Enhanced EVENODD Encoding Techniques: Innovations for Improved Data Resilience

Sheng Bi*

Faculty of Engineering, The University of New South Wales, 1466/2025, Sydney, Australia

Abstract. The scale and complexity of modern storage systems necessitate robust data protection mechanisms, where EVENODD coding algorithms play a critical role. Despite its widespread application, traditional EVENODD algorithms exhibit limitations that demand advanced solutions for today's data storage challenges. This paper introduces an innovative approach to the encoding and decoding processes of EVENODD code, highlighting significant enhancements that offer better protection and operational efficiency. The study explores new methods that reduce computational complexity and accelerate data recovery processes, ensuring superior data integrity and system reliability. The enhancements are vital for large-scale storage systems where high data availability and rapid recovery are paramount. The analysis also delves into the practical implementations and potential challenges of these novel methods, proposing strategic solutions to integrate them effectively in existing systems. This research aims to set a foundational framework for future studies and practical applications in data storage and protection technology, pushing the boundaries of current data resilience capabilities.

1 Introduction

The demand for reliable data storage continues to grow with the rapid expansion of digital information, presenting challenges that require robust error correction and data integrity mechanisms. EVENODD codes, developed as an effective solution for error detection in data transmissions, play a crucial role in maintaining the integrity of data across various transmission and storage environments. As these systems increase in complexity, the underlying mechanisms that ensure data protection must evolve to meet the higher standards of reliability and efficiency expected in modern applications [1].

Despite the widespread adoption of EVENODD codes, the traditional methods of encoding and decoding are now being pushed to their limits by the demands of larger, more dynamic storage architectures. These methods must be optimized to handle increased volumes of data without compromising on speed or computational overhead. There exists an urgent need to refine these processes to accommodate the burgeoning data loads and to reduce the latency in recovery during failures [2].

* Corresponding author: z5524857@ad.unsw.edu.au

This paper addresses the gaps in current EVENODD encoding and decoding techniques by introducing innovative modifications designed to streamline these processes. Enhancements focus on reducing computational complexity, thus expediting data recovery and bolstering system resilience against disk failures. By integrating new approaches such as the use of skew check columns and optimizing XOR operations, the modified EVENODD codes aim to set a new standard for data protection in high-reliability storage systems. These developments not only enhance the performance but also extend the applicability of EVENODD codes across a broader spectrum of storage technologies [3].

2 Relevant theories

2.1 Definition of EVENODD code

EVENODD code is one of the famous parity checking schemes. It is an erasure code used in data storage systems that can tolerate the failure of two disks. It is especially widely used in RAID-6 systems [4, 5]. It is designed to provide data redundancy to protect the system from up to two disk failures while maintaining data integrity and recoverability. Its working principle is based on the following points. First, EVENODD code use parity checking to generate redundant information in data blocks, which is called “Parity1” and “Parity2”. By constructing these two parities, it is convenient to reconstruct lost data in data disks. Second, the data is organized into a two-dimensional matrix with the two generated redundant information as the last two columns. In this case, each row contains data and verification information. Two special parity lines are used to protect the data - one is traditional row-level parity (Parity 1), and the other is specially constructed to provide additional redundancy (Parity 2). Third, by cleverly designing parity information, EVENODD can reconstruct lost data when two disks fail simultaneously, something many other storage solutions cannot achieve. Finally, EVENODD is more efficient in terms of computational complexity and is suitable for application in storage systems.

2.2 Traditional encoding and decoding methods of EVENODD code

For traditional encoding mechanism of EVENODD, it involves arranging data across ‘m’ data disks where m needs to be prime and two parity disks. Every disk is a column of m-1 symbols and the encoding process two main parity XOR calculations [6].

For parity1, also known as row parity, this step performs XOR operations across all the data in each row and generates a row parity which is stored in the first parity disks. It is a simple process [7]. For parity2, often referred to as diagonal parity, this enhancement is crucial for improving fault tolerance. A special sum S is required first:

$$\mathrm{S} = \sum_{t=1}^{m-1} a_{m-1-t,t} \tag{1}$$

Then, use $a_{l,m+1} = \mathrm{S} + \sum_{t=0,t\neq l+1}^{m-1} a_{<l-t>m,t}$. Here notation <x>m = y means x = y mod m where y is between 0 and m-1. For example, 0 mod 4 = 0, 3 mod 5 = 3, 5 mod 2 = 3, -3 mod 5 = 2 and -7 mod 5 = 3. The calculation method for this step is actually s plus the sum of several diagonals.

Table 1. Data Distribution and Parity Generation in EVENODD Code System

Disks/blocks	0	1	2	3	4	parity1	parity2
0	0	1	0	0	1		
1	1	1	0	1	0		

2	1	0	0	1	0		
3	1	1	1	0	1		

As show in the table 1. For parity1, just need calculate the sum of each row. Let us take an example of the first row, should calculate 0 + 1 + 0 + 0 + 1 = 0. Hence parity1 in the first row is 0.

Table 2. Initial EVENODD Parity Calculation with Data and Parity1

Disks/blocks	0	1	2	3	4	parity1	parity2
0	0	1	0	0	1	0	
1	1	1	0	1	0		
2	1	0	0	1	0		
3	1	1	1	0	1		

As show in the table 2. Then, let us calculate parity 2 in the first row. need calculate S first:

$$S = \sum_{t=1}^{m-1} a_{m-1-t,t} \quad (2)$$

Here m = 5, so $S = \sum_{t=1}^{4} a_{4-t,t} = a_{3,1} + a_{2,2} + a_{1,3} + a_{0,4} = 1 + 0 + 1 + 1 = 1$. Now, let us take l = 0 and calculate $a_{0,6} = S + \sum_{t=0,t\neq 1}^{4} a_{<0-t>5,t} = S + a_{0,0} + a_{3,2} + a_{2,3} + a_{1,4} = 1 + 1 + 1 + 0 = 1$. Now get parity2 in the first row.

Table 3. Computation of Parity2 for EVENODD Code

Disks/blocks	0	1	2	3	4	parity1	parity2
0	0	1	0	0	1	0	1
1	1	1	0	1	0		
2	1	0	0	1	0		
3	1	1	1	0	1		

As show in the table 3. Through the above method, can calculate all parities.

Table 4. Full Parity Calculation for EVENODD Encoding

Disks/blocks	0	1	2	3	4	parity1	parity2
0	0	1	0	0	1	0	1
1	1	1	0	1	0	1	1
2	1	0	0	1	0	0	0
3	1	1	1	0	1	0	0

As show in the table 4. If some data disks failed, can now decode them by parity1 and parity2.

For example, if two data disks failed, set x and y.

Table 5. EVENODD Decoding Process for Dual Disk Failures

Disks/blocks	0	1	2	3	4	parity1	parity2
0	? x	? y	0	0	1	0	1
1	1	1	0	1	0	1	1
2	1	0	0	1	0	0	0
3	1	1	1	0	1	0	0

As shown in Table 5, by using parity1 in the first row, x + y + 0 + 0 + 1 equals 0, which results in x + y = 1.By using parity2 in the first row, 1 + x + 0 + 1 + 1 equals 1, leading to the determination that x = 0 and y = 1. Thus, two data failures are successfully recovered.

3 System analysis and application research

3.1 Basic arrangement of traditional EVENODD code in coding

In conclusion, EVENODD is an erasure coding scheme developed for reliable data storage within distributed systems. The traditional EVENODD Code can withstand the failure of two data disks. Here is a basic outline of the EVENODD code arrangement. Firstly, the scheme incorporates two additional parity disks to safeguard data. Secondly, it computes the sum of each row using XOR operations and stores these in parity1 [8]. Thirdly, it calculates a special value, S, and combines S with the sums of values along different diagonals to determine the parity2 for each row. Lastly, should one or two data disks fail, the system is capable of reconstructing the lost data by solving linear equations using the parity information.

3.2 Extended encoding methods of EVENODD code

An extended encoding method for the EVENODD code is introduced. This method involves determining the first and second skew check columns of the EVENODD code to be constructed. The first 2 [k/2] check digits of the first skew check column undergo XOR operations with the first common factor. Simultaneously, the first 2 (k-1) check digits of the second skew check column undergo XOR operations with the second common factor [9]. By selectively performing XOR operations with common factors on these two sets of skew parity columns, the method effectively reduces the encoding complexity and update complexity of the EVENODD code. Assuming k is a positive integer representing the encoding of the information matrix for the EVENODD code to be constructed, the encoding method includes the following steps..

Calculate the first common factor S_1 and the second common factor S_2 .

$$S_1 = \sum_{j=1}^{k-1} b_{p-1-j,j} \tag{3}$$

$$S_2 = \sum_{j=1}^{k-1} b_{p-1-2j,j} \tag{4}$$

Construct the first skew check column. Assuming p is the prime number encoded by the information matrix of the EVENODD code to be constructed and let $a_{i,k+1}$ stands for the first skew check column. The formula of the first skew check column is that $b_{i,k+1} = \sum_{j=1}^{k-1} b_{(i-j)p,j} + S_1$ for $0 \le i \le 2(k/2) - 1$ (5)

$$b_{i,k+1} = \sum_{j=0}^{k-1} b_{(i-j)p,j} \quad \text{for} \quad 2(k/2) \le i \le p-2. \tag{6}$$

Construct the second skew check column. Let $b_{i,k+2}$ stands for the first skew check column. The formula of the first skew check column is:

$$b_{i,k+2} = \sum_{j=0}^{k-1} a_{(i-2j)p,j} + S_2 \quad \text{for } 0 \le i \le 2(k-1) \tag{7}$$

$$b_{i,k+1} = \sum_{j=0}^{k-1} a_{(i-2j)p,j} \quad \text{for} \quad 2(k-1) \le i \le p-2 \tag{8}$$

The first common factor performs XOR operations with the first 2[k/2] parity bits of the first diagonal parity column; the second common factor performs XOR operations with the first 2(k-1) parity bits of the second diagonal parity column. Then store the results obtained. successfully finish the encoding process. As show in the table 6.

Table 6. Detailed EVENODD Encoding Process with Extended Parity Calculations

D0	D1	D2	D3	D4	P5	P6	P7
b00	b01	b02	b03	b04	b00+b01+b02+b03+b04	b00+b02+b63+b04+S1	b00+b91+b72+b53+b34+S2
b10	b11	b12	b13	b14	b10+b11+b12+b13+b14	b10+b82+b63+b84+S1	b10+b82+b63+b74+S2
b20	b21	b22	b23	b24	b20+b21+b22+b23+b24	b20+b21+b02+b33+S1	b20+b01+b02+b73+b54+S2
b30	b31	b32	b33	b34	b30+b31+b32+b33+b34	b30+b31+b02+b03+S1	b30+b71+b82+b33+b54+S2
b40	b41	b42	b43	b44	b40+b41+b42+b43+b44	b40+b41+b32+b33+S1	b40+b41+b52+b33+S2
b50	b51	b52	b53	b54	b50+b51+b52+b53+b54	b50+b81+b62+b73+S1	b60+b11+b12+b13+b14+S2
b60	b61	b62	b63	b64	b60+b61+b62+b63+b64	b60+b11+b62+b63+S1	b60+b71+b82+b83+b34+S2
b70	b71	b72	b73	b74	b70+b71+b72+b73+b74	b70+b11+b82+b03+S1	b70+b61+b82+b83+b34+S2
b80	b81	b82	b83	b84	b80+b81+b82+b83+b84	b80+b81+b62+b53+S1	b80+b81+b82+b33+b14+S2
b90	b91	b92	b93	b94	b90+b91+b92+b93+b94	b90+b91+b92+b93+S1	b90+b81+b82+b83+b34+S2

As show in the table 7. Let me simplify it. Now, successfully finish the encoding process.

Table 7. Simplified EVENODD Encoding Results 1

D0	D1	D2	D3	D4	P5	P6	P7
b00	b01	b0	b0	b0	S0	A	K
b10	b11	b1	b1	b1	S1	B	L
b20	b21	b2	b2	b2	S2	C	M
b30	b31	b3	b3	b3	S3	D	N
b40	b41	b4	b4	b4	S4	E	O
b50	b51	b5	b5	b5	S5	F	P
b60	b61	b6	b6	b6	S6	G	Q
b70	b71	b7	b7	b7	S7	H	R
b80	b81	b8	b8	b8	S8	I	S
b90	b91	b9	b9	b9	S9	J	T

3.3 Extended decoding methods of EVENODD code

For the decoding of this method, it includes the following steps:

Obtain the EVENODD code which should be decoded. Determine the surviving correct information of the EVENODD code [10]. The surviving correct information includes all correct information columns and horizontal parity columns, the first and the second diagonal parity column. Determine the position of data disks failures, and select the appropriate check columns to decode.

As show in the table 8. Let me take an example for one failure.

Table 8. EVENODD Decoding Process for Initial Disk Failure Detection

D0	D1	D2	D3	D4	P5	P6	P7
?	b01	b02	b03	b04	Sum0	A	K
b10	b11	b12	b13	b14	Sum1	B	L
b20	b21	b22	b23	b24	Sum2	C	M
b30	b31	b32	b33	b34	Sum3	D	N
b40	b41	b42	b43	b44	Sum4	E	O
b50	b51	b52	b53	b54	Sum5	F	P
b60	b61	b62	b63	b64	Sum6	G	Q
b70	b71	b72	b73	b74	Sum7	H	R
b80	b81	b82	b83	b84	Sum8	I	S
b90	b91	b92	b93	b94	Sum9	J	T

Here, the EVENODD code is obtained and it is straightforward to determine the surviving correct information of the EVENODD code and identify the position of data disk failure, as the only failure is in the position of b00. There are three methods to recover it. The first method uses P5, similar to the traditional decoding process, which will not be detailed here. The second method utilizes P6, deriving that A = b00 + b92 + b83 + b74 + S1, offering a convenient approach for decoding. The third method involves P7, where it is understood that K = b00 + b91 + b72 + b53 + b34 + S2. However, this method is slightly more complex than the second, making the second method more appropriate. Decoding is performed using the second method. The equation ? + b92 + b83 + b74 + S1 equals A = b00 + b92 + b83 + b74 + S1, leading to the conclusion that ? = b00.

Table 9. EVENODD Decoding Process with Multiple Unknown Disk Failures

D0	D1	D2	D3	D4	P5	P6	P7
?	?	?	b03	b04	Sum0	A	K
b10	b11	b12	b13	b14	Sum1	B	L
b20	b21	b22	b23	b24	Sum2	C	M
b30	b31	b32	b33	b34	Sum3	D	N
b40	b41	b42	b43	b44	Sum4	E	O
b50	b51	b52	b53	b54	Sum5	F	P
b60	b61	b62	b63	b64	Sum6	G	Q
b70	b71	b72	b73	b74	Sum7	H	R
b80	b81	b82	b83	b84	Sum8	I	S
b90	b91	b92	b93	b94	Sum9	J	T

Table 10. EVENODD Decoding Process for Complex Failure Scenarios

D0	D1	D2	D3	D4	P5	P6	P7
? x	? y	? z	? k	b04	Sum0	A	K
b10	b11	b12	b13	b14	Sum1	B	L
b20	b21	b22	b23	b24	Sum2	C	M
b30	b31	b32	b33	b34	Sum3	D	N
b40	b41	b42	b43	b44	Sum4	E	O
b50	b51	b52	b53	b54	Sum5	F	P
b60	b61	b62	b63	b64	Sum6	G	Q
b70	b71	b72	b73	b74	Sum7	H	R
b80	b81	b82	b83	b84	Sum8	I	S
b90	b91	b92	b93	b94	Sum9	J	T

As shown in Table 9, consider a scenario where there are three errors in data disks. Using A, M, and O allows for their recovery. The decoding process closely mirrors earlier steps and will not be detailed further. As shown in Table 10, consider a more challenging scenario with four errors in the same row. First, x, y, and z are calculated similarly to previous cases. Subsequently, k can be resolved using Sum0, successfully restoring four failures.

3.4 Algorithm design and application of skew check column

The algorithm design and application of the skew check column aims to improve the efficiency and reliability of the data storage system. In a period when the amount of data is exploding, traditional storage is unable to deal with the challenges. Distributed storage enhances storage capabilities through data dispersion, but also increases the risk of failure. EVENODD code, as an MDS array code, effectively reduces the complexity and update difficulty by encoding the information matrix into an expanded coding matrix and using the common factor performing XOR operation on the skew check column. This design improves storage efficiency and fault tolerance, it is suitable for large-scale data storage requirements and it is a key technology to enhance system performance which has broad application prospects.

4 Conclusion

This paper has introduced innovative enhancements to the EVENODD encoding and decoding techniques, aimed at boosting data resilience in modern large-scale storage systems. Through the deployment of advanced methods such as optimized XOR operations and skew check columns, the proposed modifications have significantly reduced computational complexity and accelerated data recovery times. These advancements ensure robust data protection and enhanced operational efficiency, critical for systems where high data availability and rapid recovery are essential. The research not only fortifies the theoretical framework for future innovations in data storage but also extends the practical applicability of EVENODD codes, setting new benchmarks in data protection technologies. Moving forward, the study will delve deeper into refining these encoding and decoding methods to better suit the dynamic demands of distributed storage environments. Empirical testing and scalability assessments will play a pivotal role in validating the effectiveness of these modified EVENODD techniques across different system conditions and workloads. Future investigations will also explore adaptive coding strategies that dynamically respond to changes in system status and workload, ensuring sustained high performance and fault tolerance. By pushing the boundaries of current technology, these efforts aim to maintain the reliability and efficiency of data storage systems, preparing them to meet the evolving data demands and technological challenges of the future.

References

1. Y. Huang, R. Wang, M. Ju, X. Zhu, Y. Xie, "Reconstructing global daily XCO2 at 1°× 1° spatial resolution from 2016 to 2019 with multisource satellite observation data," Journal of Applied Remote Sensing, **18**(2), 028502-028502 (2024).
2. Z. Zhao, X. Zhu, X. Wei, X. Wang, J. Zuo, "Application of Workflow Technology in the Integrated Management Platform of Smart Park," in 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), vol. **4**, pp. 1433-1437, IEEE (2021).
3. R. Wang, J. Zhu, S. Wang, T. Wang, J. Huang, X. Zhu, Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking, Int. J. Multimed. Inf. Retr. **13**(4), 39 (2024).

4. 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).
5. Y. Zeng, Implementation and Efficacy of EVENODD Coding in High-Reliability Data Storage Architectures, Sci. Technol. Eng. Chem. Environ. Prot. **1**(9) (2024).
6. C. Qu, A Comparative Study on Fault-Tolerant Coding Techniques in RAID6 Analyzing the Efficiency and Limitation of EVENODD and RDP Codes, Sci. Technol. Eng. Chem. Environ. Prot. **1**(9) (2024).
7. X. Zhu, Z. Zhao, X. Wei, X. Wang, J. Zuo, Action recognition method based on wavelet transform and neural network in wireless network, Proc. 2021 5th Int. Conf. Digit. Signal Process., 60-65 (2021).
8. H. Xu, X. Zhu, Z. Zhao, X. Wei, X. Wang, J. Zuo, Research of Pipeline Leak Detection Technology and Application Prospect of Petrochemical Wharf, 2020 IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), Chongqing, China, 263-271 (2020).
9. L. Lin, Storage and Data Processing Technology in the Era of Information Explosion: from Traditional to Edge Computing, Highlights Sci. Eng. Technol. **87**, 66-75 (2024).
10. Y. Zhang, H. Zhao, X. Zhu, Z. Zhao, J. Zuo, "Strain Measurement Quantization Technology based on DAS System," in 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp. 214-218, IEEE (2019).