

A bit on the bit

The bit is a proper unit of measurement and should be recognized as such, argues Iulia Georgescu.

According to the guide to the International System of Units (SI) issued by the National Institute of Standards and Technology (NIST)¹, there are a “few highly specialized units” that are nonetheless still “acceptable for use with the SI”. One of them is the bit — the unit of information. I would like to challenge this rather unfair treatment of one of the most important units today, the very one that defines our information society.

The bit (a contraction of ‘binary digit’), also known as the shannon, quantifies the information capacity of a two-state system: 0/1, true/false, yes/no, on/off. The term was coined by John Tukey at Bell Labs in 1947, and first used formally by Claude Shannon a year later in his seminal paper laying the foundations of information theory². In general, the amount of information that can be stored in a classical system with N states is $\log_b N$, where the logarithmic base b represents the choice of a unit for measuring information. For $b = 2$ one has the bit, for $b = 3$ the trit, for $b = 10$ the hartley or ban or decimal digit, and for $b = e$, the base of the natural logarithm, one has the nat. The bit is the best known among these units and also the ‘smallest’ in terms of the logarithmic base defining it.

For practical purposes the bit is a bit too small, so people usually think in terms of chunks of eight bits, or bytes. An array of eight bits can represent 2^8 values, enough to encode alphanumeric characters, punctuation and more. The byte (B), which got its name from Werner Buchholz at IBM in 1956, has become the standard across computer architectures and is used to characterize the size of the active memory or storage capacity of computing devices.

The beauty of the bit is that it is not just a mathematical concept. It reflects a real physical quantity, and I am not necessarily referring to the memory size of your hard drive. Information itself is physical. Rolf Landauer was the first to observe that information is not a purely abstract concept, but that it is always tied to a physical representation and is therefore inevitably subject to the laws of physics — in particular, the laws of

thermodynamics³. Logical irreversibility means thermodynamic irreversibility, that is, erasing a classical bit of information requires a small amount of heat to be dissipated. The smallest amount of heat needed to erase a bit is known as the Landauer limit, $kT \log_2 2$, which is proportional to the Boltzmann constant k and the temperature T of the system. Landauer’s principle has been verified in various experiments, but current computing devices operate far from the Landauer limit due to heat dissipation from inherent inefficiencies.

Far from being a “highly specialized unit”, the bit has become one of the most common units around. Every day we upload and download megabytes to gigabytes of data from and to our computers and mobile devices. We complain about the slow internet

connection speed, in megabits per second, and we may judge the performance of the next mobile phone in terms of its memory capacity: when is 64 GB ever enough? Units of measurement have been historically linked to economics and trading. The bit now underlies electronic commerce from electronic goods in the form of music, films, e-books or software to large databases containing more or less sensitive information and digital currencies. Gigabytes of data can now be more valuable than kilograms of grain or metres of silk.

Another reason why the bit as a unit should be taken more seriously is the extremely fast increase in the amount of information produced by humankind, sometimes referred to as the information explosion, which will clearly have a significant social, economic and cultural impact. To get a feeling of this data deluge recall that high-energy physics experiments, such as the Large Hadron Collider, produce tens of petabytes ($1 \text{ PB} = 10^{15} \text{ B}$) of data each year. When fully functional, the Square Kilometer Array radio telescope is expected to generate an exabyte (10^{18} B) of data per day. And by the end of this year the global internet traffic will exceed one zettabyte (10^{21} B). But if you worry about running out of computer memory for this astonishing amount of data, fear not: Seth Lloyd estimated the information capacity of the Universe to be 10^{90} – 10^{120} bits⁴. □

IULIA GEORGESCU is a senior editor at *Nature Physics*.

References

1. Thompson, A. & Taylor, B. N. *The NIST Guide for the use of the International System of Units* Ch. 5 (2009); <http://go.nature.com/2akL1Pn>
2. Shannon, C. E. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
3. Landauer, R. *IBM J. Res. Dev.* **5**, 183–191 (1961).
4. Lloyd, S. *Phys. Rev. Lett.* **88**, 237901 (2002).

