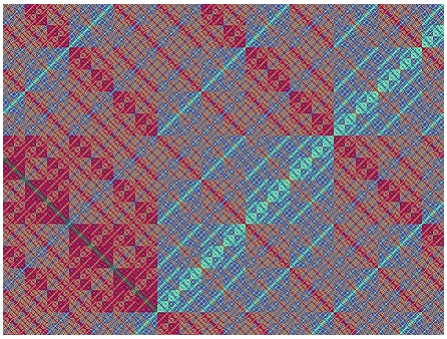
WIKIPEDIA

# Coding theory

**Coding theory** is the study of the properties of codes and their respective fitness for specific applications. Codes are used for data compression, cryptography, error detection and correction, data transmission and data storage. Codes are studied by various scientific disciplines—such as information theory, electrical engineering, mathematics, linguistics, and computer science—for the purpose of designing efficient and reliable data transmission methods. This typically involves the removal of redundancy and the correction or detection of errors in the transmitted data.



A two-dimensional visualisation of the Hamming distance, a critical measure in coding theory.

There are four types of coding:[1]

1. Data compression (or, *source coding*)
2. Error control (or *channel coding*)
3. Cryptographic coding
4. Line coding

Data compression attempts to remove redundancy from the data from a source in order to transmit it more efficiently. For example, Zip data compression makes data files smaller, for purposes such as to reduce Internet traffic. Data compression and error correction may be studied in combination.

Error correction adds extra data bits to make the transmission of data more robust to disturbances present on the transmission channel. The ordinary user may not be aware of many applications using error correction. A typical music CD uses the Reed-Solomon code to correct for scratches and dust. In this application the transmission channel is the CD itself. Cell phones also use coding techniques to correct for the fading and noise of high frequency radio transmission. Data modems, telephone transmissions, and the NASA Deep Space Network all employ channel coding techniques to get the bits through, for example the turbo code and LDPC codes.

## Contents

**History of coding theory**

**Source coding**
- Definition
- Properties
- Principle
- Example

**Channel coding**
- Linear codes
  - Linear block codes
  - Convolutional codes

**Cryptographical coding**

**Line coding**

**Other applications of coding theory**
- Group testing

Analog coding
**Neural coding**
**See also**
**Notes**
**References**

# History of coding theory

In 1948, Claude Shannon published "A Mathematical Theory of Communication", an article in two parts in the July and October issues of the *Bell System Technical Journal*. This work focuses on the problem of how best to encode the information a sender wants to transmit. In this fundamental work he used tools in probability theory, developed by Norbert Wiener, which were in their nascent stages of being applied to communication theory at that time. Shannon developed information entropy as a measure for the uncertainty in a message while essentially inventing the field of information theory.

The binary Golay code was developed in 1949. It is an error-correcting code capable of correcting up to three errors in each 24-bit word, and detecting a fourth.

Richard Hamming won the Turing Award in 1968 for his work at Bell Labs in numerical methods, automatic coding systems, and error-detecting and error-correcting codes. He invented the concepts known as Hamming codes, Hamming windows, Hamming numbers, and Hamming distance.

# Source coding

The aim of source coding is to take the source data and make it smaller.

## Definition

Data can be seen as a random variable $X : \Omega \to \mathcal{X}$, where $x \in \mathcal{X}$ appears with probability $\mathbb{P}[X = x]$.

Data are encoded by strings (words) over an alphabet $\Sigma$.

A code is a function

$C : \mathcal{X} \to \Sigma^*$ (or $\Sigma^+$ if the empty string is not part of the alphabet).

$C(x)$ is the code word associated with $x$.

Length of the code word is written as

$l(C(x))$.

Expected length of a code is

$$l(C) = \sum_{x \in \mathcal{X}} l(C(x))\mathbb{P}[X = x]$$

The concatenation of code words $C(x_1, \ldots, x_k) = C(x_1)C(x_2)\ldots C(x_k)$.

The code word of the empty string is the empty string itself:

$C(\epsilon) = \epsilon$

### Properties

1. $C : \mathcal{X} \to \Sigma^*$ is non-singular if injective.
2. $C : \mathcal{X}^* \to \Sigma^*$ is uniquely decodable if injective.
3. $C : \mathcal{X} \to \Sigma^*$ is instantaneous if $C(x_1)$ is not a prefix of $C(x_2)$ (and vice versa).

### Principle

Entropy of a source is the measure of information. Basically, source codes try to reduce the redundancy present in the source, and represent the source with fewer bits that carry more information.

Data compression which explicitly tries to minimize the average length of messages according to a particular assumed probability model is called entropy encoding.

Various techniques used by source coding schemes try to achieve the limit of Entropy of the source. $C(x) \geq H(x)$, where $H(x)$ is entropy of source (bitrate), and $C(x)$ is the bitrate after compression. In particular, no source coding scheme can be better than the entropy of the source.

### Example

Facsimile transmission uses a simple run length code. Source coding removes all data superfluous to the need of the transmitter, decreasing the bandwidth required for transmission.

## Channel coding

The purpose of channel coding theory is to find codes which transmit quickly, contain many valid code words and can correct or at least detect many errors. While not mutually exclusive, performance in these areas is a trade off. So, different codes are optimal for different applications. The needed properties of this code mainly depend on the probability of errors happening during transmission. In a typical CD, the impairment is mainly dust or scratches.

CDs use cross-interleaved Reed–Solomon coding to spread the data out over the disk.[2]

Although not a very good code, a simple repeat code can serve as an understandable example. Suppose we take a block of data bits (representing sound) and send it three times. At the receiver we will examine the three repetitions bit by bit and take a majority vote. The twist on this is that we don't merely send the bits in order. We interleave them. The block of data bits is first divided into 4 smaller blocks. Then we cycle through the block and send one bit from the first, then the second, etc. This is done three times to spread the data out over the surface of the disk. In the context of the simple repeat code, this may not appear effective. However, there are more powerful codes known which are very effective at correcting the "burst" error of a scratch or a dust spot when this interleaving technique is used.

Other codes are more appropriate for different applications. Deep space communications are limited by the thermal noise of the receiver which is more of a continuous nature than a bursty nature. Likewise, narrowband modems are limited by the noise, present in the telephone network and also modeled better as a continuous disturbance. Cell phones are subject to rapid fading. The high frequencies used can cause rapid fading of the signal even if the receiver is moved a few inches. Again there are a class of channel codes that are designed to combat fading.

## Linear codes

The term **algebraic coding theory** denotes the sub-field of coding theory where the properties of codes are expressed in algebraic terms and then further researched.

Algebraic coding theory is basically divided into two major types of codes:

1. Linear block codes
2. Convolutional codes

It analyzes the following three properties of a code – mainly:

- code word length
- total number of valid code words
- the minimum distance between two valid code words, using mainly the Hamming distance, sometimes also other distances like the Lee distance

### Linear block codes

Linear block codes have the property of linearity, i.e. the sum of any two codewords is also a code word, and they are applied to the source bits in blocks, hence the name linear block codes. There are block codes that are not linear, but it is difficult to prove that a code is a good one without this property.[3]

Linear block codes are summarized by their symbol alphabets (e.g., binary or ternary) and parameters $(n,m,d_{min})$[4] where

1. n is the length of the codeword, in symbols,
2. m is the number of source symbols that will be used for encoding at once,
3. $d_{min}$ is the minimum hamming distance for the code.

There are many types of linear block codes, such as

1. Cyclic codes (e.g., Hamming codes)
2. Repetition codes
3. Parity codes
4. Polynomial codes (e.g., BCH codes)
5. Reed–Solomon codes
6. Algebraic geometric codes
7. Reed–Muller codes
8. Perfect codes

Block codes are tied to the sphere packing problem, which has received some attention over the years. In two dimensions, it is easy to visualize. Take a bunch of pennies flat on the table and push them together. The result is a hexagon pattern like a bee's nest. But block codes rely on more dimensions which cannot easily be visualized. The powerful (24,12) Golay code used in deep space communications uses 24 dimensions. If used as a binary code (which it usually is) the dimensions refer to the length of the codeword as defined above.

The theory of coding uses the *N*-dimensional sphere model. For example, how many pennies can be packed into a circle on a tabletop, or in 3 dimensions, how many marbles can be packed into a globe. Other considerations enter the choice of a code. For example, hexagon packing into the constraint of a rectangular box will leave empty space at the corners. As the dimensions get larger, the percentage of empty space grows smaller. But at certain dimensions, the packing uses all the space and these codes are the so-called "perfect" codes. The only nontrivial and useful perfect codes are the distance-3 Hamming codes with parameters satisfying $(2^r - 1, 2^r - 1 - r, 3)$, and the [23,12,7] binary and [11,6,5] ternary Golay codes.[3][4]

Another code property is the number of neighbors that a single codeword may have.[5] Again, consider pennies as an example. First we pack the pennies in a rectangular grid. Each penny will have 4 near neighbors (and 4 at the corners which are farther away). In a hexagon, each penny will have 6 near neighbors. When we increase the dimensions, the number of near neighbors increases very rapidly. The result is the number of ways for noise to make the receiver choose a neighbor (hence an error) grows as well. This is a fundamental limitation of block codes, and indeed all codes. It may be harder to cause an error to a single neighbor, but the number of neighbors can be large enough so the total error probability actually suffers.[5]

Properties of linear block codes are used in many applications. For example, the syndrome-coset uniqueness property of linear block codes is used in trellis shaping,[6] one of the best known shaping codes. This same property is used in sensor networks for distributed source coding.

### Convolutional codes

The idea behind a convolutional code is to make every codeword symbol be the weighted sum of the various input message symbols. This is like convolution used in LTI systems to find the output of a system, when you know the input and impulse response.

So we generally find the output of the system convolutional encoder, which is the convolution of the input bit, against the states of the convolution encoder, registers.

Fundamentally, convolutional codes do not offer more protection against noise than an equivalent block code. In many cases, they generally offer greater simplicity of implementation over a block code of equal power. The encoder is usually a simple circuit which has state memory and some feedback logic, normally XOR gates. The decoder can be implemented in software or firmware.

The Viterbi algorithm is the optimum algorithm used to decode convolutional codes. There are simplifications to reduce the computational load. They rely on searching only the most likely paths. Although not optimum, they have generally been found to give good results in low noise environments.

Convolutional codes are used in voiceband modems (V.32, V.17, V.34) and in GSM mobile phones, as well as satellite and military communication devices.

# Cryptographical coding

Cryptography or cryptographic coding is the practice and study of techniques for secure communication in the presence of third parties (called adversaries).[7] More generally, it is about constructing and analyzing protocols that block adversaries;[8] various aspects in information security such as data confidentiality, data integrity, authentication, and non-repudiation[9] are central to modern cryptography. Modern cryptography exists at the intersection of the disciplines of mathematics, computer science, and electrical engineering. Applications of cryptography include ATM cards, computer passwords, and electronic commerce.

Cryptography prior to the modern age was effectively synonymous with *encryption*, the conversion of information from a readable state to apparent nonsense. The originator of an encrypted message shared the decoding technique needed to recover the original information only with intended recipients, thereby precluding unwanted persons from doing the same. Since World War I and the advent of the computer, the methods used to carry out cryptology have become increasingly complex and its application more widespread.

Modern cryptography is heavily based on mathematical theory and computer science practice; cryptographic algorithms are designed around computational hardness assumptions, making such algorithms hard to break in practice by any adversary. It is theoretically possible to break such a system, but it is infeasible to do so by any known practical means. These schemes are therefore termed computationally secure; theoretical advances, e.g., improvements in integer factorization algorithms, and faster computing technology require these solutions to be continually adapted. There exist information-theoretically secure schemes that provably cannot be broken even with unlimited computing power—an example is the one-time pad—but these schemes are more difficult to implement than the best theoretically breakable but computationally secure mechanisms.

# Line coding

A line code (also called digital baseband modulation or digital baseband transmission method) is a code chosen for use within a communications system for baseband transmission purposes. Line coding is often used for digital data transport.

Line coding consists of representing the digital signal to be transported by an amplitude- and time-discrete signal that is optimally tuned for the specific properties of the physical channel (and of the receiving equipment). The waveform pattern of voltage or current used to represent the 1s and 0s of a digital data on a transmission link is called *line encoding*. The common types of line encoding are unipolar, polar, bipolar, and Manchester encoding.

# Other applications of coding theory

Another concern of coding theory is designing codes that help synchronization. A code may be designed so that a phase shift can be easily detected and corrected and that multiple signals can be sent on the same channel.

Another application of codes, used in some mobile phone systems, is code-division multiple access (CDMA). Each phone is assigned a code sequence that is approximately uncorrelated with the codes of other phones. When transmitting, the code word is used to modulate the data bits representing the voice message. At the receiver, a demodulation process is performed to recover the data. The properties of this class of codes allow many users (with different codes) to use the same radio channel at the same time. To the receiver, the signals of other users will appear to the demodulator only as a low-level noise.

Another general class of codes are the automatic repeat-request (ARQ) codes. In these codes the sender adds redundancy to each message for error checking, usually by adding check bits. If the check bits are not consistent with the rest of the message when it arrives, the receiver will ask the sender to retransmit the message. All but the simplest wide area network protocols use ARQ. Common protocols include SDLC (IBM), TCP (Internet), X.25 (International) and many others. There is an extensive field of research on this topic because of the problem of matching a rejected packet against a new packet. Is it a new one or is it a retransmission? Typically numbering schemes are used, as in TCP."RFC793" (http://tools.ietf.org/html/rfc793). *RFCs*. Internet Engineering Task Force (IETF). September 1981.

## Group testing

Group testing uses codes in a different way. Consider a large group of items in which a very few are different in a particular way (e.g., defective products or infected test subjects). The idea of group testing is to determine which items are "different" by using as few tests as possible. The origin of the problem has its roots in the Second World War when the United States Army Air Forces needed to test its soldiers for syphilis. It originated from a ground-breaking paper by Robert Dorfman.

## Analog coding

Information is encoded analogously in the neural networks of brains, in analog signal processing, and analog electronics. Aspects of analog coding include analog error correction,[10] analog data compression[11] and analog encryption.[12]

# Neural coding

Neural coding is a neuroscience-related field concerned with how sensory and other information is represented in the brain by networks of neurons. The main goal of studying neural coding is to characterize the relationship between the stimulus and the individual or ensemble neuronal responses and the relationship among electrical activity of the neurons in the ensemble.[13] It is thought that neurons can encode both digital and analog information,[14] and that neurons follow the principles of information theory and compress information,[15] and detect and correct[16] errors in the signals that are sent throughout the brain and wider nervous system.

# See also

- Coding gain
- Covering code
- Error correction code
- Folded Reed–Solomon code
- Group testing
- Hamming distance, Hamming weight
- Lee distance
- List of algebraic coding theory topics
- Spatial coding and MIMO in multiple antenna research
  - Spatial diversity coding is spatial coding that transmits replicas of the information signal along different spatial paths, so as to increase the reliability of the data transmission.
  - Spatial interference cancellation coding
  - Spatial multiplex coding
- Timeline of information theory, data compression, and error correcting codes

# Notes


1. James Irvine; David Harle (2002). "2.4.4 Types of Coding". *Data Communications and Networks* (https://books.google.com/books?id=ZigejECe4r0C). p. 18. ISBN 9780471808725. "There are four types of coding"
2. Todd Campbell. "Answer Geek: Error Correction Rule CDs" (http://abcnews.go.com/Technology/story?id=119305&page=1).
3. Terras, Audrey (1999). *Fourier Analysis on Finite Groups and Applications* (https://books.google.com/books?id=-B2TA669dJMC&pg=PA195). Cambridge University Press. ISBN 978-0-521-45718-7.
4. Blahut, Richard E. (2003). *Algebraic Codes for Data Transmission* (https://books.google.com/books?id=n0XHMY58tL8C&pg=PA60). Cambridge University Press. ISBN 978-0-521-55374-2.
5. Christian Schlegel; Lance Pérez (2004). *Trellis and turbo coding* (https://books.google.com/books?id=9wRCjfGAaEcC&pg=PA73). Wiley-IEEE. p. 73. ISBN 978-0-471-22755-7.
6. Forney, G.D., Jr. (March 1992). "Trellis shaping" (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=119687). *IEEE Transactions on Information Theory*. **38** (2 Pt 2): 281–300. doi:10.1109/18.119687o (https://doi.org/10.1109%2F18.119687o) (inactive 2019-02-20).
7. Rivest, Ronald L. (1990). "Cryptology". In J. Van Leeuwen (ed.). *Handbook of Theoretical Computer Science*. **1**. Elsevier.
8. Bellare, Mihir; Rogaway, Phillip (21 September 2005). "Introduction". *Introduction to Modern Cryptography*. p. 10.

9. Menezes, A. J.; van Oorschot, P. C.; Vanstone, S. A. (1997). *Handbook of Applied Cryptography* (http://www.cacr.math.uwaterloo.ca/hac/). ISBN 978-0-8493-8523-0. Archived (https://web.archive.org/web/20050307081354/http://www.cacr.math.uwaterloo.ca/hac/) from the original on 2005-03-07.
10. Chen, Brian; Wornell, Gregory W. (July 1998). "Analog Error-Correcting Codes Based on Chaotic Dynamical Systems" (http://allegro.mit.edu/dspg/publications/Journals/pdf/98Chen.pdf) (PDF). *IEEE Transactions on Communications*. **46** (7): 881–890. CiteSeerX 10.1.1.30.4093 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.30.4093). doi:10.1109/26.701312 (https://doi.org/10.1109%2F26.701312).
11. Hvala, Franc Novak Bojan; Klavžar, Sandi (1999). "On Analog Signature Analysis". *Proceedings of the conference on Design, automation and test in Europe*. CiteSeerX 10.1.1.142.5853 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.142.5853). ISBN 1-58113-121-6.
12. Shujun Li; Chengqing Li; Kwok-Tung Lo; Guanrong Chen (April 2008). "Cryptanalyzing an Encryption Scheme Based on Blind Source Separation" (http://epubs.surrey.ac.uk/532452/1/IEEETCASI2008.pdf) (PDF). *IEEE Transactions on Circuits and Systems I*. **55** (4): 1055–63. doi:10.1109/TCSI.2008.916540 (https://doi.org/10.1109%2FTCSI.2008.916540).
13. Brown EN, Kass RE, Mitra PP (May 2004). "Multiple neural spike train data analysis: state-of-the-art and future challenges". *Nat. Neurosci.* **7** (5): 456–61. doi:10.1038/nn1228 (https://doi.org/10.1038%2Fnn1228). PMID 15114358 (https://www.ncbi.nlm.nih.gov/pubmed/15114358).
14. Thorpe, S.J. (1990). "Spike arrival times: A highly efficient coding scheme for neural networks" (http://pop.cerco.ups-tlse.fr/fr_vers/documents/thorpe_sj_90_91.pdf) (PDF). In Eckmiller, R.; Hartmann, G.; Hauske, G. (eds.). *Parallel processing in neural systems and computers* (https://books.google.com/books?id=b9gmAAAAMAAJ) (PDF). North-Holland. pp. 91–94. ISBN 978-0-444-88390-2. Retrieved 30 June 2013.
15. Gedeon, T.; Parker, A.E.; Dimitrov, A.G. (Spring 2002). "Information Distortion and Neural Coding" (http://www.math.ualberta.ca/ami/CAMQ/table_of_content/vol_10/10_1c.htm). *Canadian Applied Mathematics Quarterly*. **10** (1): 10. CiteSeerX 10.1.1.5.6365 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.6365).
16. Stiber, M. (July 2005). "Spike timing precision and neural error correction: local behavior". *Neural Computation*. **17** (7): 1577–1601. arXiv:q-bio/0501021 (https://arxiv.org/abs/q-bio/0501021). doi:10.1162/0899766053723069 (https://doi.org/10.1162%2F0899766053723069). PMID 15901408 (https://www.ncbi.nlm.nih.gov/pubmed/15901408).


# References


- Elwyn R. Berlekamp (2014), *Algebraic Coding Theory*, World Scientific Publishing (revised edition), ISBN 978-9-81463-589-9.
- MacKay, David J. C.. *Information Theory, Inference, and Learning Algorithms (https://web.archive.org/web/20160217105359/http://www.inference.phy.cam.ac.uk/mackay/itila/book.html)* Cambridge: Cambridge University Press, 2003. ISBN 0-521-64298-1
- Vera Pless (1982), *Introduction to the Theory of Error-Correcting Codes*, John Wiley & Sons, Inc., ISBN 0-471-08684-3.
- Randy Yates, *A Coding Theory Tutorial (https://web.archive.org/web/20110710143034/http://www.digitalsignallabs.com/tutorial.pdf)*.