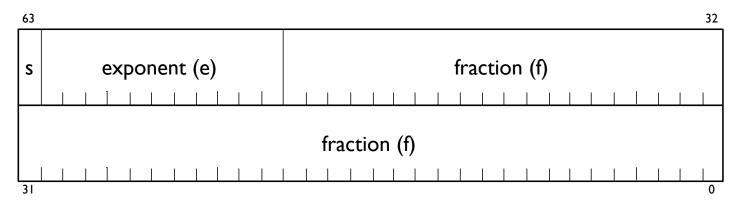


Single and Double Precision

32-Bit Single Precision (bias = 127)



64-Bit Double Precision (bias = 1023)



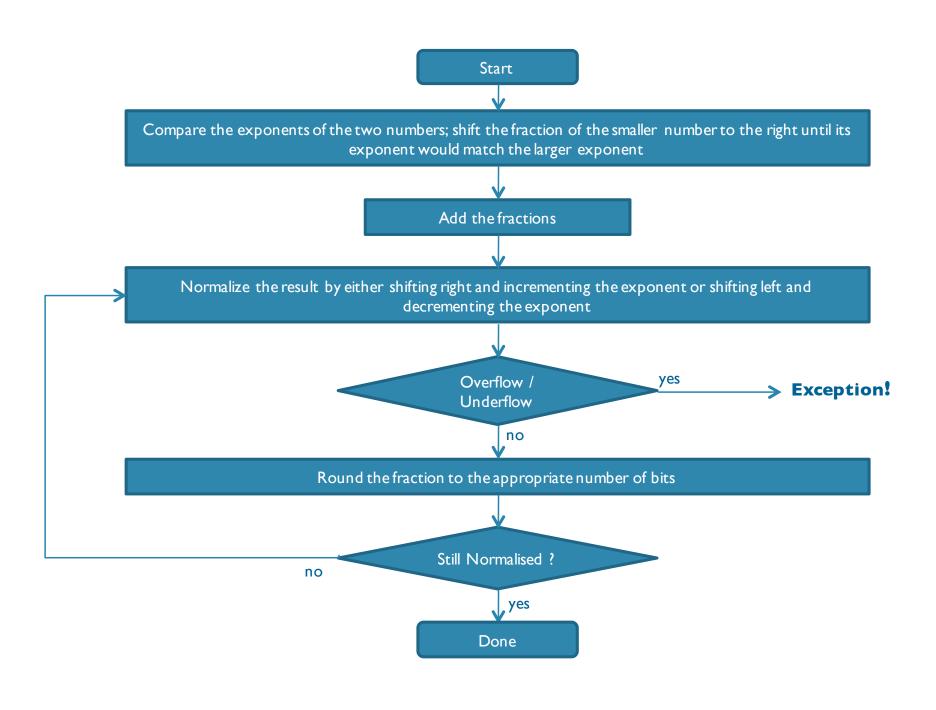
We can add the fractions of two floating point values if their exponents are the same

If their exponents are not the same to begin, shift the fraction of the value with the smaller exponent to compensate

```
e.g.
```

```
1.01101 \times 2^{3} + 1.00110 \times 2^{-2}
```

- $= 1.01101 \times 2^3 + 0.0000100110 \times 2^3$
- $= 1.0111000110 \times 2^{3}$



What about adding negative values (S==1)?

Proceed as before but before adding, the fractions of values with S==1 should be converted to their 2's Compliment

A Closer Loo Warning: Intel Inside

San Francisco Chronicle

BUSINESS

Intel's Pentium Problem Persica

TALKING BUSINESS

The yield curve has had so many contractions that it about to give birth. The birth ... might be recession.

Intel to work on software patch for Pentium bug

will work around the i

Grove says l sorry about

argest chipmaker.

largest chipmaker.

The problem, fixed months ago, received national attention Thursday after two weeks of discussion by researchers and mathematicians on the Internet, the global computer network, and some coverage in high-technology induces readility at the control of the c

industry publications.

Reacting for the first time, investors sent Intel's stock down 2 percent, or 1¼, to 63% on the Nasdaq stock market, which was closed for Thanksgiving on Thurs-

of a long-term problem for the Santa Clara company.

"Interesting but irrelevant," said David Wu, an analyst with S.G. Warburg. "These things are quickly forgotten. People won't remember it by Mooday."

The flaw occurred in early ver-

sions of the Pentium. Once every few billion calculations, the chip might produce a wrong answer to



Pentium bu

INTEL'S MISTAKE

Flawed Chip **Bruises Intel**

Investors react, stock plunges

Intel Corp. stock dropped sharply yesterday after a minor flaw in the company's Pentium mi-croprocessor triggered a rash of negative publicity for the world's

in certain equations

Pentium woes continue Faulty FPU flubs math Multithreading gets lost on PIOO systems

INTEL CORP. PROBABLY FOUND IT difficult to find

Photo finish: Although Photoshop outdoes Picture P Bug Dodge Booed has an im Intel Knocked For

Response On Flaw

SAN MATEO, CALIF. Amid criticism or man carly shadded a flaw in an early version of its Pentium microprocessor, Intel Corp's processor, and a processor proces Amid criticism of how it has nine digits beyond the deci-handled a flaw in an early wersion of its Pentium micro-arise only in obscure math

pany has claimed.

The division flaw occurs in normal spreadsheet calculations up to the third signifi
Intel claims. "We believe the cant digit, according to a re-

IBM to stop

Pentium PCs

shipping

ent it and doler said much of a aim for its

at the Uni-s Informa-ne patches L, Page 6F

Chip Shot

Computer Giants' War Over Flaw in Pentium Jolts the PC Industry

Who Is Twisting the Truth? Intel Stands by Product

I'd Be Totally Confused

By BART ZIEGLER and DON CLARK

ERSONAL TECHNOLOGY

i't Serving Who Bought ium Campaign

As IBM Halts Shipments

calling Intel's toll-free number (800-625-6680), even if you're not an MTI professor. Intel doesn't make it easy. But in a test I conducted over



Some Scientists Are Angry Over Flaw CITYLINE: 24How In Pentium Chip, and Intel's Response

Intel's Grove Airs Apology for Pentium Over the Internet

Not to Replace All Chips. Doesn't Silence Criticism

Patterson & Hennessy, "Computer Organisation and Design", 4th edition

TECHNOLOGY

But CEO, Defending Policy

"To owners of Pentium™ processor-based computers and the PC community: We at Intel wish to sincerely apologize for our handling of the recently publicized Pentium processor flaw. The Intel Inside ® symbol means that your computer has a microprocessor second to none in quality and performance. Thousands of Intel employees work very hard to ensure that this is true. But no microprocessor is ever perfect. What Intel continues to believe is technically an extremely minor problem has taken on a life of its own. Although Intel firmly stands behind the quality of the current version of the Pentium processor, we recognize that many users have concerns. We want to resolve these concerns. Intel will exchange the current version of the Pentium processor for an updated version, in which this floating-point divide flaw is corrected, for any owner who requests it, free of charge anytime during the life of their computer."

Andrew S. Grove, Craig R. Barrett, and Gordon E. Moore President, Vice President, and Chairman of the Board, respectively, of Intel

Patterson & Hennessy, "Computer Organisation and Design", 4th edition