CE 311K: Errors

Krishna Kumar

University of Texas at Austin

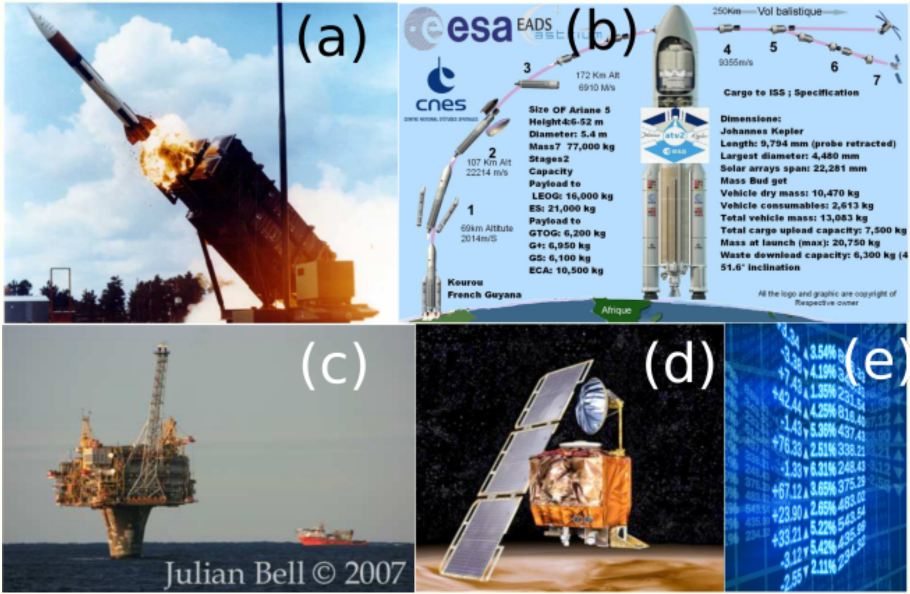
krishnak@utexas.edu

September 22, 2019

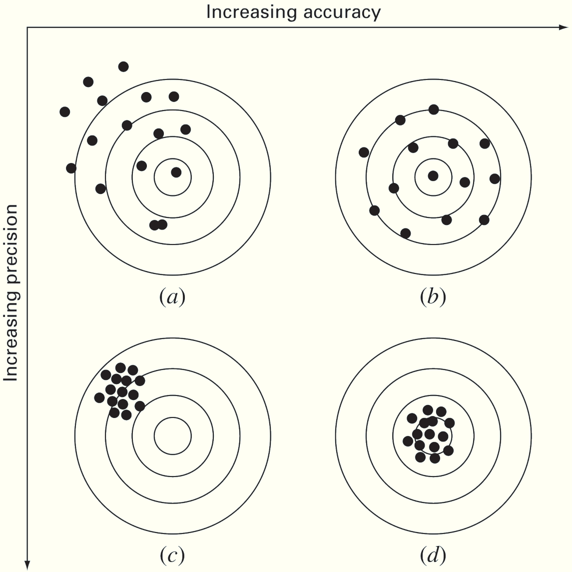
What causes errors?

"⚠ *the computer calculated it, so it must be right*"

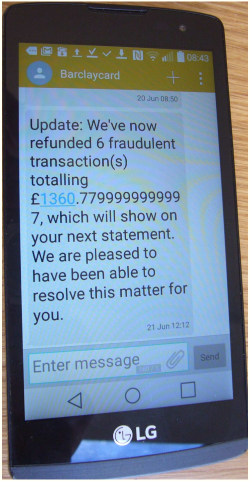
ⓘ The cost of errors



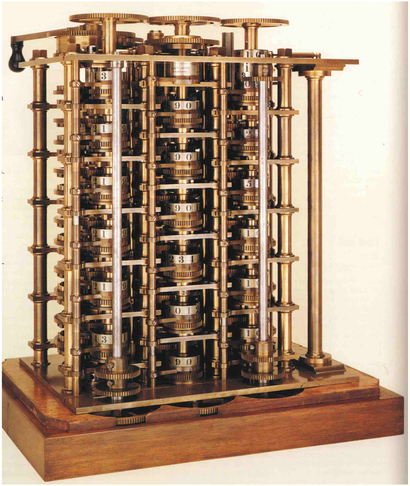
Quantifying errors: accuracy v precision



Significant figures

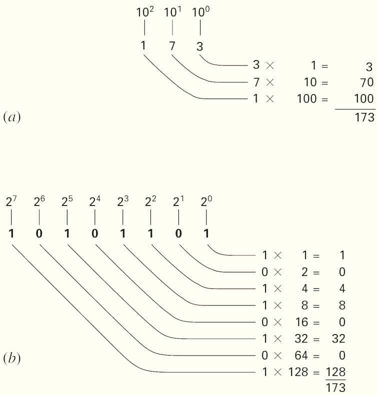


Decimal or binary



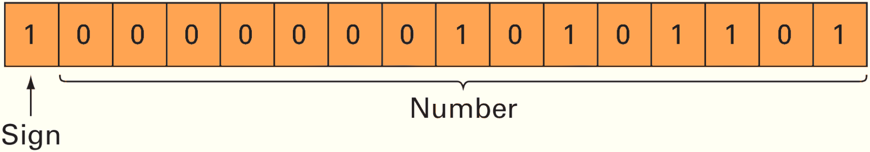
Charles Babbage's machine used base 10. CompScis decided a little later that base 2 is more funky.

Decimal or binary



Representing 173 in Decimal and Binary system

Bits and Bytes



Representing -173 on a 16-bit computer using the signed magnitude method.

Bit - **Bi**nary Digi**t**. 8 bits is a byte(?) ASCII needed 7 bits to represent all English alphabets.

The Gangam Style problem

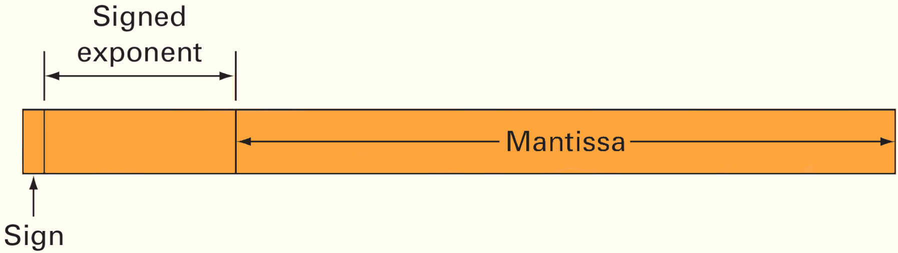
When 2,147,483,647 views of Gangam Style broke YouTube

Floating point representation

Fractional quantities are typically represented in computers using floating-point form. In this approach, the number is expressed as a fractional part, called a *mantissa* or *significand*, and an integer part, called an *exponent* or *characteristic*,

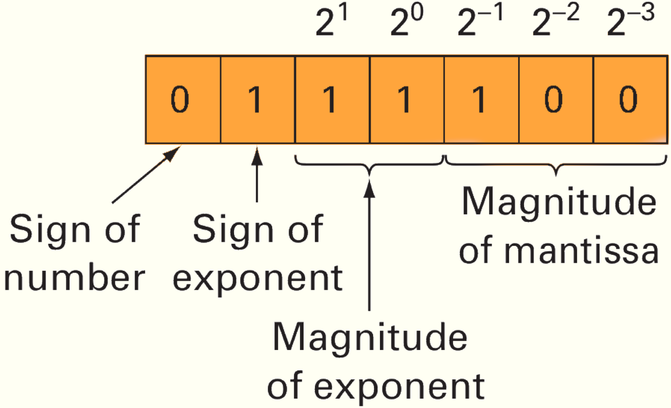
$$m.b^e$$

where m is the mantissa, b is the base of the number system being used, and e the exponent. For instance, the number 156.78 could be represented as 0.15678×10^3 in a floating-point base-10 system.



Smallest floating point for a 7-bit representation

Employ the first bit for the sign of the number, the next three for the sign and the magnitude of the exponent, and the last three for the magnitude of the mantissa:



Numerical errors

Numerical errors arise from the use of approximations to represent exact mathematical operations and quantities.

? Quantifying errors

Suppose that you have the task of measuring the lengths of a bridge and a rivet and come up with 9999 and 9 cm, respectively. If the true values are 10,000 and 10 cm, respectively. Quantify the errors.

Quantifying errors

Absolute and relative errors

Let x be a real number. We will use x^* to denote its approximation. We define two ways of measuring error introduced by this approximation.

- Absolute error:

- Relative error:

The relationships (WARNING: Severe abuse of notation!) $x^* = x \pm \varepsilon_x$ and $x^* = x(1 \pm \eta_x)$ are also commonly used to mean that x^* may take any value in the interval $[x - \varepsilon_x, x + \varepsilon_x]$.

Error accumulation

Let

$$x^* = x \pm \varepsilon_x \quad \text{and} \quad y^* = y \pm \varepsilon_y$$

Adding these two yields:

Exercise: What about subtraction?

Error accumulation: Loss of significance

Beware: when addition or subtraction causes partial or total cancellation, the relative error of the result can be much larger than that of the operands. We call this **loss of significance**.

For example, consider we store values to 3 significant digits and we take the innocent-looking $x = 9.99$, $y = 9.98$.

Error accumulation

This gets even worse when the loss of significance happens in a fraction's denominator. Consider an extension of the previous example:

$$\frac{1}{x - y}$$