REPORTS @SCM
AN ELECTRONIC JOURNAL OF THE SOCIETAT CATALANA DE MATEMÀTIQUES

Long Density Parity Check codes

Tomàs Ortega
University of California, Irvine
tomas.ortega@uci.edu

Resum *(CAT)*
Els codis de Low Density Parity Check (LDPC), o codis de comprovació de paritat de baixa densitat, van ser proposats per Robert Gallager al 1963 a la seva tesi doctoral [3]. Tot i que la tesi demostra que existeixen LDPC asimptòticament òptims, es van abandonar pel seu cost computacional. Gràcies a la teoria de grafs expansors, sabem que poden ser codificats i descodificats en temps lineal. Aquest TFM [4] repassa la història dels LDPC i en presenta una nova família creada a partir d'estructures d'incidència anomenades Quadrangles Generalitzats.

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *error correcting codes, expander graphs, communication theory.*

Abstract

To send information from a transmitter to a receiver through a noisy channel, the data is sent with some redundancy in order to correct the errors that might occur. We will assume in this work that the data to be sent is a sequence of bits. The transmitter sends blocks of n bits, k of which are data bits, and the remaining $n-k$ are redundancy bits. The way we map vectors of k data bits to vectors of n transmitted bits is called a *(binary) code*. The vectors of n transmitted bits of a given code are called its *codewords*. The fraction of data bits per codeword, k/n, is called the code's *rate*.

A receiver, upon receiving a vector of n bits (also known as a word), has to decide which vector of k data bits encoded this word. This is called decoding a word. To make the receiver's job easier, codes are designed in such a way that different codewords do not resemble each other. Thus, if some errors occur in the channel, the receiver can still distinguish which codeword was sent and perform the decoding successfully. Given two vectors, the *Hamming distance* between them is the number of coordinates where the vectors differ. Thus, if the *minimum Hamming distance* between codewords is large enough, when we decode the received word we can correct some errors that were introduced in the noisy channel.

Given a family of codes, we say that they are *asymptotically good* if their rate is bounded by a constant larger than zero, and their minimum Hamming distance grows linearly with block length (n). In 1948, Shannon used the probabilistic method to show that asymptotically good codes exist [5]. However, his method did not give explicit examples on how to obtain them. Moreover, these codes might not be easily encodable or decodable.

Error correcting codes need to be practical, which means that encoding and decoding must be cheap both in computation and storage. The most common solution is to use *linear codes*, which are characterized

by the property that any linear combination of codewords is also a codeword. After Shannon's landmark paper, the race to find asymptotically good linear codes began. During the sixties and seventies, algebraic constructions proved that such codes exist, but they were not encodable and decodable in linear time.

In 1963, Gallager discovered Low Density Parity Check (LDPC) codes [3], which he found experimentally to have good performance and had a link to graph theory through random graphs. However, Gallager lacked the tools to give explicit arguments of all the good properties of these codes, namely the concept of *expander graphs*. These codes were somewhat forgotten, since the thought at the time was that they were not practical due to the computing power they required.

In the seventies, the concept of expander graphs was introduced, which allowed Tanner, Sipser and Spielman [8, 6] to produce stronger results than the ones Gallager had obtained with random graphs. These led to Spielman's discovery in 1996 of the first family of asymptotically good, explicit codes, with encoding and decoding time linear in block length [7], which he coined *expander codes*.

Nowadays, LDPC codes are extensively used. Most notably, they appear in Digital Video Broadcasting, Wi-Fi and 5G standards [2, 1]. They are also widely employed for various storage system applications. While Spielman's decoding algorithm gives stronger analytical results, usually variants of Gallager's probabilistic decoding method are used, as the latter has better performance in practice.

The work [4] is an introduction to LDPC codes, both in theory and practice. It also presents a new family of LDPC codes constructed from the point-line incidence structure of Generalized Quadrangles. These codes are quasi-cyclic, which means that their parity-check matrix can be represented as a series of cyclic permutations of a smaller base matrix. This structure aids in efficient hardware implementation of decoding algorithms. Numerical experiments show that GQ-LDPC codes out-perform random codes.

Acknowledgements

The author would like to thank Professor Simeon Ball for his guidance.

References

[1] J.H. Bae, A. Abotabl, H.-P. Lin, K.-B. Song, J. Lee, An overview of channel coding for 5G NR cellular communications, *APSIPA Transactions on Signal and Information Processing* **8** (2019), 1–14.

[2] D.J. Costello, G. David Forney, Channel coding: The road to channel capacity, *Proceedings of the IEEE* **95(6)** (2007), 1150–1177.

[3] R. G. Gallager, Low density parity check codes, monograph (1963).

[4] T. Ortega, Low density parity check codes, Master's thesis, Universitat Politècnica de Catalunya, 2021.

[5] C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal* **27(3)** (1948), 379–423.

[6] M. Sipser, D.A. Spielman, Expander codes, *IEEE Transactions on Information Theory* **42(6)** (1996), 1710–1722.

[7] D.A. Spielman, Linear-time encodable and decodable error-correcting codes, *IEEE Transactions on Information Theory* **42(6)** (1996), 1723–1731.

[8] R. Tanner, A recursive approach to low complexity codes, *IEEE Transactions on Information Theory* **27(5)** (1981), 533–547.