Proceedings of the 3rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/141/2025.21699

Research on Signal Detection Algorithms and Circuit Structures in Large-scale MIMO Systems

Difei Wang[1,a,*]
[1]*Shandong University, Jinan City, Shandong Province, 250100, China*
a. 2706549926@qq.com
**corresponding author*

Abstract: Large-scale Multiple-Input Multiple-Output (MIMO) systems are crucial for improving both the spectral efficiency and reliability of communication links. Additionally, they contribute to reducing the power consumption of base stations (BS), which is a significant concern in modern wireless networks. In this context, data detection plays an essential role, particularly in the uplink of large-scale MIMO systems. However, one of the major challenges lies in the substantial increase in computational complexity required for base stations to process the data efficiently in such systems. The conventional optimal detection techniques, such as ML detection, SD detection, MMSE detection, and ZF detection, are widely employed in MIMO systems. While these methods offer optimal performance in terms of detection accuracy, they suffer from extremely high computational demands, which can be impractical for real-time implementation in large-scale networks. Under the premise of approaching the performance of the optimal calculation methods, this paper analyzes three new algorithms with lower computational complexity, including iterative algorithms, matrix approximation algorithms, and K-Best algorithms. This paper conducts a detailed study on the advantages and disadvantages of the calculation methods and circuit structures of these algorithms.

Keywords: High-capacity MIMO detection, Iterative methods, Matrix approximation algorithms, K-Best algorithm

1. Introduction

The integration of Multiple-Input Multiple-Output (MIMO) technology with spatial multiplexing has become a cornerstone of modern wireless communication systems, including standards such as 3GPP LTE and IEEE 802.11n. This combination significantly improves both the data transmission rates and the reliability of communication links, making it a vital technology for high-performance networks. High-capacity MIMO, an advanced extension of traditional MIMO, further enhances spectral efficiency by supporting a greater number of antennas at the base station. This increase in antenna density not only contributes to higher network capacity but also extends the coverage area, potentially reducing the overall complexity of the base station infrastructure [1].

Despite these advantages, data detection in the uplink of High-capacity MIMO systems presents a considerable challenge. The need for high-performance detection algorithms capable of managing the high-dimensional data associated with the large number of antennas deployed at the base station is paramount. Moreover, small-cell systems like 3GPP LTE and LTE-A, which employ SC-FDMA,

DOI: 10.54254/2755-2721/141/2025.21699

introduce additional complexities. SC-FDMA increases the dimensionality of the detection problem, further raising the computational burden on the base station [1-2]. Traditional optimal detection methods, such as ML detection, SD, MMSE detection, and ZF detection, are commonly used in these systems. However, these methods suffer from high computational complexity, particularly when it comes to inverting high-dimensional matrices. The complexity increases exponentially with the number of transmission streams, making them less feasible for large-scale deployments in High-capacity MIMO systems [1-4].

This paper seeks to provide a comprehensive analysis of various detection algorithms and their corresponding circuit architectures, focusing on their strengths, limitations, and potential avenues for optimization. The algorithms discussed in this paper are categorized into four distinct groups, each with its own unique approach to reducing computational complexity and improving system performance. The first category includes iterative methods, which are known for their relatively low computational demands. These methods are particularly well-suited for statistical detection tasks in large-scale MIMO systems. By iteratively approximating optimal solutions, these techniques allow for efficient hardware implementation, especially when combined with specialized VLSI architecture designs. This approach minimizes resource consumption while maintaining accuracy [4-7]. The second group comprises matrix approximation algorithms, which utilize techniques like Neumann series approximation (NSA), tridiagonal matrix approximation, and matrix partitioning to reduce the complexity of matrix operations. These methods focus on simplifying the inversion process, enabling faster signal detection with fewer computational resources. By using matrix approximations to invert matrices efficiently, we can significantly lower the overall computational cost of the system [1][3][7][8]. The third category features the K-Best algorithm, which leverages the diversity and probabilistic characteristics of received signals, particularly in challenging low signal-to-noise ratio (SNR) environments. This algorithm estimates the most likely transmitted signal state by considering a limited number of the best candidates, thus enhancing the performance of the MIMO system without the need for exhaustive searching. Through sorting and selection techniques, the K-Best algorithm achieves high detection accuracy while minimizing computational overhead [9]. Finally, this paper also offers a detailed examination of the advantages and drawbacks of each algorithm, along with suggestions for potential improvements to further enhance their performance and efficiency in real-world applications.

2. Gauss-Seidel Iterative algorithm and Circuit Structure

2.1. Gauss-Seidel Iterative Algorithm (GS)

In a typical MIMO system, the transmission stream of each user undergoes encoding through a channel encoder, followed by mapping onto symbol points via modulation. The series of modulated values are energy-normalized before being transmitted as a signal vector, denoted as *s,* through the wireless channel. The wireless channel's behavior is modeled as follows:

$$y = H_s + n \tag{1}$$

In this model, the vector s encompasses the signals transmitted by all users, while y represents the received vector at the base station (BS). The matrix H corresponds to the uplink channel matrix, and n denotes the noise added to the signal at the base station. At the BS, a soft-output detection algorithm is used to compute the LLRs of the encoded bits, based on the received vector y and the channel matrix H. One of the most common MIMO detection methods applied in this context is the MMSE detection algorithm. For the MMSE approach, the estimated transmitted signal vector $\hat{s}$ is computed as:

DOI: 10.54254/2755-2721/141/2025.21699

$$\hat{s} = \left(G + N_0 I_K\right)^{-1} H^H y = W^{-1} y^{MF}, \tag{2}$$

Here, $y^{MF} = H^H y$ represents the output from the matched filter. $G = H^H H$ is the Gram matrix, and W is the minimum mean square error filtering matrix.

The Gauss-Seidel method (GS) [4] is applied to solve the N-dimensional linear equation $Ax=b$ Here, A is a Hermitian positive definite matrix of dimension $N\times N$, x is the vector of unknown solutions, and b represents the measurement vector. Since W is Hermitian positive definite, the GS method is used to iteratively solve for x, leading to the following decomposition:

$$W = D + L + L^H, \tag{3}$$

Here,D, L and L^H represent the diagonal, lower triangular, and upper triangular parts of the matrix W, respectively. The estimation of the transmitted signal vector s is then computed through iterations, expressed as:

$$S^{(i)} = (D + L)^{-1}\left(y^{MF} - L^H s^{(i-1)}\right), i = 1,2,\cdots, \tag{4}$$

Here, i denotes the iteration index, and $S^{(0)}$ is the initial guess for the solution. The choice of the initial solution is critical to the algorithm's convergence speed. A poor choice for the initial guess may lead to slow convergence or increased computational complexity. Therefore, it is generally recommended to avoid using a zero vector as the starting point. Instead, an approximation method such as the Neumann Series (NS) can be employed to improve the initial guess. The Neumann Series expansion provides an approximation for W^{-1}:

$$W^{-1} = \sum_{n=0}^{\infty}\left(I_k - X^{-1}w\right)^n X^{-1}, \tag{5}$$

By defining $X = D$ and retaining only the first two terms of the NS, we obtain the following approximations for the inverse of W^{-1}:

$$W_2^{-1} = D^{-1} - D^{-1}ED^{-1}, \tag{6}$$

Here, E represents the non-diagonal part of W. Substituting this approximation into the iteration formula results in a refined initial solution $S^{(0)}$, which improves the accuracy and efficiency of the Gauss-Seidel method.

$$S^{(0)} = W_2^{-1} y^{MF} \tag{7}$$

When comparing the performance of the Gauss-Seidel iterative algorithm to that of the NSA algorithm, it becomes evident that the GS method offers a significant advantage in terms of convergence speed. The results from the simulations are shown in Fig.1, where the performance of both algorithms under different antenna configurations is assessed.

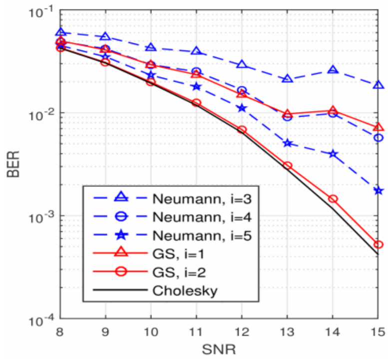


Figure 1: Performance comparison under different antenna configurations.[4]

To ensure a fair and accurate comparison between the two algorithms, it is important to consider the number of iterations used. Since the initial solution for the GS method is derived after performing two iterations of the NSA, it is necessary to adjust the iteration count for the NSA algorithm. Specifically, to make the comparison more meaningful, the NSA algorithm is set to run for 3 to 5 iterations, while the GS algorithm is limited to 1 or 2 iterations. As shown in Fig.1, when the number of iterations for both algorithms is set to 4 for the GS method and 1 for the NSA method, the performance of both algorithms is nearly identical. However, the GS method reaches this level of performance after just 1 or 2 iterations, while the NSA method requires 3 to 5 iterations to achieve comparable results. This indicates that the GS algorithm not only delivers similar results in terms of performance but also demonstrates superior convergence speed, requiring fewer iterations to reach a similar level of accuracy.

2.2. Gauss-Seidel Iterative Circuit Design

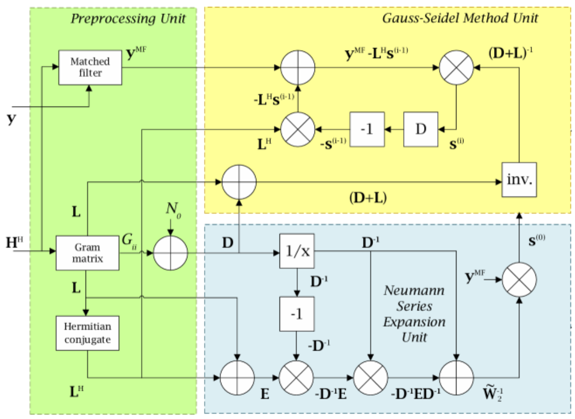


Figure 2: Architecture of the proposed GS-based soft-output data detection [4]

The circuit architecture for the Gauss-Seidel (GS) algorithm-based soft-output data detection system is illustrated in Fig.2. This architecture consists of three main components: the preprocessing unit, the second-order iterative NSA unit, and the GS algorithm unit. a) Preprocessing Unit: The primary task of the preprocessing unit is to perform matched filtering, which is represented by the equation $y^{MF} = H^{H}y$, and to calculate the matrix W. Given that W is a Hermitian matrix, the design employs a K×K lower triangular tridiagonal array to facilitate its computation. Each processing element (PE) in this array is responsible for performing a MAC operation, which involves two input values. Notably, the PE located on the main diagonal is configured to add N^0 and output the result *D*, which differs from the operation of the PEs on the off-diagonal positions. b) Second-Order Iterative Neumann

Series Approximation Unit: This unit is responsible for calculating W_2^{-1}, which serves as the initial value for $S^{(0)} = W_2^{-1}y^{MF}$. To compute the inverse of the diagonal matrix D^{-1}, a reciprocal module that utilizes a lookup table (LUT) is employed. This module is well-suited for implementation on FPGA hardware. The matrix D^{-1} is simplified and stored in vector form as d^{-1}. The diagonal nature of D^{-1} coupled with the Hermitian property of E, ensures that the product $D^{-1}ED^{-1}$ retains the Hermitian matrix property. By synchronously reading the corresponding rows of E and d^{-1} during each clock cycle of d^{-1}, the rows of W_2^{-1} can be efficiently generated. Furthermore, as all the elements involved are real numbers, this simplifies the hardware design complexity. c) Gauss-Seidel Algorithm Unit: The GS algorithm unit performs iterative operations to compute the updated transmission signal vector, $S^{(i)} = (D + L)^{-1}\left(y^{MF} - L^{H}s^{(i-1)}\right)$, ultimately solving for $\hat{s}$. Each iteration of the GS algorithm is broken down into three distinct stages. The first stage involves calculating the product L^{H} and $-s^{(i-1)}$ using the same pulse array employed for the matched filter. In the second stage, the subtraction $y^{MF} - L^{H}s^{(i-1)}$ is computed using an adder. Finally, in the third stage, another pulse array calculates $(D + L)^{-1}$ and multiplies it with the output from the second stage. Because the initial solution is typically quite close to the actual solution, only a small number of iterations are required to achieve satisfactory results.

3. Neumann Series Approximate Algorithms and Circuit Structures for Matrix Inversion Approximation

3.1. Neumann Series Approximation Algorithms for Matrix Inversion

The primary computational challenge in estimating the transmission vector $\hat{s}$ in the context of communication systems arises from the need to compute the inverse matrix W^{-1} of the minimum mean square error (MMSE) filter matrix W. In Section I, we provided a general outline of the NSA, which is a key method for matrix inversion. In this section, we will offer a detailed explanation of how we arrive at the Neumann series formula (6) and the iterative procedures involved in its computation.

To begin, let matrix W be represented as an invertible matrix X, such that the following condition holds:

$$\lim_{n\to\infty}\left(I - X^{-1}W\right)^{n} = 0 \tag{8}$$

This equation suggests that the matrix W^{-1} can be approximated using the Neumann series expansion. More specifically, the inverse of W can be expressed as:

$$W^{-1} = \sum_{n=0}^{\infty}\left(X^{-1}(X - W)\right)^{n} X^{-1} \tag{9}$$

Next, let us consider the matrix *W*, which is comprised of two components: the diagonal part *D* and the off-diagonal part *E*. Therefore, *W* can be written as: W = D + E. In (3), matrix *W* is assumed to be a diagonal-dominant matrix, which holds when the number of antennas *N* at the base station (BS) is sufficiently large. As a result, matrix *W* can be approximated as a dominant diagonal matrix *D*. Thus, let us define *X*=*D*. If matrix *D* is invertible, and the condition in Equation (8) is satisfied, we can rewrite the inverse of *W* as:

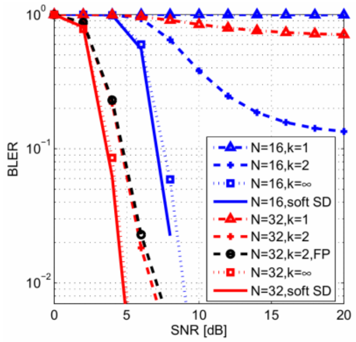
$$W^{-1} = (D + E)^{-1} = \sum_{n=0}^{\infty}\left(-D^{-1}E\right)^{n} D^{-1} \tag{10}$$

By truncating the series after k terms, we obtain a finite approximation of W^{-1}, which is expressed as:

$$W_k^{-1} = \sum_{n=0}^{k-1} \left(-D^{-1}E\right)^n D^{-1}, \tag{11}$$

From this approximation, we observe that when k=2, the formula corresponds exactly to Equation (6), which simplifies the operations by squaring the matrix M (where M is the matrix formed by the single-antenna users). As a result, using the Neumann series formula (6) allows for a rough estimation of W, thus reducing the complexity of linear detection operations.

Fig. 3 showcases the performance outcomes of the Neumann series-based matrix inversion approximation method applied to a massive multiple-input, multiple-output (MIMO) uplink system. The system utilizes a MIMO-orthogonal frequency-division multiplexing (OFDM) setup featuring 128 subcarriers and a 16-QAM modulation scheme. Additionally, the configuration assumes a 10-meter-long linear antenna array at the base station (BS), which plays a critical role in the system's performance evaluation.



(a) $M = 4$ users.

Figure 3: Uplink BLER performance in large-scale MIMO; curves for N = ∞ represent exact matrix inversion, while 'FP' curves indicate fixed-point implementations [1]

This simulation provides an evaluation of the bit error rate (BER) performance using the Neumann series matrix inversion approximation, comparing it to the exact matrix inversion method. The parameter M=4 indicates that the system is modeled with 4 single-antenna users. As depicted in the figure, when the number of iterations is minimized, the NSA performs more closely to the exact matrix inversion method as the total number of base station (BS) antennas N increases. In fact, the proposed method significantly improves upon exact inversion techniques, such as maximum likelihood (ML) detection, particularly in large-scale multiple-input multiple-output (MIMO) systems with lower computational complexity. However, the error generated by the approximation is more pronounced in smaller systems. This highlights the strong dependency of the method on the number of antennas at the base station. Larger antenna configurations lead to more accurate results, demonstrating that the efficiency of this approximation method increases with system size.

3.2. Neumann Series Approximation Circuit Architecture for Matrix Inversion

Fig.4 illustrates the high-level architecture of a circuit designed to perform the matrix inversion approximation using NS. This architecture is a VLSI circuit capable of efficiently computing binomial approximations with an exceptionally high throughput.

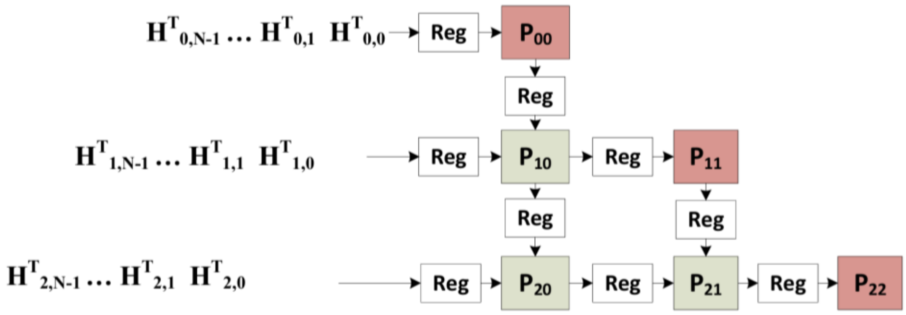


Figure 4: High-level architecture of the systolic array for M =3. [1]

This architecture is primarily designed to compute the Gram matrix G =H^HH. Since this matrix is symmetric, the computational effort can be reduced by only processing the lower triangular part of it. To achieve this, a specialized M×M lower triangular pulse array is employed for calculating the Gram matrix. In this setup, the value of *M* is chosen to be 3 for the specific case in consideration. Each element within the pulse array, referred to as a processing unit (denoted as Pkl), consists of a multiply-addition (MAC) unit. The input matrix, denoted as H^T, is fed into the system by shifting one column to the right for each successive cycle, ensuring that the data is processed in a sequential manner. Within each processing unit, MAC operations are carried out on two input operands, where multiplication and addition are performed simultaneously. To ensure the accuracy of computations in the architecture, there is a strategic delay of the values in the *i*-th row of H^T to the (*i*-1)-th cycle, which allows for a more precise data processing sequence. Once the input data reaches the diagonal processing unit in the array, it undergoes a conjugation operation, after which the result is passed to the lower half of the pulse array for further processing.

4. K- best Algorithm and Circuit Structure

4.1. K-Best Algorithm

The K-Best algorithm [9] is a widely used technique in MIMO signal detection. This algorithm utilizes a tree-like search structure, where the search gradually converges towards the optimal solution by retaining the K best candidate nodes at each layer, based on the PED. By focusing on a limited number of candidates, the K-Best algorithm provides higher throughput and a shorter search path compared to a full search approach, significantly reducing the computational complexity.

The core element in the K-Best algorithm is the PED value of the candidate nodes, as this directly influences the efficiency of the algorithm. In the study referenced by [9], two advanced types of K-Best detectors were introduced to further optimize performance. The first is an enumeration method that incorporates pre-screening, which effectively minimizes the number of candidate subtrees that need to be explored. The second improvement is the use of a SBOESA, which reduces the sorting complexity associated with the algorithm.

For the enumeration method of pre-screening, its core formula is as follows:

$$L_i = \tilde{y}_i - \sum_{j=i+1}^{M} r_{i,j} S_j \tag{12}$$

$$C_i = \frac{L_i}{r_{i,i}} \tag{13}$$

In these equations, $r_{i,j}$ represents an element of the R matrix, where $r_{i,i}$ is the diagonal element obtained from the QR decomposition of the matrix. The detection center point C_i is pivotal for pre-screening candidate nodes, as it helps eliminate less promising candidates from further consideration.

This reduces the number of sub-nodes that need to be processed, thus improving the efficiency of the detection process. Fig.5 illustrates the simulation results of bit error rate (BER) performance in a 4x4 MIMO system using 16-QAM, comparing scenarios with and without the pre-screening method.

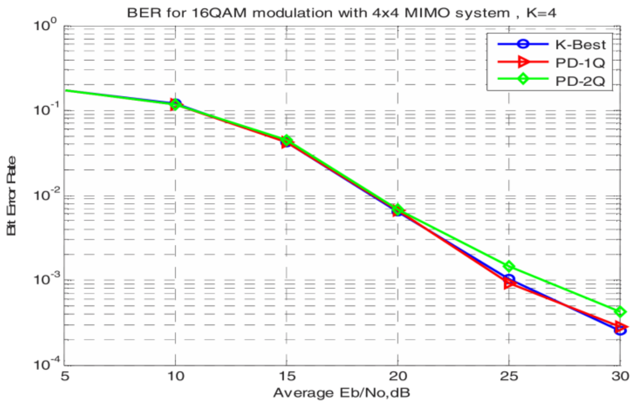


Figure 5: Simulations of the BER performances [9]

As shown in Fig.5, the proposed PD-1Q design demonstrates a comparable bit error rate performance to the direct K-Best design, even when no pre-filtering is applied. This indicates that the inclusion of the pre-screening step does not degrade performance in the 4x4 16-QAM mode. Consequently, when circuit complexity increases, the number of nodes eligible for pre-screening can be adjusted to optimize performance.

The Batcher Odd-Even Sort algorithm is a parallel sorting method that operates by dividing the input sequence into odd and even indexed elements, which are then sorted separately. The sorted sequences are subsequently merged to form the final sorted output. SBOESA, a simplified variation of this algorithm, refines the process by sorting only specific nodes using the odd-even method rather than sorting the entire output sequence. This approach significantly reduces the number of comparators needed for sorting operations. In SBOESA, the sorting network is designed to maximize the utilization efficiency of comparators. By carefully planning the sorting process, unnecessary comparison paths are eliminated, resulting in a more efficient use of resources. This reduction in comparator usage contributes to lowering the overall circuit complexity while maintaining high-performance levels in MIMO systems.

4.2. K-Best Circuit Architecture

The circuit architecture for the K-Best detector is illustrated in Fig.6. This detector employs a combination of pipeline and parallel processing techniques to significantly enhance throughput, making it well-suited for high-performance MIMO systems.

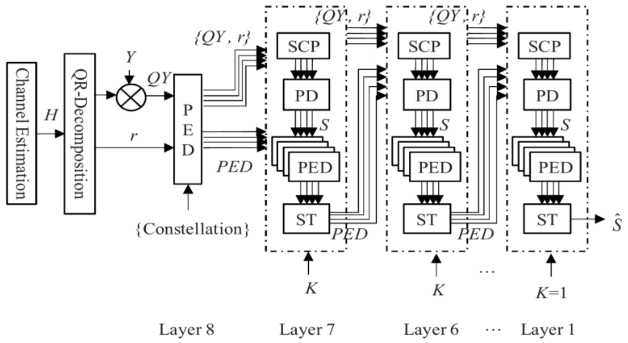


Figure 6: Architecture of the proposed 4x4 64QAM K-Best detector [9]

It consists of four functional blocks: Selective Clustering Points (SCP), Pre-Detection (PD), Partial Euclidean Distance (PED), and Sorting (ST) modules. In each detection layer, they are replicated and processed through parallel and pipeline operations. a) SCP Module: This module temporarily retains potential nodes generated during the current layer's operation. Once the PED calculation is completed for the current layer, the SCP module expands these retained nodes to the next layer. b) PD module: This module is used for the prescreening of sub-nodes to reduce the number of PEDs that need to be calculated. c) PED module: A key component of the K-Best detector, it calculates the PED of retained nodes. d) ST module: According to the proposed SBOESA, the PED values are sorted. The sorted nodes are arranged in ascending order of PED values, and the system selects the K nodes with the smallest PED values as candidates for the next layer of detection.

5. Analysis of Advantages and Disadvantages and Suggestions for Improvement

5.1. Analysis of Advantages and Disadvantages of Gauss-Seidel Iterative Algorithm and Circuit Structure

The main advantages of the Gauss-Seidel iterative algorithm are as follows: Its convergence speed is relatively fast, especially when the coefficient matrix is diagonally dominant; it can significantly contribute to achieving convergence. The architecture based on the Gauss-Seidel method is highly efficient for hardware implementation. Through FPGA implementation, this algorithm demonstrates advantages in hardware resource usage and flexibility. The architecture will flexibly adjust the number of iterations to achieve various objectives, making Gauss-Seidel iterative algorithms applicable in a range of scenarios. The main drawback of the Gauss-Seidel iterative algorithm lies in the selection of the initial solution, which significantly impacts the fusion speed and the quality of the final solution. If the initial solution is improperly chosen, it may result in slow convergence or failure to reach the correct solution. An approximate initial solution based on the Neumann series expansion is proposed to accelerate convergence, though this method still requires significant computational effort. The Gauss-Seidel iterative method typically requires the coefficient matrix to be both symmetric and positive definite. Although this condition holds in many practical applications, some systems of equations may require alternative methods or preprocessing.

The Gauss-Seidel iterative circuit structure offers high efficiency and flexibility. This structure can fully utilize the mathematical characteristics of the Gauss-Seidel iteration to achieve rapid convergence, enhance signal detection efficiency, and operate without relying on a specific number of iterations. However, due to the inherent characteristics of the Gauss-Seidel iteration, this circuit structure may have limitations in parallel processing, making it difficult to fully leverage modern hardware's parallel computing capabilities. Moreover, designing and implementing this circuit structure can be complex and require advanced technical skills.

5.2. An In-Depth Analysis of the Neumann Series Algorithm and Circuit Architecture for Approximating Matrix Inversion

The Neumann series algorithm for approximating matrix inversion brings several notable benefits. One of the key advantages is its ability to significantly lower the computational complexity compared to traditional methods. By selecting an optimal number of terms in the NS, the algorithm achieves performance that closely matches more advanced detection techniques, such as the SD (Successive Detection) method, but without the substantial computational cost. This makes it a suitable choice for real-time applications where computational resources are limited. However, the Neumann series algorithm does have some drawbacks. The primary disadvantage lies in the approximation error that the algorithm may introduce. While the error is generally minor under typical operating conditions, it can become problematic when certain parameters, such as the number of antennas in the base station

DOI: 10.54254/2755-2721/141/2025.21699

(denoted by N) or the signal-to-noise ratio (SNR), are not ideal. In scenarios where these factors are low, the algorithm's performance can deteriorate, leading to a higher likelihood of system failure. Additionally, the efficiency and computational demands of this method are heavily influenced by the specific characteristics of the MIMO (Multiple Input Multiple Output) system, such as the number of users with single antennas and the properties of the channel matrix.

When examining the Neumann series algorithm's associated circuit architecture, several advantages become apparent. This architecture efficiently handles matrix inversion approximation tasks by using a pipeline approach that sequentially computes the Gram matrix and the approximate inverse matrix. This pipelined approach is crucial for large-scale MIMO systems, which require the processing of substantial data volumes and complex matrix operations. Furthermore, the architecture's design is adaptable, allowing for parameter adjustments based on the configuration of the MIMO system. This makes it a versatile solution for a variety of system sizes and operational complexities. However, the implementation of this architecture comes with its own set of challenges. A major disadvantage is the complexity of ensuring precise control over the data flow and operation sequencing within the pipeline to guarantee accurate results. Additionally, the design of specialized hardware components, such as multiply-accumulate (MAC) units and multipliers, is necessary to perform the required mathematical operations. Another potential issue arises when the input matrix is singular, which can lead to substantial errors in the approximation results produced by this circuit architecture.

5.3. Examination of the K-Best Algorithm and Its Circuit Architecture: Advantages and Limitations

The advantages of the K-Best algorithm primarily include the following: when the value of K is large, the K-Best algorithm's performance can approach that of the ML detection algorithm, thus providing higher detection accuracy. The algorithm's performance and complexity can be balanced by adjusting the value of K. As K increases, the algorithm can explore more candidate nodes, thereby improving detection performance. Furthermore, this algorithm can be applied to various modulation schemes and MIMO configurations, such as 64QAM, demonstrating good scalability. The disadvantages of this algorithm mainly include the following: when K is large, the complexity of the Euclidean distance (PED) increases significantly, and the demand for storing candidate nodes also rises. Additionally, the traditional K-Best algorithm may have limitations in sorting and search schemes, such as efficiency problems encountered when using depth-first or breadth-first search.

The primary advantages of the K-Best detector circuit architecture include the following: the K-Best detector effectively reduces the number of leaf nodes that need to be accessed by adopting strategies like pre-screening and simplified sorting, thus improving detection speed. By implementing Batcher's parity sorting algorithm (SBOESA), the K-Best detector can streamline the sorting process and further reduce hardware costs. Moreover, the K-Best detector can select the optimal K value combination by simulating the BER performance under different K value schemes, achieving the best balance between detection performance and hardware efficiency. The main disadvantages of this circuit architecture include the following: the performance of the detector is largely dependent on the selected K value and sorting algorithm. Improper parameter selection may lead to performance degradation or reduced hardware efficiency, thus requiring more hardware resources or higher power consumption.

6. Conclusion

This study introduces three distinct algorithms for detecting data in the uplink of large-scale MIMO systems, along with the corresponding circuit architectures for each. These algorithms are designed

DOI: 10.54254/2755-2721/141/2025.21699

to significantly alleviate the computational complexity encountered in real-world applications. The Gauss-Seidel iterative method, coupled with its VLSI architecture, accelerates the convergence process and enhances overall system performance by utilizing the Hermitian positive definite property of the MMSE filtering matrix. Meanwhile, the Neumann series matrix inversion approximation algorithm simplifies the matrix inversion process by truncating the series at a finite number of terms, thus reducing the computational burden. The K-Best algorithm, through a pre-screening technique and the use of Batcher's parity sorting algorithm (SBOESA), reduces the number of candidate sub-nodes that must be accessed, which helps lower the complexity of sorting operations. Future investigations into data detection for large-scale MIMO systems will focus on further refinement and optimization of algorithms based on Gauss-Seidel iterations and Neumann series matrix inversion methods. By exploring the mathematical underpinnings of these algorithms in greater detail, the research aims to uncover more efficient iterative and approximation techniques. The results of these studies are anticipated to significantly improve both the speed and accuracy of data detection in large-scale MIMO systems.

References

[1] *Wu, M., Yin, B., Vosoughi, A., et al. (2013). Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink. 2013 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE. https://doi.org/10.1109/ISCAS.2013.6572301*

[2] *Yin, B., Wu, M., Wang, G., et al. (2014). A 3.8Gb/s large-scale MIMO detector for 3GPP LTE-Advanced. IEEE International Conference on Acoustics. IEEE. https://doi.org/10.1109/ICASSP.2014.6854328*

[3] *Wang, F., Zhang, C., Yang, J., et al. (2015). Efficient matrix inversion architecture for linear detection in high-capacity MIMO systems. IEEE. https://doi.org/10.1109/ICDSP.2015.7251869*

[4] *Wu, Z., Zhang, C., Xue, Y., et al. (2016). Efficient architecture for soft-output high-capacity MIMO detection with Gauss-Seidel method. 2016 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE. https://doi.org/10.1109/ISCAS.2016.7538940*

[5] *Yin, B., Wu, M., Cavallaro, J. R., et al. (2015). VLSI design of large-scale soft-output MIMO detection using conjugate gradients. IEEE. https://doi.org/10.1109/ISCAS.2015.7168929*

[6] *Zhang, P., Liu, L., Peng, G., et al. (2016). Large-scale MIMO detection design and FPGA implementations using SOR method. IEEE. https://doi.org/10.1109/ICCSN.2016.7586650*

[7] *Ji, Y., Wu, Z., Shen, Y., et al. (2018). A low-complexity high-capacity MIMO detection algorithm based on matrix partition. IEEE. https://doi.org/10.1109/SiPS.2018.8598436*

[8] *Zhang, Y., Jing, S., Zhang, Z., et al. (2019). Efficient belief propagation detection based on channel hardening for high-capacity MIMO. IEEE. https://doi.org/10.1109/ICASSP.2019.8682557*

[9] *He, J. J., & Fan, C. P. (2015). Design and VLSI implementation of novel pre-screening and simplified sorting based K-best detection for MIMO systems. IEEE. https://doi.org/10.1109/VLSI-DAT.2015.7114511.*