WIKIPEDIA

Units of information

In computing and telecommunications, a **unit of information** is the capacity of some standard data storage system or communication channel, used to measure the capacities of other systems and channels. In information theory, units of information are also used to measure the entropy of random variables and information contained in messages.

The most commonly used units of data storage capacity are the bit, the capacity of a system that has only two states, and the byte (or octet), which is equivalent to eight bits. Multiples of these units can be formed from these with the SI prefixes (power-of-ten prefixes) or the newer IEC binary prefixes (power-of-two prefixes).

Contents

Primary units

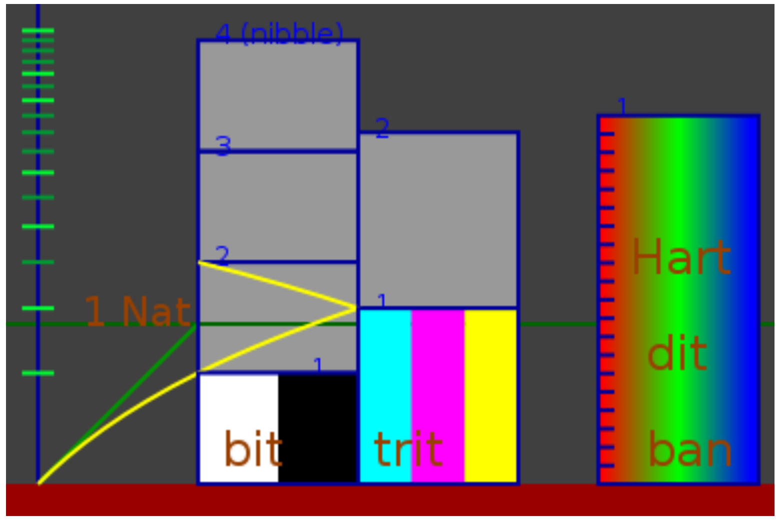
In 1928, Ralph Hartley observed a fundamental storage principle,[1] which was further formalized by Claude Shannon in 1945: the information that can be stored in a system is proportional to the logarithm of N possible states of that system, denoted $\log_b N$. Changing the base of the logarithm from b to a different number c has the effect of multiplying the value of the logarithm by a fixed constant, namely $\log_c N = (\log_c b) \log_b N$. Therefore, the choice of the base b determines the unit used to measure information. In particular, if b is a positive integer, then the unit is the amount of information that can be stored in a system with N possible states.

When b is 2, the unit is the shannon, equal to the information content of one "bit" (a portmanteau of binary digit[2]). A system with 8 possible states, for example, can store up to $\log_2 8 = 3$ bits of information. Other units that have been named include:

- Base $b = 3$: the unit is called "trit", and is equal to $\log_2 3$ (≈ 1.585) bits.[3]

- Base $b = 10$: the unit is called *decimal digit*, *hartley*, *ban*, *decit*, or *dit*, and is equal to $\log_2 10$ (≈ 3.322) bits.[1][4][5][6]
- Base $b = e$, the base of natural logarithms: the unit is called a *nat*, *nit*, or *nepit* (from Neperian), and is worth $\log_2 e$ (≈ 1.443) bits.[1]

The trit, ban, and nat are rarely used to measure storage capacity; but the nat, in particular, is often used in information theory, because natural logarithms are mathematically more convenient than logarithms in other bases.



Comparison of units of information: bit, trit, nat, ban. Quantity of information is the height of bars. Dark green level is the "Nat" unit.

Units derived from bit

Several conventional names are used for collections or groups of bits.

Byte

Historically, a byte was the number of bits used to encode a character of text in the computer, which depended on computer hardware architecture; but today it almost always means eight bits – that is, an octet. A byte can represent 256 (2^8) distinct values, such as non-negative integers from 0 to 255, or signed integers from −128 to 127. The IEEE 1541-2002 standard specifies "B" (upper case) as the symbol for byte (IEC 80000-13 uses "o" for octet in French, but also allows "B" in English, which is what is actually being used). Bytes, or multiples thereof, are almost always used to specify the sizes of computer files and the capacity of storage units. Most modern computers and peripheral devices are designed to manipulate data in whole bytes or groups of bytes, rather than individual bits.

Nibble

A group of four bits, or half a byte, is sometimes called a nibble or nybble. This unit is most often used in the context of hexadecimal number representations, since a nibble has the same amount of information as one hexadecimal digit.[7]

Word, block, and page

Computers usually manipulate bits in groups of a fixed size, conventionally called *words*. The number of bits in a word is usually defined by the size of the registers in the computer's CPU, or by the number of data bits that are fetched from its main memory in a single operation. In the IA-32 architecture more commonly known as x86-32, a word is 16 bits, but other past and current architectures use words with 8, 9, 12, 18, 24, 26, 32, 36, 39, 40, 48, 56, 60, 64, 80 bits or others.

Some machine instructions and computer number formats use two words (a "double word" or "dword"), or four words (a "quad word" or "quad").

Computer memory caches usually operate on *blocks* of memory that consist of several consecutive words. These units are customarily called *cache blocks*, or, in CPU caches, *cache lines*.

Virtual memory systems partition the computer's main storage into even larger units, traditionally called *pages*.

Systematic multiples

Terms for large quantities of bits can be formed using the standard range of SI prefixes for powers of 10, e.g., kilo $= 10^3 = 1000$ (as in kilobit or kbit), mega- $= 10^6 = 1\,000\,000$ (as in megabit or Mbit) and giga $= 10^9 = 1\,000\,000\,000$ (as in gigabit or Gbit). These prefixes are more often used for multiples of bytes, as in kilobyte (1 kB = 8000 bit), megabyte (1 MB = 8 000 000 bit), and gigabyte (1 GB = 8 000 000 000 bit).

However, for technical reasons, the capacities of computer memories and some storage units are often multiples of some large power of two, such as $2^{28} = 268\,435\,456$ bytes. To avoid such unwieldy numbers, people have often repurposed the SI prefixes to mean the nearest power of two, e.g., using the prefix *kilo* for $2^{10} = 1024$, *mega* for $2^{20} = 1\,048\,576$, and *giga* for $2^{30} = 1\,073\,741\,824$, and so on. For example, a random access memory chip with a capacity of 2^{28} bytes would be referred to as a 256-megabyte chip. The table below illustrates these differences.

Multiples of bits

Decimal				Binary					
Value		SI		Value		IEC		JEDEC	
1000	10^3	kbit	kilobit	1024	2^{10}	Kibit	kibibit	Kbit	kilobit
1000^2	10^6	Mbit	megabit	1024^2	2^{20}	Mibit	mebibit	Mbit	megabit
1000^3	10^9	Gbit	gigabit	1024^3	2^{30}	Gibit	gibibit	Gbit	gigabit
1000^4	10^{12}	Tbit	terabit	1024^4	2^{40}	Tibit	tebibit	-	
1000^5	10^{15}	Pbit	petabit	1024^5	2^{50}	Pibit	pebibit	-	
1000^6	10^{18}	Ebit	exabit	1024^6	2^{60}	Eibit	exbibit	-	
1000^7	10^{21}	Zbit	zettabit	1024^7	2^{70}	Zibit	zebibit	-	
1000^8	10^{24}	Ybit	yottabit	1024^8	2^{80}	Yibit	yobibit	-	

See also: Nibble · Byte · Orders of magnitude of data

Symbol	Prefix	SI Meaning	Binary meaning	Size difference
k	kilo	$10^3 = 1000^1$	$2^{10} = 1024^1$	2.40%
M	mega	$10^6 = 1000^2$	$2^{20} = 1024^2$	4.86%
G	giga	$10^9 = 1000^3$	$2^{30} = 1024^3$	7.37%
T	tera	$10^{12} = 1000^4$	$2^{40} = 1024^4$	9.95%
P	peta	$10^{15} = 1000^5$	$2^{50} = 1024^5$	12.59%
E	exa	$10^{18} = 1000^6$	$2^{60} = 1024^6$	15.29%
Z	zetta	$10^{21} = 1000^7$	$2^{70} = 1024^7$	18.06%
Y	yotta	$10^{24} = 1000^8$	$2^{80} = 1024^8$	20.89%

In the past, uppercase *K* has been used instead of lowercase *k* to indicate 1024 instead of 1000. However, this usage was never consistently applied.

On the other hand, for external storage systems (such as optical discs), the SI prefixes were commonly used with their decimal values (powers of 10). There have been many attempts to resolve the confusion by providing alternative notations for power-of-two multiples. In 1998 the International Electrotechnical Commission (IEC) issued a standard for this purpose, namely a series of binary prefixes that use 1024 instead of 1000 as the main radix:[8]

Symbol	Prefix			
Ki	kibi, *binary kilo*	1 kibibyte (KiB)	2^{10} bytes	1024 B
Mi	mebi, *binary mega*	1 mebibyte (MiB)	2^{20} bytes	1024 KiB
Gi	gibi, *binary giga*	1 gibibyte (GiB)	2^{30} bytes	1024 MiB
Ti	tebi, *binary tera*	1 tebibyte (TiB)	2^{40} bytes	1024 GiB
Pi	pebi, *binary peta*	1 pebibyte (PiB)	2^{50} bytes	1024 TiB
Ei	exbi, *binary exa*	1 exbibyte (EiB)	2^{60} bytes	1024 PiB

The JEDEC memory standards however define uppercase K, M, and G for the binary powers 2^{10}, 2^{20} and 2^{30} to reflect common usage.[9]

Size examples

- 1 bit – answer to a yes/no question.
- 1 byte – a number from 0 to 255.
- 90 bytes: enough to store a typical line of text from a book.
- 512 bytes = ½ KiB: the typical sector of a hard disk.
- 1024 bytes = 1 KiB: the classical block size in UNIX filesystems.
- 2048 bytes = 2 KiB: a CD-ROM sector.
- 4096 bytes = 4 KiB: a memory page in x86 (since Intel 80386).
- 4 kB: about one page of text from a novel.

- 120 kB: the text of a typical pocket book.
- 1 MiB – a 1024×1024 pixel bitmap image with 256 colors (8 bpp color depth).
- 3 MB – a three-minute song (133 kbit/s).
- 650–900 MB – a CD-ROM.
- 1 GB – 114 minutes of uncompressed CD-quality audio at 1.4 Mbit/s.
- 8/16 GB – two common sizes of USB flash drives.
- 4 TB – the size of a $100 hard disk (as of early 2018).
- 12 TB Largest hard disk drive (as of early 2018)
- 16 TB Largest commercially available solid state drive (as of early 2018)
- 60 TB Largest solid state drive constructed (as of early 2018)
- 1.3 ZB – prediction of the volume of the whole internet in 2016.

Obsolete and unusual units

Several other units of information storage have been named:[7]

- 1 bit: unibit,[10][11] sniff.
- 2 bits: dibit,[12][13][10][14] crumb,[15] quad, quarter, taste, tayste, tidbit, tydbit, lick, lyck, semi-nibble.
- 3 bits: tribit,[12][13][10] triad, triade,[16][17] tribble.
- 5 bits: pentad, pentade,[18] nickel, nyckle.
- 6 bits: byte (in early IBM machines using BCD alphamerics), hexad, hexade,[18][19] sextet.[20]
- 7 bits: heptad, heptade.[18]
- 8 bits: octet, now usually called byte
- 10 bits: declet,[21][22][23][24] decle,[25] deckle, dyme.
- 12 bits: slab.[26][27][28]
- 15 bits: parcel (on CDC 6600 and CDC 7600).
- 16 bits: doublet,[29] wyde,[3][30] parcel (on Cray-1), plate, playte, chomp, chawmp (on a 32-bit machine).
- 18 bits: chomp, chawmp (on a 36-bit machine).
- 32 bits: quadlet,[29][31][32] tetra,[30] dinner, dynner, gawble (on a 32-bit machine).
- 48 bits: gobble, gawble (under circumstances that remain obscure).
- 64 bits: octlet,[29] octa.[30]
- 96 bits: bentobox (in ITRON OS)
- 128 bits: hexlet.[29][33]
- 16 bytes: paragraph (on Intel x86 processors).
- 6 trits: tryte.[34]
- combit, comword.[35][36][37]

Some of these names are jargon, obsolete, or used only in very restricted contexts.

See also

- Metric prefix
- File size

References

1. Norman Abramson (1963), *Information theory and coding*. McGraw-Hill.

2. Mackenzie, Charles E. (1980). *Coded Character Sets, History and Development* (https://books.google.com/books?id=6-tQAAAAMAAJ). *The Systems Programming Series* (1 ed.). Addison-Wesley Publishing Company, Inc. p. xii. ISBN 0-201-14460-3. LCCN 77-90165 (https://lccn.loc.gov/77-90165). Retrieved 2016-05-22. [1] (https://web.archive.org/web/20160526172151/https://textfiles.meulie.net/bitsaved/Books/Mackenzie_CodedCharSets.pdf)
3. Knuth, Donald Ervin. *The Art of Computer Programming: Seminumerical algorithms*. **2**. Addison Wesley.
4. Shanmugam (2006), *Digital and Analog Computer Systems*.
5. Gregg Jaeger (2007), [2] (https://www.springer.com/physics/quantum+physics/book/978-0-387-35725-6) *Quantum information: an overview*
6. I. Ravi Kumar (2001), *Comprehensive Statistical Theory of Communication*.
7. *Nybble* (http://dictionary.reference.com/browse/nybble) at dictionary reference.com; sourced from Jargon File 4.2.0, accessed 2007-08-12
8. ISO/IEC standard is ISO/IEC 80000-13:2008. This standard cancels and replaces subclauses 3.8 and 3.9 of IEC 60027-2:2005. The only significant change is the addition of explicit definitions for some quantities. ISO Online Catalogue (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=31898)
9. JEDEC Solid State Technology Association (December 2002). "Terms, Definitions, and Letter Symbols for Microcomputers, Microprocessors, and Memory Integrated Circuits" (http://www.jedec.org/download/search/JESD100B01.pdf) (PDF). *JESD 100B.01*. Retrieved 2009-04-05
10. Horak, Ray (2007). *Webster's New World Telecom Dictionary* (https://books.google.com/books?id=L18YaEomzjMC&pg=PA402). John Wiley & Sons. p. 402. ISBN 9-78047022571-4.
11. http://www.yourdictionary.com/unibit#computer
12. Steinbuch, Karl W.; Wagner, Siegfried W., eds. (1967) [1962]. Written at Karlsruhe, Germany. *Taschenbuch der Nachrichtenverarbeitung* (in German) (2 ed.). Berlin / Heidelberg / New York: Springer-Verlag OHG. pp. 835–836. LCCN 67-21079 (https://lccn.loc.gov/67-21079). Title No. 1036.
13. Steinbuch, Karl W.; Weber, Wolfgang; Heinemann, Traute, eds. (1974) [1967]. Written at Karlsruhe / Bochum. *Taschenbuch der Informatik - Band III - Anwendungen und spezielle Systeme der Nachrichtenverarbeitung*. *Taschenbuch der Nachrichtenverarbeitung* (in German). **3** (3 ed.). Berlin / Heidelberg / New York: Springer Verlag. pp. 357–358. ISBN 3-540-06242-4. LCCN 73-80607 (https://lccn.loc.gov/73-80607).
14. Bertram, H. Neal (1994). *Theory of magnetic recording* (1 ed.). Cambridge University Press. ISBN 0-521-44973-1. 9-780521-449731. "[…] The writing of an impulse would involve writing a dibit or two transitions arbitrarily closely together. […]"
15. Weisstein, Eric. W. "Crumb" (http://mathworld.wolfram.com/Crumb.html). MathWorld. Retrieved 2015-08-02.
16. Paul, Reinhold (2013). "Elektrotechnik und Elektronik für Informatiker - Grundgebiete der Elektronik" (https://books.google.com/books?id=-fzKBgAAQBAJ&hl=de). *Leitfaden der Informatik*. B.G. Teubner Stuttgart / Springer. ISBN 3-32296652-6. 9-78332296652-0. Retrieved 2015-08-03.
17. Böhme, Gert; Born, Werner; Wagner, B.; Schwarze, G. (2013-07-02) [1969]. Reichenbach, Jürgen, ed. *Programmierung von Prozeßrechnern* (https://books.google.com/books?id=T-S0BgAAQBAJ). Reihe Automatisierungstechnik (in German). **79**. VEB Verlag Technik Berlin, reprint: Springer Verlag. doi:10.1007/978-3-663-02721-8 (https://doi.org/10.1007%2F978-3-663-02721-8). ISBN 978-3-663-00808-8. 9/3/4185.
18. Speiser, Ambrosius Paul (1965) [1961]. *Digitale Rechenanlagen - Grundlagen / Schaltungstechnik / Arbeitsweise / Betriebssicherheit* [*Digital computers - Basics / Circuits / Operation / Reliability*] (in German) (2 ed.). ETH Zürich, Zürich, Switzerland: Springer-Verlag / IBM. pp. 6, 34, 165, 183, 208, 213, 215. LCCN 65-14624 (https://lccn.loc.gov/65-14624). 0978.

19. Steinbuch, Karl W., ed. (1962). Written at Karlsruhe, Germany. *Taschenbuch der Nachrichtenverarbeitung* (in German) (1 ed.). Berlin / Göttingen / New York: Springer-Verlag OHG. p. 1076. LCCN 62-14511 (https://lccn.loc.gov/62-14511).
20. Svoboda, Antonín; White, Donnamaie E. (2016) [1979-08-01]. *Advanced Logical Circuit Design Techniques* (http://www.donnamaie.com/Advanced_logic_ckt/Advanced_Logical_Circuit_Design_Techniques%20sm.pdf) (PDF) (retyped electronic reissue ed.). Garland STPM Press (original issue) / WhitePubs (reissue). ISBN 0-8240-7014-3. Archived (https://web.archive.org/web/20170414163013/http://www.donnamaie.com/Advanced_logic_ckt/Advanced_Logical_Circuit_Design_Techniques%20sm.pdf) (PDF) from the original on 2017-04-14. Retrieved 2017-04-15. [3] (http://www.donnamaie.com/) [4] (https://books.google.com/books?id=g3uzAAAAIAAJ)
21. *IEEE 754-2008 - IEEE Standard for Floating-Point Arithmetic* (http://ieeexplore.ieee.org/servlet/opac?punumber=4610933). 2008-08-29. doi:10.1109/IEEESTD.2008.4610935 (https://doi.org/10.1109%2FIEEESTD.2008.4610935). ISBN 978-0-7381-5752-8. Retrieved 2016-02-10.
22. Muller, Jean-Michel; Brisebarre, Nicolas; de Dinechin, Florent; Jeannerod, Claude-Pierre; Lefèvre, Vincent; Melquiond, Guillaume; Revol, Nathalie; Stehlé, Damien; Torres, Serge (2010). *Handbook of Floating-Point Arithmetic* (1 ed.). Birkhäuser. doi:10.1007/978-0-8176-4705-6 (https://doi.org/10.1007%2F978-0-8176-4705-6). ISBN 978-0-8176-4704-9. LCCN 2009939668 (https://lccn.loc.gov/2009939668).
23. Erle, Mark A. (2008-11-21). *Algorithms and Hardware Designs for Decimal Multiplication* (https://books.google.com/books?id=_4OKPgcPfEYC) (Thesis). Lehigh University: ProQuest (published 2009). ISBN 9781109042283. 1109042280. Retrieved 2016-02-10.
24. Kneusel, Ronald T. (2015). *Numbers and Computers* (https://books.google.com/books?id=FPlICAAAQBAJ). Springer. ISBN 9783319172606. 3319172603. Retrieved 2016-02-10.
25. Zbiciak, Joe. "AS1600 Quick-and-Dirty Documentation" (http://spatula-city.org/~im14u2c/intv/jzintv-1.0-beta3/doc/utilities/as1600.txt). Retrieved 2013-04-28.
26. "315 Electronic Data Processing System" (http://www.thecorememory.com/NCR_315_EDPS.pdf) (PDF). NCR. November 1965. NCR MPN ST-5008-15. Archived (https://web.archive.org/web/20160524113545/http://www.thecorememory.com/NCR_315_EDPS.pdf) (PDF) from the original on 2016-05-24. Retrieved 2015-01-28.
27. Bardin, Hillel (1963). "NCR 315 Seminar" (http://www.thecorememory.com/NCR_315_Seminar.pdf) (PDF). *Computer Usage Communique*. **2** (3). Archived (https://web.archive.org/web/20160524113611/http://www.thecorememory.com/NCR_315_Seminar.pdf) (PDF) from the original on 2016-05-24.
28. Schneider, Carl (2013) [1970]. *Datenverarbeitungs-Lexikon* (https://books.google.com/books?id=3aN9BwAAQBAJ) [*Lexicon of information technology*] (in German) (softcover reprint of hardcover 1st ed.). Wiesbaden, Germany: Springer Fachmedien Wiesbaden GmbH / Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH. pp. 201, 308. doi:10.1007/978-3-663-13618-7 (https://doi.org/10.1007%2F978-3-663-13618-7). ISBN 978-3-409-31831-0. Retrieved 2016-05-24. "*slab*, Abk. aus *syllable* = Silbe, die kleinste adressierbare Informationseinheit für 12 bit zur Übertragung von zwei Alphazeichen oder drei numerischen Zeichen. *(NCR)* […] Hardware: Datenstruktur: NCR 315-100 / NCR 315-RMC; Wortlänge: Silbe; Bits: 12; Bytes: –; Dezimalziffern: 3; Zeichen: 2; Gleitkommadarstellung: fest verdrahtet; Mantisse: 4 Silben; Exponent: 1 Silbe (11 Stellen + 1 Vorzeichen) [*slab*, abbr. for *syllable* = syllable, smallest addressable information unit for 12 bits for the transfer of two alphabetical characters or three numerical characters. *(NCR)* […] Hardware: Data structure: NCR 315-100 / NCR 315-RMC; Word length: Syllable; Bits: 12; Bytes: –; Decimal digits: 3; Characters: 2; Floating point format: hard-wired; Significand: 4 syllables; Exponent: 1 syllable (11 digits + 1 prefix)]"

29. *IEEE Std 1754-1994 - IEEE Standard for a 32-bit Microcontroller Architecture* (http://ieeexplore.ieee.org/servlet/opac?punumber=3435). The Institute of Electrical and Electronic Engineers, Inc. pp. 5–7. doi:10.1109/IEEESTD.1995.79519 (https://doi.org/10.1109%2FIEEESTD.1995.79519). ISBN 1-55937-428-4. Retrieved 2016-02-10. (NB. The standard defines doublets, quadlets, octlets and hexlets as 2, 4, 8 and 16 bytes, giving the numbers of bits (16, 32, 64 and 128) only as a secondary meaning. This might be important given that bytes were not always understood to mean 8 bits (octets) historically.)
30. Knuth, Donald Ervin (2004-02-15) [1999]. *Fascicle 1: MMIX* (http://profs.scienze.univr.it/~manca/storia-informatica/mmix.pdf) (PDF). *The Art of Computer Programming* (0th printing, 15th ed.). Stanford University: Addison-Wesley. Archived (https://web.archive.org/web/20170330001534/http://profs.scienze.univr.it/~manca/storia-informatica/mmix.pdf) (PDF) from the original on 2017-03-30. Retrieved 2017-03-30.
31. Böszörményi, László; Hölzl, Günther; Pirker, Emaneul (February 1999). Written at Salzburg, Austria. Zinterhof, Peter; Vajteršic, Marian; Uhl, Andreas, eds. *Parallel Cluster Computing with IEEE1394–1995*. Parallel Computation: 4th International ACPC Conference including Special Tracks on Parallel Numerics (ParNum '99) and Parallel Computing in Image Processing, Video Processing, and Multimedia. *Proceedings: Lecture Notes in Computer Science 1557*. Berlin, Germany: Springer Verlag.
32. Nicoud, Jean-Daniel (1986). *Calculatrices* (https://books.google.com/books?id=3hGZ2qxxby8C&pg=PR3). *Traité d'électricité de l'École polytechnique fédérale de Lausanne* (in French). **14** (2 ed.). Lausanne: Presses polytechniques romandes. ISBN 2880740541.
33. *Proceedings*. Symposium on Experiences with Distributed and Multiprocessor Systems (SEDMS). **4**. USENIX Association. 1993.
34. Brousentsov, N. P.; Maslov, S. P.; Ramil Alvarez, J.; Zhogolev, E.A. "Development of ternary computers at Moscow State University" (http://www.computer-museum.ru/english/setun.htm). Retrieved 2010-01-20.
35. US4319227 (https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US4319227), Malinowski, Christopher W.; Heinz Rinderle & Martin Siegle, "Three-state signaling system", issued 1982-03-09, assigned to Department of Research and Development, AEG-Telefunken, Heilbronn, Germany
36. http://www.google.com/patents/US4319227
37. http://patentimages.storage.googleapis.com/pdfs/US4319227.pdf

External links

- Representation of numerical values and SI units in character strings for information interchanges (http://people.csail.mit.edu/jaffer/MIXF/MIXF-08)
- Bit Calculator (http://www.bit-calculator.com) – Make conversions between bits, bytes, kilobits, kilobytes, megabits, megabytes, gigabits, gigabytes, terabits, terabytes, petabits, petabytes, exabits, exabytes, zettabits, zettabytes, yottabits, yottabytes.
- Paper on standardized units for use in information technology (http://www.cl.cam.ac.uk/~mgk25/information-units.txt)
- Data Byte Converter (https://www.gbmb.org/)