Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/115/2025.18516

Linear Detection and Circuit Structure Based on ZF, MMSE Matrix Iteration and Operations

Kun Li[1,a,*]

[1]*Dalian University of Technology, Dalian City, Liaoning Province, 116024, China*
a. 1563190815@163.com
**corresponding author*

Abstract: The Multiple-Input Multiple-Output (MIMO) technology substantially boosts the reliability and data transfer rates of communication links by enabling the parallel transmission of multiple data signals over the same frequency bandwidth, utilizing multiple transmit and receive antennas.Despite these enhancements, traditional MIMO systems are approaching their theoretical throughput limits, which underscores the need for developing new algorithms to tackle computational complexity.This paper categorizes the primary technical approaches into three types: iterative approximation methods, matrix operation-based methods, and the emerging Joint channel Estimation and Data detection (JED) algorithms.It delves into the Neumann series approximation for efficient matrix inversion in large-scale MIMO systems, the Gauss-Seidel iterative approach for soft-output detection, and LDLT decomposition-based algorithms for matrix inversion efficiency. Additionally, the paper emphasizes the JED algorithm, such as PROX, which presents a low-complexity solution for large Single-Input Multiple-Output (SIMO) wireless systems.

Keywords: Massive MIMO, Signal detection algorithm, MMSE detection, FPGA implementation, VLSI.

1. Introduction

Multi-input Multi-output (MIMO) technology is an essential component of High-Speed Packet Access (HSPA), Long Term Evolution (LTE), and LTE Advanced, and it is also a core technology within LTE[1]. Compared with single-antenna systems, MIMO technology can transmit multiple data signals in parallel over the same frequency bandwidth using multiple transmit and receive antennas, to significantly improve the reliability and Data transfer rate of communications links. By spatial multiplexing, multiple data streams can be transmitted on different spatial paths at the same time without mutual interference, thus improving the spectral efficiency and system capacity. This technique is expected to bring significant improvements in spectrum efficiency, link reliability, and coverage of small-scale MIMO systems [2].

Nevertheless, the theoretical limit of throughput for conventional point-to-point and multi-user MIMO wireless systems is being reached. For example, common Minimum mean square error and maximum likelihood approximation methods, because of their computational complexity, we must create new algorithms to meet this challenge. Fortunately, through relentless effort and wisdom, our predecessors have developed a variety of solutions to meet the needs of current MIMO systems. At present, we can categorize the main technical approaches into three types: first, iterative

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/115/2025.18516

approximation methods; second, matrix operation-based methods; and lastly, the emerging JED algorithm.

First, a major approach is iterative approximation, such as Neumann series approximation: this method introduces a hardware architecture that efficiently inverts matrices for linear detection in large-scale MIMO systems. The architecture employs the Neumann series approximation to achieve matrix inversion through an iterative calculation process. This approach boasts reduced hardware complexity, increased operational frequency, and offers both flexibility and scalability. [3].Second, the Gauss-Seidel (GS) iterative approach, which proposes an efficient hardware architecture based on the GS approach for soft-output detection in large-scale MIMO systems, by using truncated Neumann series approximation as initialization to accelerate the convergence rate and flexibly adjust the number of iterations to meet different application requirements [4] .

In addition, there are algorithms designed for matrix evolution, such as the algorithm based on LDLT decomposition: this method proposes an extended recursive diagonal algorithm based on LDLT decomposition, by using the characteristic of 2×2 matrix inversion, the large-scale matrix inversion is realized, the purpose of this paper is to provide an efficient, low resource consuming and fast computing matrix inversion method for MIMO system implemented on FPGA hardware[5]. A distributed feed-forward architecture based on FPGA and parallel equalization algorithm can also be used to realize decentralized equalization of large-scale multi-user multi-input multi-output (MU-MIMO) systems using large-scale MIMO approximate message passing (Lama), to reduce the computational complexity and interconnection bandwidth bottlenecks in traditional centralized architectures[6].

In conclusion, a selection of innovative algorithms for Joint channel Estimation and Data detection (JED) has been developed, including PROX, which is characterized by its low complexity. This algorithm, along with its matching FPGA design, is specifically tailored for large-scale Single-Input Multiple-Output (SIMO) wireless systems. The BCR framework facilitates the efficient computation of solutions for the approximate maximum likelihood JED problem, while a scalable VLSI architecture is employed to ensure both high performance and minimal complexity in implementation. These innovative methods provide a new direction for the development of MIMO technology and help to further improve the performance of wireless communication systems [7].

2. Prerequisite Knowledge

In this article we will focus on a large-scale multi-user (MU) MIMO system controlled by a base station (BS) equipped with N antennas and communicating with M (where M is less than N) antenna-equipped users. We will pay special attention to uplink transmission, that is, multiple users transmit data to the base station at the same time.

2.1. Ssystem Model

In uplink MIMO systems, each user's bitstream undergoes encoding by a channel encoder prior to being mapped to specific symbols. The normalized signal vector is then transmitted through a wireless channel. The system model can be summarized as follows:

$$Y = H * x + n \tag{1}$$

Among them, Y represents the signal vector received by the base station (BS), H is the channel matrix of the uplink, x is the transmission vector including the transmission symbols of all users, and n is used to simulate additive noise at the base station.

DOI: 10.54254/2755-2721/115/2025.18516

2.2. MMSE Detection for Massive MIMO

In MU-MIMO systems, detection techniques are required to distinguish the data streams sent by each user. Linear MMSE detection is a popular detection method that can find a balance between reducing error rates and maintaining controllable algorithm complexity. The standard linear MMSE processing flow first constructs the Gram matrix $G=H^H H$, and then calculates the Gram matrix W used for regularization as:

$$w = G + N_0 I_{Nt} \tag{2}$$

Subsequently, the transmitted signal x̂ can be deduced as follows:

$$\hat{x} = W^{-1} y^{MF} \tag{3}$$

Among them, $y^{MF}=H^H$y represents the matched filter vector. Then calculate the approximate log likelihood ratio (LLR) of the encoded bitstream based on the estimated value $\hat{x}$[8].

In the uplink of massive MIMO systems, when the ratio of the number of receiving antennas to the number of transmitting antennas (N/M) is large, the linear MMSE detection method can provide near ideal performance. This is because when the number of antennas M remains constant and N tends towards infinity, the column vectors of the channel matrix H can be almost considered orthogonal. This characteristic guarantees that the Gram matrix G and the matrix W, which is regularized, are both positive definite and Hermitian.

3. Iterative Approximation Algorithm and Circutt

The significant computational complexity in circuit design is often caused by solving the inverse matrix of regularization matrix Specifically, the computational process of M-matrix regularization requires a lot of computational steps, which is especially obvious in large-scale MIMO systems. The computational complexity of M-matrix regularization is often prohibitive In order to solve this problem, I propose an iterative approximation, which can simplify the calculation process to a great extent and thus reduce the complexity of the operation.

3.1. Algorithm details and process

Compared with direct matrix inversion, we can use Neumman series approximation to simplify the operation [9] Specifically, if A

$$\lim_{n\to\infty}(I - x^{-1}M)^n = 0 \text{ or}$$

$$\lim_{n\to\infty}(I - xM)^n = 0 \tag{4}$$

Then the invertible matrix x can be approximated as:

$$M^{-1} = \Sigma_{n=0}^{\infty}\left(x^{-1}(x - M)\right)^n X^{-1} \tag{5}$$

In this case, M can be decomposed into main diagonal D and diagonal E such that $M=D+E$, and Matrix M is a diagonal dominant matrix, that is, $M \approx D$ Matrix H plays a key role in Gaussian channel When the system enters the large antenna limit state, that is, the number of antennas N tends to infinity and the number of users M is constant, we can see that the matrix A will approach the unit matrix I This characteristic is very important for large-scale MIMO systems because it makes it possible to realize matrix inversion with low complexity By taking advantage of this property, we can realize efficient processing in large-scale MIMO systems by noticing the complexity of matrix operations.

DOI: 10.54254/2755-2721/115/2025.18516

Matrix M tends to diagonal matrix M in large-scale MIMO systems, so we can use Neumann series to approximate the inverse matrix of M We can reconstruct the inverse matrix formula of M=D+E under the assumption that M is invertible and under certain conditions (4)

$$M^{-1} = (D+E)^{-1} = \Sigma_{n=0}^{\infty}\left(-D^{-1}E\right)^{n} D^{-1} \tag{6}$$

By keeping just the initial k terms of the Neumann series, we derive the k-term approximation for the inverse of matrix M^{-1}

$$\widetilde{M}_K^{-1} = \Sigma_{n=0}^{k-1}\left(-D^{-1}E\right)^{n} D^{-1} \tag{7}$$

The detection algorithm operates virtually the same as the matched filter detector in the specific case where K is 1 This is because Formula (2) only scales each component in the yMF and does not introduce additional calculation steps. When k is equal to 2, we can use a simplified method to approximate the inverse of matrix M, that is, $\widetilde{M_2^{-1}} = D^{-1} - D^{-1}ED^{-1}$. The computational complexity of this approximation is O(M^2), which is significant compared with O(M^3) operation required for accurate inversion M^{-1}).By substituting M^{-1} for M^{-1}, We thereby significantly streamline the computation involved in linear detection for large-scale MIMO systems.Circuit structure

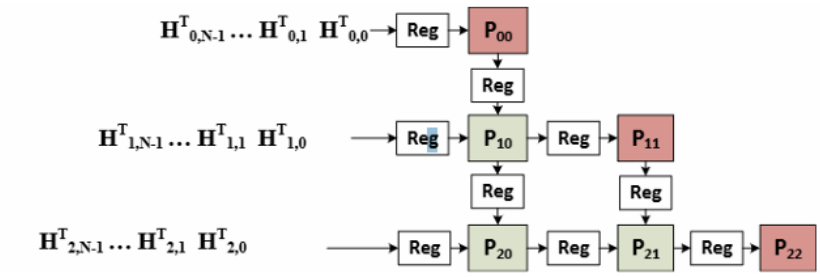

Figure 1: High-level architecture of a pulsating array with M = 3 [10]

In this architecture, the inversion task can be divided into two parts, namely

1)Calculate the Gram Matrix G= =H^H*H

2) Calculate $\widetilde{M_2^{-1}} = D^{-1} - D^{-1}ED^{-1}$

And only the lower triangle part can be processed.

And for the circuit architecture:

1)P00 is the multiplication and accumulation of the first row and the first column;

2)Reg stores and transmits basic raw data for operation;

3) The input data on the left side is not aligned. In order to realize the beat, wait for the upper layer data to be passed in together with the layer data and do the operation.

4. Method and circuit for matrix calculus operation

From the previous analysis, we understand that the main reason for the limited performance of MIMO systems lies in solving the inverse of the regularization matrix. In order to simplify the calculation process, we can use the specific properties of the matrix to greatly optimize the calculation steps.

4.1. Algorithm details and process

We propose a novel extended recursive diagonal algorithm based on LDLT decomposition for efficient matrix inversion. The algorithm first transforms an arbitrary matrix into a symmetric matrix

DOI: 10.54254/2755-2721/115/2025.18516

in order to apply the LDLT factorization. Then, by using the characteristics of the 2×2 inverse lower triangular matrix, the inversion of the whole N × N matrix is realized by starting from the 3 × 3 block matrix and gradually expanding to the N×N block matrix by an iterative method. When the algorithm is implemented on FPGA hardware, it exhibits lower resource consumption and faster processing speed.

In this method, the given arbitrary matrix is first preprocessed into a symmetric matrix $A = LDL^T$ for LDLT factorization.

$$A = LDL^T \tag{8}$$

The matrix L is an N × N lower triangular matrix, the elements on the diagonal are all 1, and the elements on the off-diagonal are represented asl_1, l_2, l_3etc, each value only appears in the position directly above the corresponding row, and all the elements below the diagonal are 0; the matrix D is a matrix with a constant diagonal and all other elements are 0.Through LDLT, it is easy to get:

$$A^{-1} = \left(L^{-1}\right)^T \times D^{-1} \times L^{-1} \tag{9}$$

After the LDLT decomposition, the recursive diagonal algorithm is applied. The resulting lower triangular matrix L and diagonal matrix D are used for subsequent calculations.

4.2. Circuit structure and details

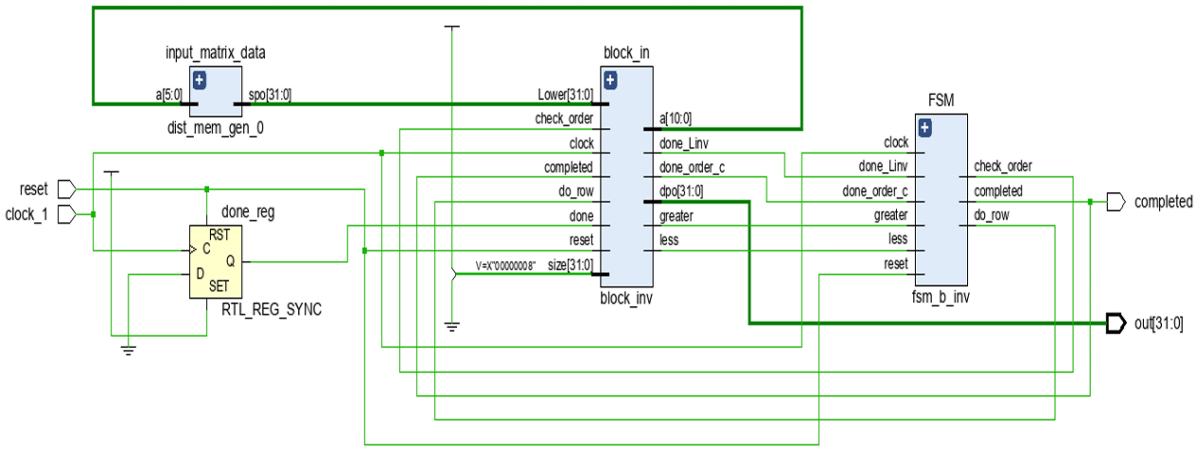

Figure 2: Detailed description and schematic design of the extended recursive diagonal intelligent algorithm [5]

The circuit mainly includes two parts: the data path and the controller. The data path is responsible for processing the input matrix data and performing LDLT decomposition and recursive diagonal algorithm to calculate the inverse of the matrix. The controller manages the entire calculation process, including the clock signal and reset signal, and checks the execution sequence and completion of the algorithm. In addition, the circuit also includes a distributed memory generator for storing input matrices, diagonal matrices, lower triangular matrices, their inverses, and intermediate computation results. High level synthesis (HLS) is used to optimize the use of hardware resources and speed up the calculation. The hardware implementation of the circuit structure significantly reduces the resource consumption and shortens the processing time, and is particularly suitable for being applied to a MIMO system to improve the communication performance.

DOI: 10.54254/2755-2721/115/2025.18516

5. Other new and innovative algorithms

In addition to the above methods, the predecessors have also developed some new and innovative solutions. For example, We have unveiled an FPGA design tailored for soft-output data detection in large-scale MIMO systems. In an effort to lower the computational intensity of data detection tasks, the design incorporates a customized rendition of the Conjugate Gradient Least Squares (CGLS) algorithm. Unlike existing linear detection algorithms for large-scale MIMO systems, this approach avoids the two most complex tasks of computing the Gram matrix and matrix inversion, while still being able to compute soft outputs. The architecture leverages a configurable array of processing elements to perform the CGLS algorithm efficiently on hardware[11].

5.1. Algorithm details and process

The Conjugate Gradient Least Squares (CGLS) algorithm is a proficient iterative approach for resolving extensive linear equation systems. In this paper, the CGLS algorithm is applied to address the ensuing regularized least-squares optimization problem:

$$\hat{x} = arg\,min\|y - H\hat{x}\|_2^2 + \frac{N_0}{E_s} \tag{10}$$

Among them, by solving the above problem directly in an iterative manner, CGLS is able to compute X with low complexity without forming the Gram matrix G and the inverse matrix A^{-1}. With each cycle of the CGLS algorithm, a comparatively inexpensive and standard matrix-vector multiplication is needed, amounting to O(MN) operations. For systems of dimension M by N, CGLS reaches the precise solution within N iterations. However, in practical applications, achieving a precise approximation of the problem often requires less than N iterations.

5.2. Circuit structure and details

Algorithm 1 CGLS-based soft-output MMSE detection

1.**inputs: H** and y

2: **initialization:**

3: b = $H^H y$and $\bar{\mathbf{H}} = [H^T, \sqrt{N_0/E_S}\, I_U]$

4: $v_o = 0, r_o = b, p_o = r_o, and\ t_o = H\bar{p}_o$

5: **for** k = 1, ..., K do

6: $\alpha_k = \|r_{k-1}\|_2^2 / \|t_{k-1}\|_2^2$

7: $v_k = v_{k-1} + \alpha_k P_{k-1}$

8: $e_k = \bar{H}\hat{}Ht_{k-1}$

9: $r_k = r_{k-1} - \alpha_k e_k$

10: $\beta_k = \|r_k\|_2^2 / \|r_{k-1}\|_2^2$

11: $P_k = r_k + \beta_k P_{k-1}$

12: $t_k = \bar{H} P_k$

13: compute $\mu_{i|k}$, $\forall i$, as in Sec. III

14: compute $\rho_{i|k}$, $\forall i$, as in Sec. III

DOI: 10.54254/2755-2721/115/2025.18516

15: **end for**

16: **outputs**: $\overline{x_k} = v_k$, $\mu_{i|k}$, $\forall i$,and $\rho_{i|k}$, $\forall i$[11]

Algorithm 1 deploys a soft-output MMSE detection algorithm that is based on CGLS, addressing the signal detection challenge in large-scale MIMO systems via an iterative method. The algorithm first receives the channel matrix H and the received signal vector y as inputs, and initializes them to calculate the initial residual vector r_0 and the search direction p_0During the iteration, the algorithm uses the current residual and search direction to update the solution vector v_k, and calculates the new residual and search direction. This process is repeated until a predetermined number of iterations k is reached.At each iteration, the algorithm also calculates the intermediate quantities $u_{i|k}$ and $\rho_{i|k}$, which are crucial for the final calculation of the log-likelihood ratio (LLR) value for each user.Finally, the algorithm outputs the estimated transmitted signal vector $\hat{x}_k$ and u _ $u_{i|k}$ and $\rho_{i|k}$ for each user, which will be used in the subsequent signal processing or decoding process.In this way, the algorithm can effectively reduce the computational complexity without explicitly calculating the Gram matrix and matrix inversion,while providing near-optimal detection performance.

5.3. Circuit structure and details

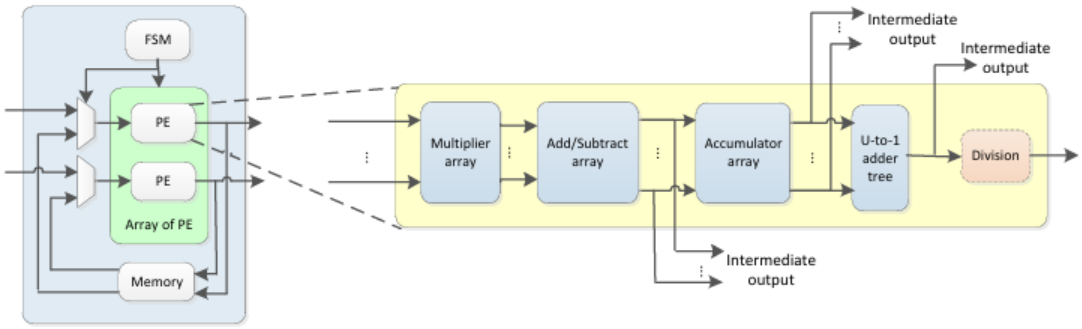

Figure 4: Reconfigurable Processing Element Array (PE) for Soft Output Data Detection Based on CGLS [11]

This diagram shows a complex digital signal processing hardware architecture consisting of multiple reconfigurable processing elements (PEs) that work together with specific array configurations to perform efficient mathematical operations. Each processing element incorporates an addition/subtraction array, an accumulator array, an addition tree, and a division module to process the signal. The addition/subtraction array is responsible for basic arithmetic operations, the accumulator array is used to accumulate intermediate results, and the addition tree combines the outputs of multiple accumulators into a single result for fast parallel data processing. The division module handles the operations that require division, such as normalization. The whole architecture is designed to provide hardware acceleration for signal detection algorithms in large-scale MIMO systems, such as CGLS algorithm, to improve the speed and efficiency of data processing through parallel processing and optimized operation flow.

6. Conclusion

Multiple-Input Multiple-Output (MIMO) technology significantly enhances the reliability and data transfer rates of communication links by transmitting multiple data signals in parallel over the same frequency bandwidth using multiple transmit and receive antennas. Despite the improvements, traditional MIMO systems are nearing their theoretical throughput limits, necessitating the

DOI: 10.54254/2755-2721/115/2025.18516

development of new algorithms to address computational complexity. This paper categorizes the main technical approaches into three types: iterative approximation methods, matrix operation-based methods, and the emerging Joint channel Estimation and Data detection (JED) algorithms. The paper discusses the Neumann series approximation for efficient matrix inversion in large-scale MIMO systems, the Gauss-Seidel iterative approach for soft-output detection, and LDLT decomposition-based algorithms for efficient matrix inversion. The paper also highlights the JED algorithm, such as PROX, which offers a low-complexity solution for large Single-Input Multiple-Output (SIMO) wireless systems. Future research in the field of MIMO technology may focus on developing more advanced algorithms and hardware architectures to further improve spectral efficiency and reduce energy consumption, such as using artificial intelligence and machine learning technologies to optimize signal processing, or exploring new antenna technologies and materials to achieve more efficient spatial multiplexing and beamforming.

References

[1] *Faridi, M. S. (2015). Spatial Correlation Based MIMO Mode Switching in LTE Downlink System. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 438-442). IEEE. DOI: 10.1109/CICT.2015.80*

[2] *[2] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas, " IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.*

[3] *Wang, F., Zhang, C., Yang, J., Liang, X., You, X., & Xu, S. (2015). Efficient Matrix Inversion Architecture for Linear Detection in Massive MIMO Systems. IEEE Transactions on Wireless Communications, 14(4), 248-252. doi:10. 1109/TWC.2015.2390351*

[4] *Wu, Z., Zhang, C., Xue, Y., Xu, S., & You, X. (2016). Efficient Architecture for Soft-Output Massive MIMO Detection with Gauss-Seidel Method. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 1886-1890). IEEE. doi:10.1109/ICASSP.2016.7472116*

[5] *Lingala, R. R., Dash, J. C., & Sarkar, D. (2022). Matrix Inversion on Hardware Board Using FPGA for Application in MIMO Systems. In Proceedings of the 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 978-1-6654-9781-7/22/$31.00).IEEE.DOI: 10.1109/ CONECCT55679.2022.9865679*

[6] *Li, K., Jeon, C., Cavallaro, J. R., & Studer, C. (2017). Decentralized Equalization for Massive MU-MIMO on FPGA. In Proceedings of the 2017 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1532-1536). IEEE. DOI: 10.1109/CONECCT.2017.8263979*

[7] *Casta˜neda, O., Goldstein, T., & Studer, C. (2017). FPGA Design of Low-Complexity Joint Channel Estimation and Data Detection for Large SIMO Wireless Systems. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 978-1-4673-6853-7/17/$31.00).IEEE.DOI: 10.1109/ ICASSP.2017.7952308*

[8] *C.Studer, S.Fateh and D.Seethaler, ASIC implementation of soft input soft-output MIMO detection using MMSE parallel interference cancellation, IEEEJ. Solid-State Circuits, vol. 46, no. 7, pp. 1754 1765, Jul.2011.*

[9] *G.Stewart, MatrixAlgorithms: Basic decompositions, 1998.*

[10] *Wu, M., Yin, B., Vosoughi, A., Studer, C., Cavallaro, J. R., & Dick, C. (2014). Approximate Matrix Inversion for High-Throughput Data Detection in the Large-Scale MIMO Uplink. IEEE Transactions on Wireless Communications.*

[11] *Yin, B., Wu, M., Cavallaro, J. R., & Studer, C. (2015). VLSI Design of Large-Scale Soft-Output MIMO Detection Using Conjugate Gradients. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1498-1502). IEEE. DOI: 10.1109/ICASSP.2015.7178730*