Chapter

Introductory Chapter: Coding Theory

Dinesh G. Harkut and Kashmira N. Kasat

1. Introduction

The field of channel coding is a fundamental part of digital communication systems. Its purpose is to enable reliable transmission of data over noisy and error-prone communication channels. Channel coding theory deals with the design of error-correcting codes that can tolerate a certain number of errors introduced during transmission, while still allowing for accurate reconstruction of the original data at the receiver end.

Channel coding is the process of adding redundancy to a message or data stream to protect against errors that may occur during transmission over a noisy communication channel. The redundant information, known as the error-correcting code, enables the receiver to detect and correct errors that occur during transmission. For example, let us say you want to send a message “HELLO” to your friend over a communication channel. During transmission, the message may get corrupted due to noise or interference on the channel. To protect against errors, you can add redundancy to the message by encoding it with an error-correcting code, such as a cyclic redundancy check (CRC) code. The CRC code adds a checksum to the message, which is computed based on the message content using a mathematical algorithm. The receiver also computes the checksum of the received message and compares it with the checksum sent by the transmitter. If the two checksums do not match, it indicates that there was an error in the transmission, and the receiver requests the transmitter to resend the message. In this example, the CRC code serves as a channel coding scheme that protects the message against errors during transmission over the noisy communication channel. By adding redundancy to the message, the receiver can detect and correct errors, ensuring reliable communication between the transmitter and receiver.

2. Historical background

The history of channel coding dates back to the early days of telegraphy and radio communication, when engineers first realized the need for error-correcting codes to ensure reliable transmission over noisy and error-prone communication channels. The first error-correcting codes were developed in the 1940s, with the work of Richard Hamming and Claude Shannon laying the foundation for modern channel coding theory [1]. Richard Hamming developed the concept of Hamming codes in 1950, which were the first practical error-correcting codes used in digital communication systems. These codes were designed to detect and correct single-bit errors, and they were widely used in early computer systems and communication protocols [1]. Claude

Shannon, a pioneer in information theory, introduced the concept of channel capacity in his landmark paper "A Mathematical Theory of Communication" published in 1948 [1]. Shannon's work established the fundamental limits of communication over noisy channels, and it paved the way for the development of more sophisticated error-correcting codes that could approach these limits. Since then, channel coding theory has continued to evolve, with new codes and techniques being developed to address the challenges of modern communication systems. Today, channel coding is an essential part of digital communication systems, enabling reliable transmission of data over a wide range of communication channels.

3. Basic concepts

Channel coding theory is based on several fundamental concepts that are essential to understanding how error-correcting codes work. These concepts include code rate, block length, minimum distance, error correction capability, and channel capacity.

- Code rate: The code rate is the ratio of the number of message bits to the total number of transmitted bits, including the redundant bits added by the error-correcting code. A higher code rate means that more information is transmitted per bit, but it also means that the error-correction capability of the code is reduced [2].

- Block length: The block length is the number of bits in a block of data that is encoded using an error-correcting code. Longer block lengths typically provide better error-correction capabilities, but they also increase the delay and complexity of the encoding and decoding processes [2].

- Minimum distance: The minimum distance of an error-correcting code is the smallest number of bit changes that must occur to transform one valid codeword into another. A higher minimum distance means that the code is more robust against errors, as it can detect and correct more errors during transmission [2].

- Error correction capability: The error correction capability of an error-correcting code is the maximum number of errors that can be corrected during transmission. This capability depends on the code rate, block length, and minimum distance of the code, as well as the characteristics of the communication channel [2].

- Channel capacity: The channel capacity is the maximum rate at which information can be transmitted over a noisy communication channel with a given error rate. This limit is determined by the channel characteristics and the laws of physics, and it provides a theoretical upper bound on the performance of any error-correcting code used over the channel [2].

These basic concepts form the foundation of channel coding theory, and they are used to design and analyze error-correcting codes for a wide range of communication systems. By understanding these concepts, engineers can develop more efficient and effective error-correcting codes that can provide reliable communication over noisy and error-prone channels. Several types of error-correcting codes are commonly used in digital communication systems. Each type has its own advantages and disadvantages, depending on the specific application requirements and constraints. The main types of error-correcting codes are:

4. Block codes

Block codes divide a message into fixed-size blocks, and each block is encoded separately using an error-correcting code. The most common block codes are Reed-Solomon codes, which are used in a wide range of applications, including CD and DVD storage, satellite communication, and digital television [3].

Advantages:

- High error-correction capability
- Simple encoding and decoding algorithms
- Robust against burst errors

Disadvantages:

- High redundancy, leading to lower code rate
- Inefficient for variable-length messages

5. Convolutional codes

Convolutional codes are based on a mathematical concept called convolution, which involves multiplying and adding a sequence of numbers. Convolutional codes are designed to operate on a continuous stream of data, and they use a sliding window to encode and decode the data [2].

Advantages:

- High error-correction capability
- Efficient for variable-length messages
- Can be used for high-speed communication systems

Disadvantages:

- Complex encoding and decoding algorithms
- Sensitive to phase distortion and timing errors
- Limited ability to correct burst errors

6. Turbo codes

Turbo codes are a type of iterative code that use multiple convolutional codes in parallel, with a feedback mechanism to refine the decoding process. Turbo codes are widely used in mobile communication systems, such as 3G and 4G cellular networks [4].

Advantages:

- Very high error-correction capability
- Efficient for variable-length messages
- Robust against noise and interference

Disadvantages:

- High complexity, requiring specialized hardware or software
- Sensitive to phase distortion and timing errors
- Limited availability of standards and implementation tools

7. LDPC codes

Low-density parity-check (LDPC) codes are a type of linear block code that use a sparse parity-check matrix to encode and decode data. LDPC codes are widely used in high-speed communication systems, such as wired and wireless networks, as well as storage systems [5].

Advantages:

- High error-correction capability
- Efficient for variable-length messages
- Robust against noise and interference

Disadvantages:

- Complex encoding and decoding algorithms
- Sensitive to channel characteristics and noise models
- Limited availability of standards and implementation tools

Overall, the choice of error-correcting code depends on the specific application requirements and constraints, such as the required error-correction capability, message length, transmission rate, and implementation complexity. Each type of error-correcting code has its own trade-offs between error-correction capability, efficiency, complexity, and robustness, and the most appropriate code must be selected based on a careful analysis of the application requirements and constraints.

8. Encoding and decoding

Encoding and decoding are fundamental operations in channel coding theory, which are used to convert an input message into a coded message and to recover the original message from the received coded message, respectively. The encoding and decoding processes are based on mathematical algorithms and principles, which are designed to provide a certain level of error-correction capability and robustness to the transmission of data over noisy and unreliable communication channels [3].

Encoding: Encoding involves transforming an input message into a coded message, which contains additional redundancy information that can be used to detect and correct errors that occur during transmission. The encoding process typically involves applying a mathematical function or algorithm to the input message, which generates a set of parity bits that are added to the message to form the coded message. For example, consider a simple block code called a parity code, which involves adding a single parity bit to a message of length n bits. The parity bit is computed as the XOR (exclusive OR) of all the bits in the message, and it is added to the end of the message to form the coded message. The encoding process can be represented by the following equation:

$$C = M \| P \tag{1}$$

where M is the original message, P is the parity bit, || denotes concatenation, and C is the coded message. For example, if the input message is 1011, the parity bit is computed as 1 XOR 0 XOR 1 XOR 1 = 1, and the coded message is 10111.

Decoding: Decoding involves recovering the original message from the received coded message, which may have been corrupted by errors during transmission. The decoding process typically involves applying a mathematical function or algorithm to the received coded message, which uses the redundant information in the message to detect and correct errors and recover the original message. For example, consider the parity code described above. To decode a received message, the receiver computes the parity bit of the received message and compares it to the received parity bit. If they are the same, the message is assumed to be error-free and the original message is recovered by removing the parity bit. If they are different, an error is detected and the receiver may attempt to correct the error by flipping the received bit that is inconsistent with the computed parity bit. The decoding process can be represented by the following equation:

$$M' = C[1:n]. \tag{2}$$

where C is the received coded message, M′ is the recovered message, and [1:n] denotes the first n bits of C. Of course, this is just a simple example of encoding and decoding with a parity code, and there are many more sophisticated and powerful coding schemes that are used in practice. However, the basic principles of encoding and decoding remain the same, and they are critical for ensuring reliable and efficient communication over noisy and unreliable channels.

There are various implementation strategies available for channel coding, and the choice of implementation strategy depends on factors such as the complexity of the coding scheme, the required data rate, and the hardware and software resources available for implementation. In this section, we will discuss some common implementation strategies for channel coding, along with their advantages and disadvantages and examples of their applications in real-time environments.

Software-based implementation: Software-based implementation of channel coding involves implementing the encoding and decoding algorithms using software running on a general-purpose processor such as a CPU or a DSP. This implementation strategy is flexible and can be easily updated or modified as needed, but may be slower and less power-efficient compared to hardware-based implementation.

> *Example: The Reed-Solomon code is a widely used block code that can correct multiple errors in a block of data. Reed-Solomon coding is often implemented in software on general-purpose processors for applications such as digital audio and video storage and transmission, and satellite communication.*

FPGA-based implementation: FPGA (Field-Programmable Gate Array) based implementation of channel coding involves implementing the encoding and decoding algorithms on an FPGA, which is a programmable hardware device that can be reconfigured to perform different functions. This implementation strategy provides high performance and low latency, but may require specialized expertise and tools for design and implementation.

> *Example: The Turbo code is a powerful and widely used convolutional code that can achieve very high data rates and error-correction capability. Turbo code decoding is often implemented on FPGAs for applications such as wireless communication, digital broadcasting, and satellite communication.*

ASIC-based implementation: ASIC (Application-Specific Integrated Circuit) based implementation of channel coding involves designing and fabricating custom hardware circuits that implement the encoding and decoding algorithms. This implementation strategy provides high performance and low power consumption, but may require high initial costs and long design and fabrication times.

> *Example: The LDPC (Low-Density Parity-Check) code is a powerful and efficient linear code that can achieve very high data rates and error-correction capability. LDPC code decoding is often implemented on ASICs for applications such as wireless communication, digital broadcasting, and storage systems.*

Hybrid implementation: Hybrid implementation of channel coding involves combining different implementation strategies such as software, FPGA, and ASIC to achieve the desired balance of performance, flexibility, and cost. This implementation strategy can provide high performance and flexibility while reducing the costs and development time compared to fully custom hardware implementation.

> *Example: The convolutional code is a popular and widely used linear code that can achieve high data rates and error-correction capability. Convolutional code decoding is often implemented using a hybrid implementation strategy that combines software and FPGA or ASIC for applications such as wireless communication and digital broadcasting.*

The choice of implementation strategy for channel coding depends on the specific requirements and constraints of the application. Software-based implementation provides flexibility and ease of development, FPGA-based implementation provides high performance and low latency, ASIC-based implementation provides high performance and low power consumption, and hybrid implementation provides a balance of performance and cost.

Channel coding theory is a fundamental aspect of modern communication systems that enables reliable transmission of digital data over noisy communication channels. The goal of channel coding is to add redundancy to the transmitted data in such a way that the receiver can correct errors caused by noise or interference in the channel. This is achieved by using error-correcting codes that add redundancy to the data stream, which can be used by the receiver to detect and correct errors. Channel coding theory involves the study of the mathematical principles underlying error-correcting codes, their encoding and decoding algorithms, and their performance analysis. Channel coding is a multidisciplinary field that draws upon concepts from mathematics, computer science, information theory, and communication engineering.

The most common types of error-correcting codes used in channel coding are block codes and convolutional codes. Block codes divide the input data into fixed-size blocks and add parity bits to each block, while convolutional codes operate on a continuous stream of input data and add redundant symbols based on a sliding window of previous symbols. One important concept in channel coding theory is the Hamming distance, which is a measure of the number of bit positions in which two binary strings differ. The minimum Hamming distance of an error-correcting code is the smallest Hamming distance between any two valid codewords, and it determines the error-correction capability of the code.

Another important concept in channel coding theory is the decoding algorithm, which is used by the receiver to recover the original data from the received codeword. There are two main types of decoding algorithms: maximum likelihood decoding and syndrome decoding. Maximum likelihood decoding involves searching for the most likely codeword given the received data, while syndrome decoding involves using the syndrome of the received codeword to correct errors. Channel coding theory also includes the analysis of the performance of error-correcting codes under different channel conditions, such as the signal-to-noise ratio (SNR) and the bit error rate (BER). The performance of a code is typically measured by its error-correction capability, which is the maximum number of errors that the code can correct, and its coding efficiency, which is the ratio of the number of information bits to the total number of transmitted bits.

The development of channel coding theory has led to the discovery of many powerful error-correcting codes that have found widespread use in communication systems. Some examples of widely used codes include the Reed-Solomon code, which is used in digital audio and video storage and transmission, the Turbo code, which is used in wireless communication and digital broadcasting, and the LDPC code, which is used in wireless communication and storage systems.

In conclusion, channel coding theory is a crucial aspect of modern communication systems that enables reliable transmission of digital data over noisy channels. It involves the study of error-correcting codes, their encoding and decoding algorithms, and their performance analysis. The development of channel coding theory has led to the discovery of many powerful error-correcting codes that have found widespread use in communication systems.

Author details

Dinesh G. Harkut[1]* and Kashmira N. Kasat[2]

1 Department of Computer Science and Engineering, Prof. Ram Meghe College of Engineering and Management, Badnera-Amravati, M.S., India

2 Department of Electronics and Telecommunication Engineering, Prof. Ram Meghe College of Engineering and Management, Badnera-Amravati, M.S., India

*Address all correspondence to: dg.harkut@gmail.com

References

[1] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;**27**(3):379-423

[2] Proakis JG, Salehi M. Digital Communications. 5th ed. New York, NY, USA: McGraw-Hill; 2008

[3] Lin C, Costello DJ Jr. Error Control Coding: Fundamentals and Applications. Upper Saddle River, NJ, USA: Pearson/ Prentice Hall; 2004

[4] Berrou C, Glavieux A, Thitimajshima P. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In: Proceedings of the International Conference on Communications (ICC'93); Geneva, Switzerland. 1993. pp. 1064-1070

[5] Gallager RG. Low-density parity-check codes. IRE Transactions on Information Theory. 1962;**8**(1):21-28