

CE 311K: Errors

Krishna Kumar

University of Texas at Austin

krishnak@utexas.edu

September 22, 2019

1 Errors

2 Bit representation

What causes errors?

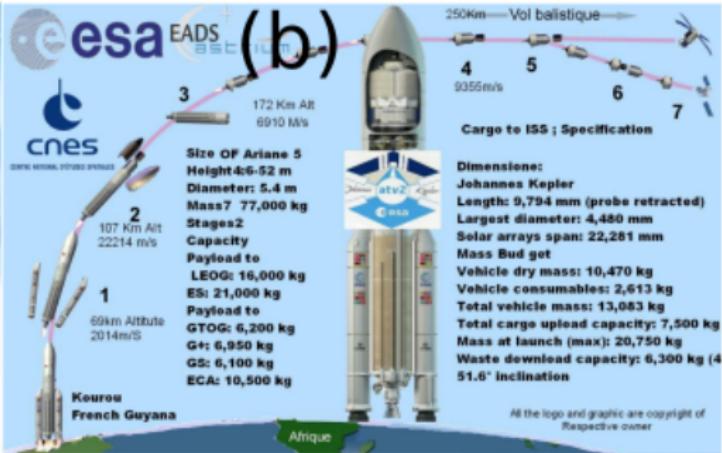
“⚠ the computer calculated it, so it must be right”

- the wrong mathematical model of reality (most subject areas lack models as precise and well-understood as Newtonian gravity)
- a parameterised model with parameters chosen to fit expected results ('over-fitting')
- the model being very sensitive to input or parameter values
- the discretisation of the continuous model for computation
- the build-up or propagation of inaccuracies caused by the finite precision of floating-point numbers
- plain old programming errors

⑧ The cost of errors



(a)



(c)



Julian Bell © 2007

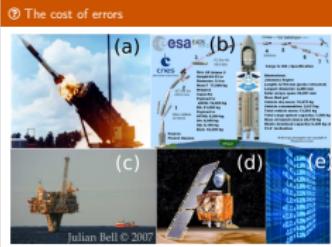


(d)



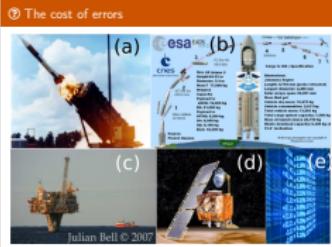
(e)

└ ? The cost of errors



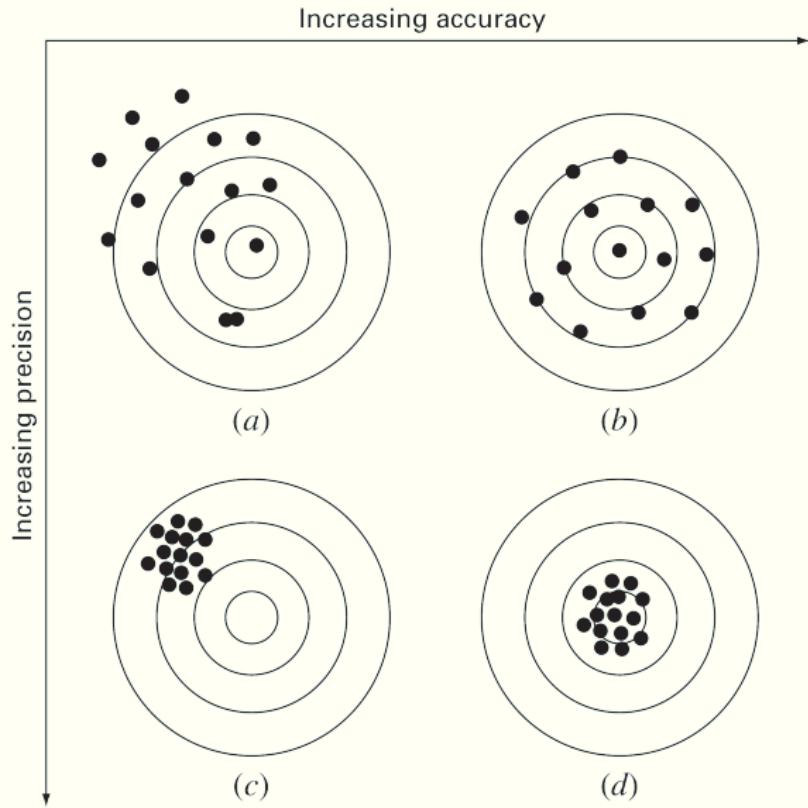
1. In 1991, a US Patriot missile failed to intercept an Iraqi Scud missile at Dhahran in Saudi Arabi, leading to a loss of life.
2. 1996 to a European Space Agency Ariane 5 unmanned rocket exploding shortly after lift-off. The rocket payload, worth US\$500 Million, was destroyed.
3. Sleipner A offshore platform sprang leak and sank on 23 August 1991. Post-accident investigation traced the error to inaccurate finite element approximation of the linear elastic model of the tricell (using the popular FDTD simulator NASTRAN). The shear stresses were underestimated by 47%. The failure involved a total economic loss of about \$700 million and causing a Richter scale 3.0 earthquake.

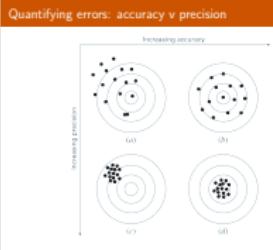
└ ? The cost of errors



1. In 1998 NASA lost its \$125 million Mars Climate Orbiter spacecraft as a result of a mistake that would shame a first-year physics student failing to convert Imperial units to metric. "*Propulsion people talk in pound-seconds of thrust and navigators talk in newton-seconds*"
2. Knight Capital Group lost \$440 million in 30 minutes on Aug 1, 2012 due to bug in their trading algorithms reportedly started pushing erratic trades through on nearly 150 different stocks

Quantifying errors: accuracy v precision



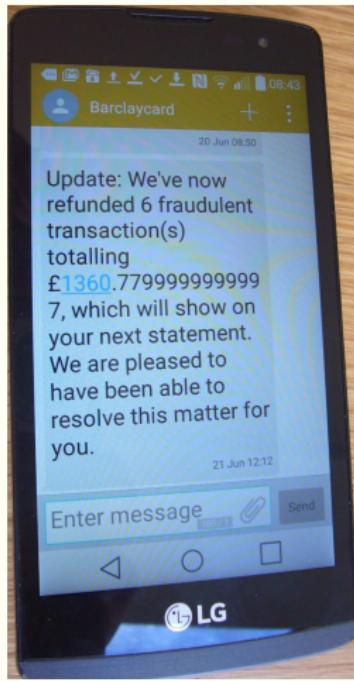


The errors associated with both calculations and measurements can be characterized with regard to their accuracy and precision.

Accuracy refers to how closely a computed or measured value agrees with the true value. Inaccuracy (also called bias) is defined as systematic deviation from the truth.

Precision refers to how closely individual computed or measured values agree with each other. Imprecision (also called uncertainty), on the other hand, refers to the magnitude of the scatter.

Significant figures

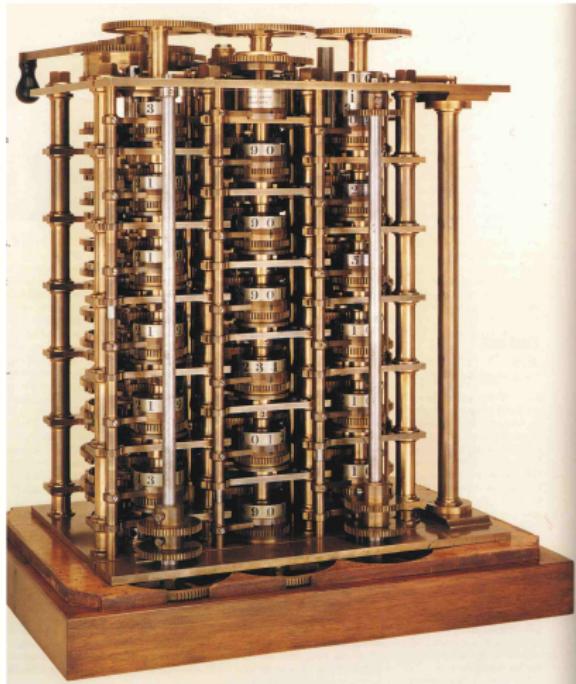


RAJ MALABAR 1 CASTLE STREET CAMBRIDGE CB2 0AH PHONE 01223312569 www.rajmalabar.com		DATE 20/11/2007	
QTY	DESCRIPTION	PRICE	AMOUNT
1	King Fisher PT	2.75	2.75
1	King Fisher PT	2.75	2.75
2	Bitter PT	2.5	5
1	Seafood Biriyani	9.99	9.99
1	Chappathi	1.48999	1.48999
	Kerala Lamb Curry	8.28999	8.28999
	Porotta	2.49	2.49
	Coca Cola/ Diet Co	1.29	1.29
	Sweet/Salty Lassi	2.25	2.25
	Kerala Lamb Curry	8.28999	8.28999
	Mon Rice	3.49	3.49
	Coca Cola/ Diet Co	1.29	1.29
	Chicken Korma	7.99	7.99
	Butter Rice	3.49	3.49

1 Errors

2 Bit representation

Decimal or binary



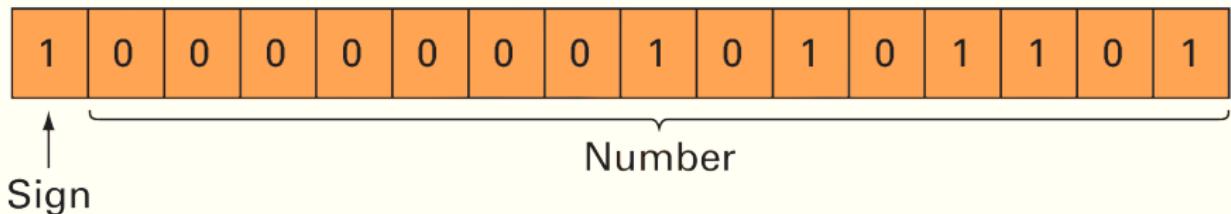
Charles Babbage's machine used base 10. CompScis decided a little later that base 2 is more funky.

Decimal or binary

$$\begin{array}{ccccccc}
 & 10^2 & 10^1 & 10^0 & & & \\
 & | & | & | & & & \\
 & 1 & 7 & 3 & & & \\
 & & \swarrow & \searrow & & & \\
 & & 3 \times & & 1 = & & 3 \\
 & & 7 \times & & 10 = & & 70 \\
 & & 1 \times & & 100 = & & 100 \\
 & & & & \hline & & 173
 \end{array}$$

Representing 173 in Decimal and Binary system

Bits and Bytes



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

Bit - Binary Digit. 8 bits is a byte(?) ASCII needed 7 bits to represent all English alphabets.

The Gangam Style problem



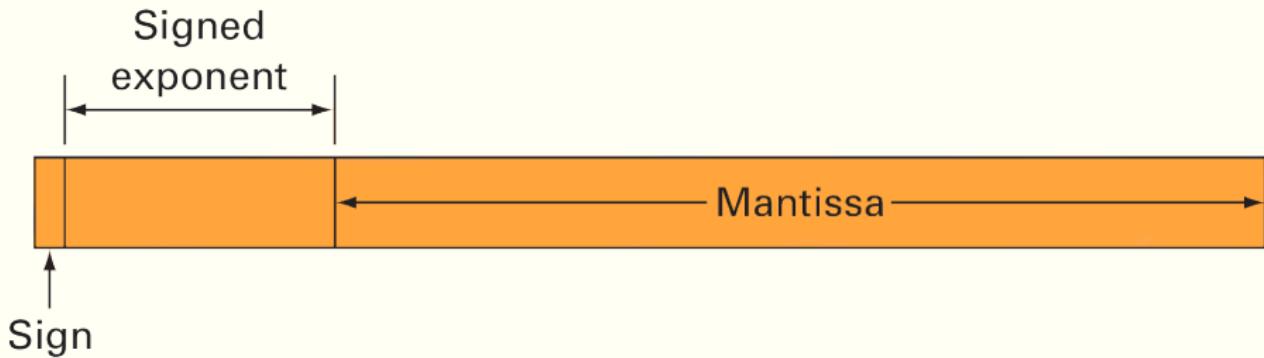
When 2,147,483,647 views of Gangam Style broke YouTube

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.



CE 311K: Errors

└ Bit representation

└ Floating point representation

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**,

$m \cdot b^e$

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 0.15678×10^4 in a floating-point base-10 system.



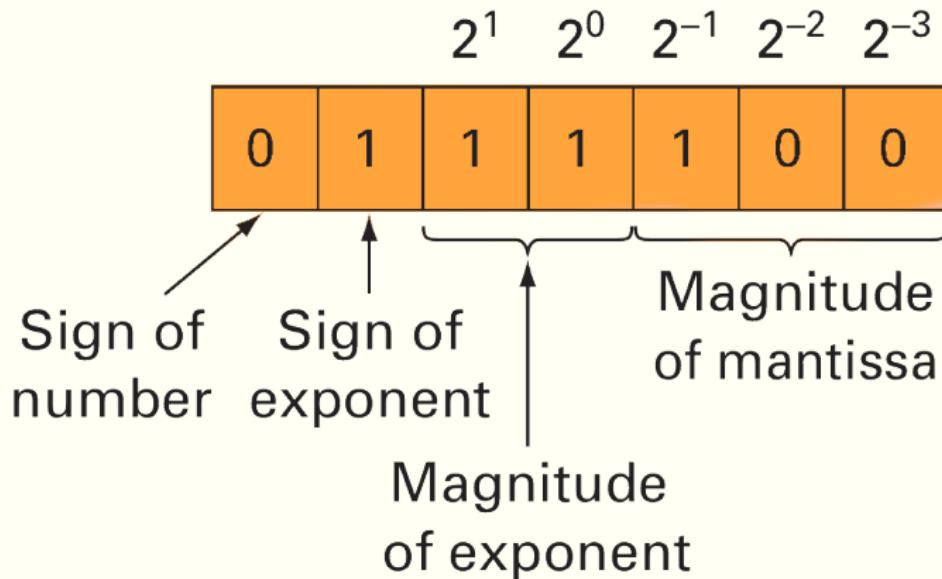
The absolute value of m is limited:

$$\frac{1}{b} \leq m < 1$$

where b the base. For example, for a base-10 system, m would range between 0.1 and 1, and for a base-2 system, between 0.5 and 1.

Smallest floating point for a 7-bit representation

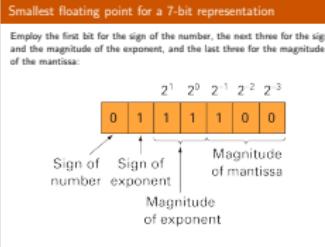
Employ the first bit for the sign of the number, the next three for the sign and the magnitude of the exponent, and the last three for the magnitude of the mantissa:



CE 311K: Errors

└ Bit representation

└ Smallest floating point for a 7-bit representation



The initial 0 indicates that the quantity is positive. The 1 in the second place designates that the exponent has a negative sign. The 1's in the third and fourth places give a maximum value to the exponent of $1 \times 2^1 + 1 \times 2^0 = -3$.

Finally, the mantissa is specified by the 100 in the last three places, which conforms to: $1 \times 2^1 + 0 \times 2^{-2} + 0 \times 2^{-3} = 0.5$. Thus, the smallest possible positive number for this system is $+0.5 \times 2^{-3}$, which is equal to 0.0625 in the base-10 system.