Evaluating the EVENODD Code: Principles, Applications, and Future Prospects in Data Storage Systems

Jinxiang Deng[*]

Automation, Nanjing University of Information Science & Technology, 210044, Nanjing, China

Abstract. In modern data storage and transmission, ensuring data integrity and reliability is critical due to potential losses or corruption caused by channel instability and system errors. Check codes have been developed to address these issues, allowing recovery of the original data even when errors occur. This paper provides a comprehensive analysis of the EVENODD code, a widely used parity code in error detection and correction applications. The fundamental principle of the EVENODD code relies on adding a binary check bit to ensure that the count of ones in the data string is either even or odd, depending on the desired configuration. Its implementation in Redundant Array of Independent Disks level 6 architecture highlights the code's ability to improve data reliability by incorporating dual parity, enhancing fault tolerance in distributed systems. The advantages and limitations of EVENODD, such as its efficiency in single-bit error detection but inability to correct multi-bit errors, are examined. Additionally, comparisons are made with similar codes, including Longitudinal Redundancy Check and Cyclic Redundancy Check, to showcase their respective strengths and use cases. The paper discusses the EVENODD code's industrial applications, particularly in satellite remote sensing and library databases, where data integrity is paramount. Future directions include optimizing the code's performance and cost-effectiveness for large-scale data storage and transmission environments, promoting secure and reliable information systems.

1 Introduction

In today's rapidly advancing digital world, data integrity and reliability are paramount, especially within large-scale data storage and transmission systems. Channel instability, hardware malfunctions, and environmental interferences can lead to data loss or corruption, posing significant challenges in ensuring consistent data quality [1]. To address these issues, error-detecting and error-correcting codes have become essential in computer science and communication engineering, providing a means of safeguarding data against transmission errors.

---

*Corresponding author: 202213360090@nuist.edu.cn

Numerous error-checking methods, including parity checks, Cyclic Redundancy Check (CRC), and Reed-Solomon (RS) codes, have been developed to enhance data security in various applications [2]. EVENODD codes, as one of the most effective parity-based methods, play a vital role in error detection and correction, especially in RAID 6 systems. Originally used as a straightforward error detection tool, EVENODD codes have been incorporated into fault-tolerant storage architectures due to their dual-parity feature, which enables efficient data recovery in the event of one or two simultaneous disk failures [3].

This paper conducts a comprehensive analysis of EVENODD codes, including their basic principles, encoding and decoding mechanisms, and application in RAID 6 architecture. Additionally, the study examines the advantages and limitations of EVENODD codes compared to similar methods, such as LRC and CRC, and highlights their industrial applications in sectors requiring high data reliability, like satellite remote sensing and library databases. Finally, the paper explores potential future directions for optimizing EVENODD codes to meet evolving demands in data storage and transmission systems, focusing on cost-effectiveness and reliability in large-scale data environments.

2 The principle and application of EVENODD code

2.1 Basic Principle

The EVENODD code consists primarily of a check digit at the end of the data. It is determined by the number of "1" in the data string. For example, if there is a string "10010101" where the number of "1" is 4, which is even, then you can add a check digit "0" to the end [4]. This is the basic composition of the even-effect check, which is the opposite of the odd check, that is, when the number of "1" is odd, "0" is added at the end.

EVENODD code has a number of advantages over other validation methods. The first is that the algorithm is simple, which can occupy a small amount of memory and validate the data [5]. Secondly, EVENODD code has a wide range of uses in many fields and is highly practical. But it also has its drawbacks. The main thing is that it can only detect single-bit errors, and it can't fix them. Errors are also prone to occur in high-bit rate environments.

2.2 Application in RAID 6

RAID was proposed by Professor D.A. Patterson of the University of California, Berkeley, in 1988, its full name in English is Redundant Arrays of Independent Disks. RAID technology Combining multiple independent hard disks into a single disk group with a huge capacity greatly improves read and write speeds while providing data protection capabilities.

The EVENODD code alone doesn't play a big role, but in RAID 6, the two-digit check bit can be used to greatly improve its efficiency [6]. In RAID 6, XOR arithmetic is predominantly used. The principle of operation is that the same is 0 and the difference is 1.

For RAID 6, there are m (m needs to be prime) data and 2 parity disks (parity1, parity2), every disk is a column of m-I symbols. This construction is particularly attractive because it only needs some simple algorithms.

As shown in Table 1 and Fig. 1, following the introduction to the EVEN-ODD code, the encoding and decoding processes are presented. Considering the case where m = 5, this means there are 5 data blocks (D_0, D_1, D_2, D_3, D_4) and two parity blocks (P_1, P_2). Each data disk contains 4 blocks..

Table 1. Schematic diagram.

block\disk	D_0	D_1	D_2	D_3	D_4	P_1	P_2
0							
1							
2							
3							

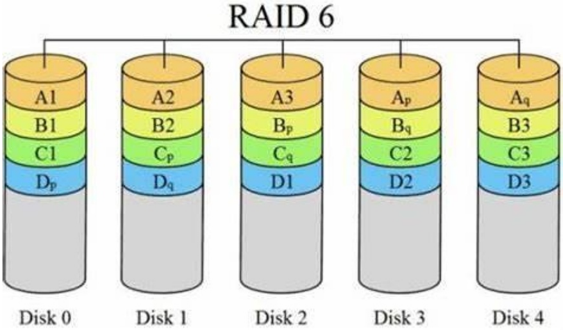


Fig. 1. Model diagram (Photo credit: Original).

The encoding process of EVEN-ODD code mainly achieves redundant storage of data by generating additional parity blocks so that data can be recovered in the event of a failure [7]. For this case, Parity 1 is a simple XOR row-wise parity, the value of p1 in each row is equal to the sum of the values in all data disks in that row. For Parity 2, it is the EVENODD invention. A special sum S is needed first: $\sum_{t=0}^{m-1} a_{m-t,t}$. In here m = 5, so it is: $\sum_{t=0}^{4} a_{5-t,t}$. Then, Parity2 by $a_{l,m+1}(\text{parity2}) = S + \sum_{t=0,t!=l+1}^{4} a_{<l-t>m,t}$. This is how p1 and p2 work during encoding.

And in the EVEN-ODD code scheme, the decoding process aims to utilize pre-computed parity information to efficiently recover data lost due to one or two simultaneous failures. When a single data block is lost, the recovery process is relatively simple, using parity1 to solve it is enough [8]. However, if two data blocks are lost, the reconstruction process is slightly more complicated. A system of equations are needed to solve it.

3 Check code similar to the EVENODD code

There are some other codes that are similar to EVENODD code like LRC (Longitudinal Redundancy Check, CRC(Longitudinal Redundancy Check) and so on. Both they and the EVENODD code use XOR operations. The LRC code is almost the same as the EVENODD code, except that the object of the check digit is different. There is a big difference between CRC and EVENODD code [9, 10]. The verification capability of CRC is greatly improved, which can not only detect multi-bit errors, but also have high error detection reliability.

3.1 LRC code

When a large amount of data is processed, the EVENODD code alone cannot be effectively processed, so the data can be divided into chunks, and each piece can be processed separately, which increases the complexity of the algorithm, but can effectively improve the accuracy of data validation. The specific calculation steps can be summarized as follows:

Data grouping: Group the data to be transmitted by the same number of bits per row.

XOR operation by column: XOR operation is performed column by column on the grouped data blocks.

Generate checksum: Stitch together the final result of each column of XOR operations to obtain a checksum.

Append Checksum: Append the checksum to the end of the block for transmission or storage.

Receiver checksum: After receiving the data block, the receiver performs XOR operations according to each column to obtain a new checksum.

Verification Result: Compares the received checksum with the calculated checksum. If the two check sums are the same, no errors occurred during the data transfer process; If the two check sums are different, an error is considered to have occurred during the data transfer process.

The above process is simple and easy to implement. As a result, the cost of such an algorithm is lower. But at the same time, because it still uses XOR operations, it can only detect but bit errors and cannot be modified. Because there are more check sums retained, if there is more data in transit, the check sum may be lost, resulting in a failed validation.

3.2 CRC code

CRC is an abbreviation for Cyclic Redundancy Check, and its operation is essentially the division of polynomials. The CRC algorithm is based on the polynomial arithmetic of GF (2) (2-element Galois field). Through a generation polynomial, it is divided by the original data, and if it can be divided, the data transfer is successful, and vice versa. The specific operation steps are as follows:

As show in the table fig. 2. Polynomial generation. Different polynomials have different error correction capabilities, If you want to use the r-bit checksum, the number of times you generate the polynomial should be r, and the following are some standard generated polynomials.

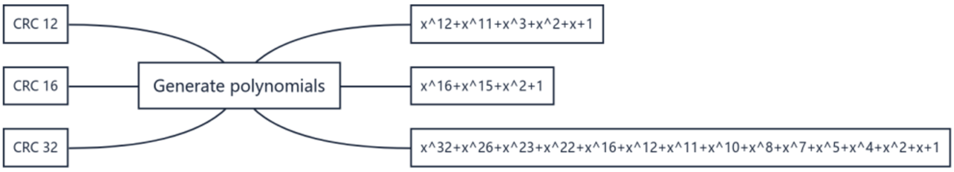


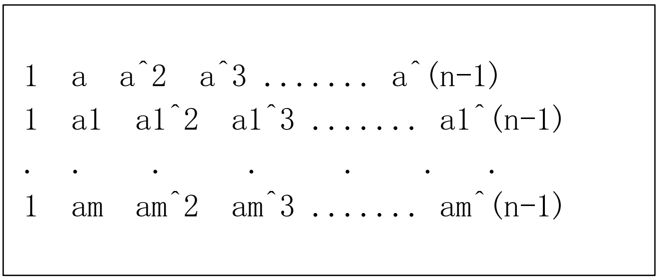
Fig. 2. Standard generated polynomials (Photo credit: Original).

Coding process. Treat the string of data bits to be transmitted as a binary polynomial, denoted as M(x). M(x) shifts r to the left to give xrM (x). Divide xrM (x) to generate the polynomial G (x), and the remainder is the CRC code.

As show in the table 1. Decoding process. After receiving the data and CRC code, the receiver composes them into a new binary polynomial, N (x). Divide N (x) to generate the polynomial G (x), if the remainder is 0, it means that the transmission is error-free; Otherwise, an error occurred during the transfer.

3.3 Reed-Solomon code

This is a kind of error correction code, which is different from the check code, which can restore lost data and has a very powerful function. The Reed-Solomon code is computed on the Galois field in order to prevent the inversion of the matrix from producing a decimal point and resulting in a loss of accuracy. Its encoding process involves a Vandermonde matrix in the form of:



```
1  a   a^2  a^3 ....... a^(n-1)
1  a1   a1^2  a1^3 ....... a1^(n-1)
.  .     .     .     .     .   .
1  am   am^2  am^3 ....... am^(n-1)
```

Fig. 3. Vandermonde matrix (Photo credit: Original).

As show in the fig. 3. By merging the identity matrix and the Vandermonde matrix vertically, the coefficient matrix is obtained, and the data matrix is multiplied by the coefficient matrix to obtain the encoded result. The decoding process is the process of data repair. The specific calculation process is as follows: delete the missing data rows, delete the corresponding rows of the coefficient matrix, and find its inverse, and then multiply it with the lost data matrix to obtain the original data.

It has the advantage of being powerful with less computation, so it is used in many fields, including the RAID mentioned above.

4 The application of evenodd code in industrial computers

Based on the theory mentioned above, there are many requirements for the application of validity codes in the actual industry. Application in satellite remote sensing. Because the industry requires high reliability and security of data, a single check code cannot be used for data validation. In practical applications, RAID is more commonly used in the industry. Among the many RAID models, RAID 5 is used to transmit and store data. Compared with other RAID models, RAID 5 retains the advantages of data security and reliability, but also has a relatively low cost. Therefore, it is often used in satellite remote sensing to store and transmit data. At present, the Think System DE4000H storage disk array is used as the storage device for satellite remote sensing data. Databases are also used in libraries, and in order to prevent data loss, libraries also use check codes to ensure the integrity of data. Some libraries also use error correction codes, such as the Reed-Solomon code. In this industry, data integrity is required, so some speed is sacrificed for accuracy.

The specific implementation steps for the second industry are described in detail. The library's database often records people's borrowing of books and statistics on literature, books, etc. Therefore, the library database needs a lot of capacity, 1TB or even more. A large amount of data flow will inevitably lead to data loss, corruption, etc. The data can be chunked and individual check digits can be added, in this case with CRC codes. When the data transmission is successful, desired data can be got quickly, when the data transmission error, it means that the CRC inspection is not passed, then the system will resend the data until it is successfully delivered.

With the rapid development of big data and cloud storage, more and more industries need to use data storage. In order to ensure its security, the data is often not sent separately and is often redundant to detect errors. The method of data error detection is not only the way to check the validity code, but also to improve the device, for example, to copy the data, use two copies of the data, and when one fails, you can choose to receive the data in the other. This approach will have higher requirements for hardware, but without the validity code as a guarantee, this approach will not be efficient. In the face of different fields, a variety of measures should be adopted to ensure the security of data.

5 Conclusion

This study has provided a thorough analysis of the EVENODD code, highlighting its principles, applications, and advantages within the context of modern data storage and transmission systems. The research emphasizes the code's significant role in enhancing fault tolerance, particularly in RAID 6 architecture, where its dual-parity design allows for reliable data recovery even in the event of dual-disk failures. By comparing EVENODD with similar parity and error-checking codes such as LRC and CRC, this study underscores the unique strengths and limitations of EVENODD in diverse industrial settings, including satellite remote sensing and large library databases, where data integrity and reliability are critical. Future research could focus on optimizing the performance and cost-effectiveness of EVENODD for large-scale storage systems. Exploring adaptive mechanisms that adjust the encoding and decoding processes based on system requirements could enhance the code's efficiency. Additionally, integrating EVENODD with emerging technologies in cloud storage and distributed computing might offer new approaches to maintaining data integrity across increasingly complex and large-scale infrastructures. By addressing these directions, future work could further establish EVENODD as a key component in robust, scalable data storage solutions.

References

1. H. Xu, X. Zhu, Z. Zhao, X. Wei, X. Wang, J. Zuo, Research of Pipeline Leak Detection Technology and Application Prospect of Petrochemical Wharf, in 2020 IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), Vol. 9, pp. 263-271 (IEEE, 2020)
2. H. Tian, X. E, J. Yang, Adaptive fault tolerance schemes for cloud storage of library. Comput. Eng. Appl. (2018)
3. X. Zhu, Y. Zhang, Z. Zhao, J. Zuo, Radio frequency sensing based environmental monitoring technology. Proc. Fourth Int. Workshop Pattern Recognit., SPIE, 11198, 187–191 (2019)
4. J. Xin, C. Zhu, F. Wen, C. Guo, Z. Zhan, Y. Hou, F. Cao, Y. Bai, Y. Yan, Research on data recovery and status reconstruction of RAID for satellite remote sensing data storage. Radio Commun. Technol. (2024)
5. Y. Zhang, H. Xu, X. Zhu, Detection and quantization technique of optical distributed acoustic coupling based on φ-OTDR. J. Shanghai Jiaotong Univ. (Sci.) **25**, 208–213 (2020)
6. Y. Zeng, Implementation and efficacy of EVENODD coding in high-reliability data storage architectures. Sci. Technol. Eng. Chem. Environ. Prot. **1**, 9 (2024)
7. Y. Zhang, X. Wang, J. Wen, X. Zhu, WiFi-based non-contact human presence detection technology Scientific Reports, **14**(1), 3605 (2024)
8. H. Zhan, Comparison and performance analysis of EVENODD and RDP fault-tolerant coding methods. Highlights Sci. Eng. Technol. **87**, 37–44 (2024)
9. X. Wei, X. Zhu, X. Wang, Z. Zhao, J. Zuo, Fuzzy Fault Tree Analysis Method and Its Application in Fault Diagnosis of Denitration System in Thermal Power Plant in Proceedings of the 2020 8th International Conference on Information Technology: IoT and Smart City, pp. 227-232 (2020)
10. Y. Zhang, H. Zhao, X. Zhu, Z. Zhao, J. Zuo, Strain measurement quantization technology based on DAS system. Proc. 2019 IEEE 3rd Adv. Inf. Manage. Commun. Electron. Autom. Control Conf. (IMCEC), 214–218 (2019)