Accepted at International Conference on Distributed Computing in Sensor Systems, IEEE DCOSS 2020

Efficient and Low-Complexity Iterative Detectors for 5G Massive MIMO Systems

Robin Chataut
Dept. of Computer Science and Engineering
University of North Texas
Denton, Texas, 76207, USA
Robin.Chataut@unt.edu

Robert Akl
Dept. of Computer Science and Engineering
University of North Texas
Denton, Texas, 76207, USA
Robert.Akl@unt.edu

***Abstract*—The global bandwidth shortage in the wireless communication sector has motivated the study and exploration of sub-6 GHz wireless access technology known as massive Multiple-Input Multiple-Output (MIMO). Massive MIMO groups together antennas at both transmitter and the receiver to provide high spectral and energy efficiency. Although massive MIMO provides enormous benefits, it has to overcome some fundamental implementation issues before it can be implemented for 5G networks. One of the fundamental issues in Massive MIMO systems is uplink signal detection, which becomes inefficient and computationally complex with a larger number of antennas. In this paper, we propose three iterative algorithms to address the issues of uplink signal detection in massive MIMO systems. The simulation results, compared to the traditional detection algorithms, show that the proposed iterative massive MIMO uplink signal detection algorithms are computationally efficient and can achieve near-optimal Bit Error Rate (BER) performance. Additionally, we propose novel hardware architectures for the proposed detection algorithms to identify the required physical components and their interrelationships.**

***Index Terms*—5G, bit error rate, massive MIMO, signal detection, spectral efficiency**

I. Introduction

With globalization, the past few years have seen tremendous growth in the wireless data traffic and fulfilling these needs, the cellular systems are deployed within a few hundred-meter distances, and wireless Local Area Networks (LAN) are placed almost everywhere. Along with increased mobile broadband service, the introduction of new concepts like the Internet of Things (IoT) and Machine-to-Machine communication (M2M) are also contributing to the increased wireless traffic. The global deployment of cellular service cultivates the cell phone users to be used to the mobile data in their day to day life tremendously. The services like video calling, online gaming, social media applications like Facebook, Twitter, WhatsApp, have changed our life drastically with the capabilities of 3G, 4G/LTE, and 5G like lower latency and high data rate [1]. In the next few years, technology like augmented and virtual reality, ultra-high-definition video, 3D video, and features like a mobile cloud will become popular to enrich the ultimate user experience. A full cell phone connected world is expected in the next few years, mainly characterized by growth in users, connectivity, data traffic volume, and a wide range of applications.

The fifth-generation (5G) network is expected to address the increasing demands in wireless data traffic and provide the user with improved speed, reliability, and accuracy. An efficient wireless access technology that can increase the wireless area throughput, reduce latency, and provide the user with a reliable connection is imminent. Massive MIMO (Multiple-Input-Multiple-Output) is the most enthralling sub-6 GHz wireless access technology to deliver the needs of future generation networks. Massive MIMO is considered as one of the key enabling technology for 5G technology, and it brings together antennas, radios, and spectrum together to enable higher capacity and speed for incoming 5G world [2]-[5]. Massive MIMO is an extension of MIMO technology, which involves using hundreds and even thousands of antennas attached to a base station to improve spectral efficiency and throughput. Massive MIMO's capacity to increase throughput and spectral efficiency has made it a crucial technology for emerging wireless standards. The key here is the considerable array gain that massive MIMO achieves with a large number of antennas [6].

A. Related work and Motivation

In massive MIMO systems, due to a large number of antennas, the uplink signal detection at the base station becomes computationally complex and reduces the achievable throughput. Also, all the signals transmitted by users superimpose at the base station to create interference, which also contributes to the reduction of throughput and spectral efficiency. There has been extensive research to find the optimal signal detection method for massive MIMO systems, which can provide better throughput performance with lower computational complexity. The conventional non-linear detectors like Sphere Decoder (SD) and Successive Interference Cancellation (SIC) yield good performance. The computational complexity is higher with more antennas, which makes them infeasible for massive MIMO systems [8].

Several linear detectors have been considered for uplink detection in massive MIMO, such as Maximum Likelihood (ML), Zero Forcing (ZF), and Minimum Mean Square Error (MMSE) [9]- [11]. ML is an optimal detector in massive MIMO, and it minimizes the probability of error, but for large antennas systems, the algorithm has prohibitive complexity.

The ZF methods mitigate the inter-antenna interference, but for ill-conditioned channel matrices, additive noise gets increased [12]. The MMSE detector has better performance than the ZF detector as it also considers the noise power during the detection [13]. Although the ML, MMSE, and ZF detection algorithms provide optimal throughput performance, they involve matrix inversion during the processing, which makes them computationally inefficient for large antenna systems like massive MIMO. The ZF and MMSE algorithms combined with the SIC was considered to cancel the interference from previously detected symbols, but the optimal performance was not achieved [14]. Several iterative methods such as Neumann Series Approximation (NSA) method [15], Richardson method [16], Successive Over-Relaxation method (SOR) [17], and Jacobi Iterative method [18] have been considered, but computational complexity was slightly reduced when compared to conventional linear methods. Other linear methods such as Gauss Seidel (GS) [19] and Conjugate Gradient (CG) [20] were also considered for massive MIMO, but they were also not found optimal for massive MIMO uplink signal detection.

In this paper, we have proposed three iterative algorithms along with their hardware architecture for uplink signal detection in massive MIMO systems. These proposed iterative algorithms are designed by considering the uplink signal detection problem as Least Square (LS) problem, Huber fitting problem, and Compressed Sensing (CS) problem, respectively [21]- [23]. These three different problems are then solved using the proposed Least Square Regression Selection (LSRS) method, Huber Fitting based Alternating Direction Method of Multipliers (HADMM) method, and modified Approximate Message Passing (MAMP) method, respectively. The numerical results through simulations show that the proposed algorithms are computationally efficient and achieve optimal Bit Error Rate (BER) performance.

B. Contributions

The main contributions of this work are outlined as follows:

- We investigate the uplink signal detection problem in massive MIMO systems and propose three iterative algorithms LSRS, HADMM, and MAMP to address uplink signal detection issues in Massive MIMO systems.
- We propose hardware architectures for the proposed detection algorithms to investigate the components required to design the system and interrelation between them. Also, these hardware architectures allow us to measure the computational complexity of the system.
- We assess the performance of the proposed algorithm by comparing it with traditional detection algorithms like ZF and MMSE. The simulations results show that proposed iterative algorithms can achieve optimal BER performance and are computationally efficient than traditional detection algorithms.

C. Paper Outline

The rest of the paper is organized as follows: Section II describes the system model required for conducting the simulations. The proposed iterative algorithms for signal detection in massive MIMO are discussed in section III. The numerical results and analysis from the simulation are provided in section IV, whereas section V presents the hardware architecture and compares the computational complexity of proposed and conventional algorithms. Finally, section V concludes the paper summarizing the key ideas and algorithms.

II. System Model for Uplink Massive MIMO Systems

A massive MIMO uplink system with M antenna at the base station is considered. These M antenna at the base station are simultaneously communicating with N users (M $\gg$ N) having a single antenna, as shown in Fig. 1. We have Rayleigh fading channel between the base station and the user terminal, where the channel gain from any base station antenna to a user terminal is described by a zero-mean circularly symmetric Gaussian random variable [24]. A sufficiently long cyclic prefix is also considered to mitigate the effect of inter-symbol interference (ISI). Each user terminal encodes their own data stream, and the encoded data stream is mapped into a constellation point with a finite alphabet set $\mathcal{O}$, such as binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), and quadrature amplitude modulation (QAM). The signal transmitted by the user terminal is $x \in \mathbb{C}^N$ and the signal received at the base station, $y \in \mathbb{C}^{M\times 1}$ is given by:

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

Where, H is channel matrix between the user terminal and the base station, and each element of H is independent and identically distributed (i.i.d) with zero mean and unit variance, i.e., $H \sim \mathcal{CN}(0,1)$. $n \in \mathbb{C}^M$ is Additive White Gaussian Noise (AWGN) and each element of n is i.i.d with zero mean and finite variance, i.e., $n \sim \mathcal{CN}(0,\sigma^2 I)$. The maximum likely-hood estimation of (1) is given as:

$$\hat{x}_{ML} = \underset{x\in X^N}{\operatorname{argmin}}\, ||y - Hx||_2^2 \tag{2}$$

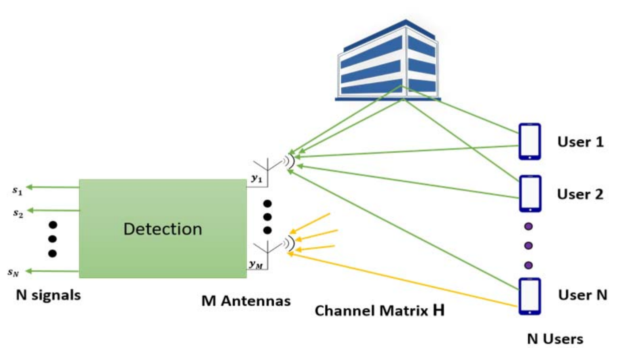


Fig. 1: An uplink Massive MIMO system with M antennas at the base station simultaneously communicating with N users having single antenna.

Equation (2) can be solved using conventional detectors such as Maximal Ratio Combining (MRC), ZF, and MMSE, but for a large number of antenna computational complexity increases exponentially as matrix inversion is required in these traditional detection algorithms. The MRC detector reduces the noise but fails to reduce the interference, the ZF detector reduces the inter-antenna interference regardless of noise enhancement, and MMSE detector reduces both interference and noise, by minimizing the error between the signal transmitted, and the signal received and achieves near-optimal performance. The estimated signal matrix for MRC, ZF, and MMSE is given as:

$$E = \begin{cases} H & \text{for MRC} \\ (H^H H)^{-1} H^H & \text{for ZF} \\ (H^H H + \frac{\sigma_n^2}{\sigma_x^2})^{-1} H^H & \text{for MMSE} \end{cases} \tag{3}$$

Here, E is the estimated signal matrix, H^H is Hermitian transpose of channel matrix H, σ_n^2 is noise variance, and σ_x^2 is the signal variance.

III. Proposed Algorithms

The traditional detection methods for massive MIMO uplink are not optimal for large antennas system since they require complex matrix inversion, which increases the computational complexity of the system. We have proposed three iterative detection algorithms, which avoids complicated matrix inversion to reduce computational complexity and achieves optimal error performance.

The problem presented in (2) can be rewritten as the LS problem:

$$f(x) = \underset{x \in X^N}{\operatorname{argmin}} \, ||y - Hx||_2^2 \tag{4}$$

Since the problem in (4) has no closed solution, an iterative method can be used to solve it. We have proposed the LSRS algorithm, which solves the least square problem by finding the best fit from the set of available regressors. Here, we try to fit a line as closely as possible to include as many points possible, and that line is called linear regressor. We try to find a line that minimizes the least squared distance to each of the points. The proposed algorithm is presented in Algorithm 1. During the initialization phase of the algorithm, all the required variables such as step size γ, dual variable λ, and Lagrangian parameter ρ are initialized. Fortunately, complicated matrix inversions and multiplications can be computed during the preprocessing, which helps to reduce the overall computational complexity of the algorithm. During the first step of the iteration, the estimated user signal x is minimized, keeping z and λ constant. This minimization is achieved using forward-backward solves [25]. Then, z is minimized, keeping x and λ constant, which is a projection onto a non-convex set. Finally, during the dual update, the dual variable λ is minimized, keeping x and z constant to detect the signal transmitted by the user, x_k. The dual update looks at the previous value of x and z and penalizes the model depending upon how much it violates the constraint γ. The dual update is significant in the algorithm, which ensures the convergence of the algorithm.

The estimation problem presented in (2) can be rewritten as the as Huber fitting problem :

$$f(x) = \underset{x \in X^N}{\operatorname{argmin}} \, \phi_{hub} ||y - Hx||_2^2 \tag{5}$$

where, Huber Penalty function ϕ_{hub} is given as:

$$\phi_{hub}(v) = \begin{cases} v^2, & \text{if } |v| \le 1 \\ (2|v| - 1), & \text{if } |v| > 1 \end{cases} \tag{6}$$

These problems are very tricky to solve in practice, but iterative methods can solve it with low complexity. We have proposed an efficient and low complex HADMM algorithm to solve the problem in (5). The ADMM algorithm is widely used for solving convex optimization problems, and it makes variable updates very smooth and more relaxed [26], [27]. In contrast, Huber fitting makes the function less sensitive to the outliers present in the signal. Huber fitting performs linear regression under the assumption that there are outliers in the received signal. During initialization, all the required variables are such as step size γ, dual variable λ, Nesterov's acceleration parameter α, and Lagrangian parameter ρ are initialized to a suitable value. All the complicated matrix inversions and multiplications are computed during the preprocessing step to reduce the computational complexity of the algorithm. During the iteration, we do ADMM like an update. First, we minimize the estimate of user signal x keeping z and λ constant. Then, using the Huber fitting function, the variable z is minimized,

Algorithm 1: LSRS Detection for Uplink 5G Massive MIMO Systems

Input Parameters:
    y, H
Initialization:
    $\widehat{\lambda}_0 = \hat{z}_0 = 0$
    $\gamma = 0.25$, $\rho = 5$
Preprocessing:
    $A = H' * H + \beta * I$
    $\hat{R} = cholesky(A, lower)$
    $R = squeeze(\hat{R})$
    $Q = squeeze(\hat{R})'$
    $\hat{H} = H' * y$
Iteration:
    for $k = 1$ to k_{max} **do**
        $P = \hat{H} + \rho * (z_k - \lambda_k)$
        $x_k = (Q/(R/P))$
        $C = (x_k | cardinality(x_k)) \le N$
        $z_k = \Pi_C(\mathrm{x_k{+}z_k})$
        $\lambda_k = \hat{\lambda}_k + \gamma * (x_k - z_k)$
    end for
Output: x_k

keeping x and λ constant. The advantage of using Huber fitting function during z minimization is that it is less sensitive to noise, less complicated than ADMM z minimization, and ensures the faster convergence of the algorithm. Finally, during the dual update, the dual variable λ is minimized, keeping x and z constant to detect the signal transmitted by the user, x_k. The dual update is the most crucial update during the iteration as it ensures the convergence of the algorithm. Here, we look at the previous value of x and z and penalize the model depending on how much we violate the γ constraint. The proposed HADMM algorithm is presented in Algorithm 2.

The convergence of algorithm 2 is further increased by using Nesterov's extrapolation method [28]. During extrapolation, we compute the value of the Nesterov's parameter α, and then, using the computed value of α, we extrapolate the dual variable λ. Since the update in the dual variable might create inexactness, a further update on variable z is required to guarantee the convergence. It is imminent that this new update introduces extra computational complexity, but this complexity is negligible when compared to the convergence achieved after extrapolation. This extrapolation in dual variable increases the convergence of algorithm from the order $1/k$ to $1/k^2$.

The problem presented in (2) can be rewritten as compressed sensing problem as:

$$f(x) = \min_{x \in X^N} ||y - Hx||_2^2 \tag{7}$$

Algorithm 2: HADMM Detection for Uplink 5G Massive MIMO Systems

Input Parameters:
y, H, N_0
Initialization:
$\lambda_0 = \widehat{\lambda}_0 = z_o = \hat{z}_0 = 0$
$\gamma = 0.25$, $\alpha_0 = 1$, $\rho = 1.5$
Preprocessing:
$\beta = N_0 * E_S^{-1}$
$A = \left(H^H * H + \beta * I\right)$
$A = R * R^T$
$\tilde{R} = R^{-1}$
$\hat{y} = H^H * y$
ADMM Iteration:
for $k = 1$ to k_{max} **do**
$x_k = (\tilde{R}^T * \tilde{R}) * (\hat{y} - \beta(\widehat{\lambda}_k - \hat{z}_k)$
$z_k = \frac{\rho}{1+\rho} * (x_k + \widehat{\lambda}_k) + \frac{1}{1+\rho} * S(x_k + \widehat{\lambda}_k, \frac{1}{1+\rho})$
$\lambda_k = \widehat{\lambda}_k + \gamma(x_k - z_k)$
$\alpha_{k+1} = \frac{1+\sqrt{1+4\alpha_k^2}}{2}$
$\widehat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}} * (\lambda_k - \lambda_{k-1})$
$\hat{z}_{k+1} = z_k + \frac{\alpha_k - 1}{\alpha_{k+1}} * (z_k - z_{k-1})$
end for
Output: x_k

Algorithm 3: MAMP Detection for Uplink 5G Massive MIMO Systems

Input Parameters:
y, H
Initialization:
$r^0 = y$
$x^0 = 0_{N \times 1}$
Iteration:
for $k = 1$ to k_{max} **do**
$\alpha = x^{k-1} + H^T * r^{k-1}$
$\theta = |\mathrm{real}(\min(\alpha)\ )|$
$x^k = \mathrm{S}(\alpha, \theta)$
$b = \frac{1}{M} * \frac{\|x\|_2}{\|x\|_1}$
$r^k = y - H * x^k + b * r^{k-1}$
end for
Output: x_k

Compressed Sensing makes use of the sparsity of the signal and recovers signals using fewer samples than the Nyquist-Shannon sampling theorem. We have proposed a novel MAMP algorithm to solve this compressed sensing problem. The AMP algorithm was initially designed to solve the absolute double shrinkage problem in compressed sensing [29]. The proposed MAMP algorithm presented in Algorithm 3 uses an iterative thresholding scheme to detect the signal transmitted by the user.

During initialization, all the required variables, such as residual vector r^0 and initial user signal x^0, are initialized. Each iteration in the algorithm estimates the user signal, using soft thresholding with parameter α and the threshold value θ. The thresholding function is defined as:

$$S(\alpha, \theta) = \begin{cases} \alpha - \theta, & \text{if } \alpha > \theta \\ 0, & \text{if } |\alpha| \leq \theta \\ \alpha + \theta, & \text{if } \alpha < -\theta \end{cases} \tag{8}$$

Then, the sparsity measure is computed by the ratio of l_2-norm and l_1-norm. Unlike the conventional AMP algorithm, we didn't use l_0-norm to calculate the sparsity measure because the performance of l_0-norm is poor in the presence of noise and its derivative doesn't contain the information of the transmitted signal. Also, sparsity measure with the ratio of l_2-norm and l_1-norm follows attributes such as Scaling, Robin-Hood, and Bill Gates [30], which helps in efficient detection of the user signal. Finally, the residual error is computed and minimized until we reach the maximum number of iterations.

IV. Numerical Results and Analysis

A thorough analysis of all the detection algorithm for uplink massive MIMO systems is presented in this section. We have compared the performance of the proposed algorithms with conventional detection algorithms like ZF and MMSE. The system model presented section II, and the simulation parameter presented Table I are used for conducting the simulations.

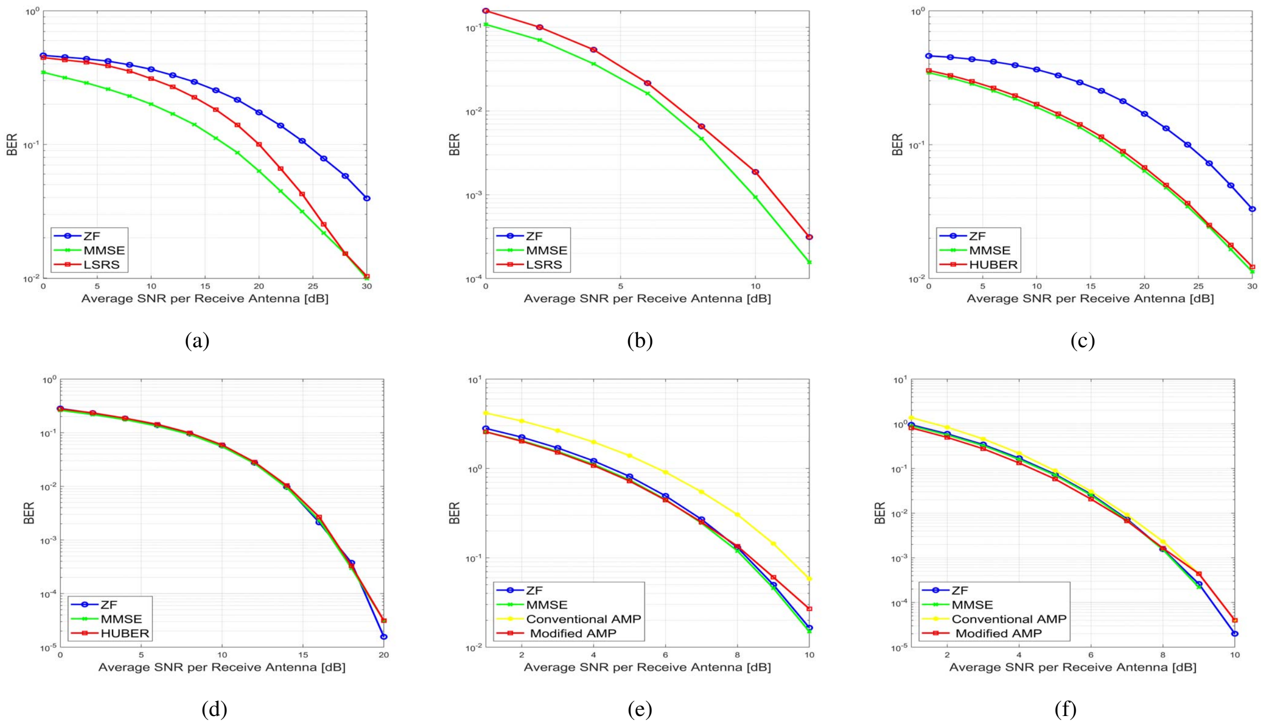


Fig. 2: Comparison of BER performance comparison of all the proposed algorithms and conventional detection algorithms (a) LSRS algorithm with M=16 (b) LSRS algorithm with M=32 (c) HADMM algorithm with M=16 (d) HADMM algorithm with M=32 (e) MAMP with M=16 (f) MAMP with M=32

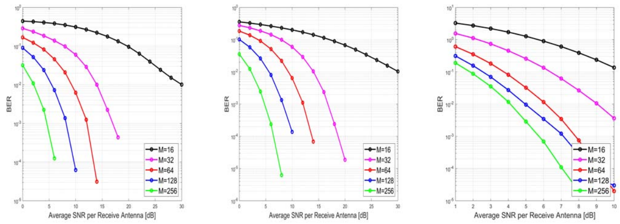
TABLE I: Simulation Parameters

Parameter	Value
System Bandwidth	20 MHz
Number of Users	8/16/32
Number of Antenna at Base Station	16/32/64/128/256
Signal Variance	2
Signal to Noise Ratio	1dB - 25dB
Channel	Rayleigh Fading
Modulation Scheme	BPSK, QPSK, 16QAM, 64QAM

All the simulations are performed in MATLAB under Mac OS, with 3.4 GHz Intel Core i7 processor and 10GB of RAM. The number of Monte Carlo run is set to 100000, and the number of iterations for each algorithm is set to 5.

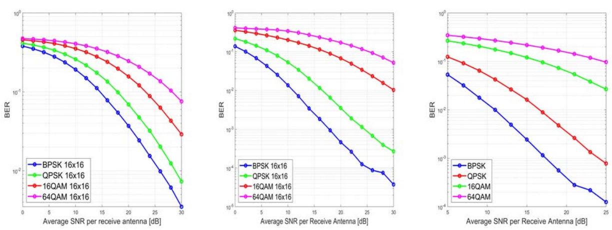
Fig. 2a and Fig. 2b shows the BER performance of the proposed LSRS algorithm with 16 users (N=16), 16-QAM modulation, and 16 to 32 antennae at the base station (M=16 or 32). For M=16, the LSRS algorithm has better performance than ZF, but MMSE outperforms the proposed algorithm, and for M=32, MMSE has better performance. In contrast, the proposed algorithm and ZF have almost similar performance. Fig. 2c and Fig. 2d shows the BER performance of the HADMM detection algorithm with 16 users (N=16), 16-QAM modulation, and 16 or 32 antennae at the base station (M=16 or 32). For M=16, the HADMM detection algorithm and the MMSE algorithm have almost similar performance, and they outperform the ZF algorithm. With M=32, all the algorithms have almost identical performance throughout the SNR curve. Fig. 2e and Fig. 2f shows the BER performance of the MAMP algorithm for 64 or 128 base station antennas (M=64 or 128) with 12 single antennas users (N=12) and using 16-QAM modulation. For M=64, the proposed MAMP algorithm has similar BER performance when compared to ZF and MMSE, and it performs better than the traditional AMP algorithm. For M=128, there is a vast improvement in the BER performance for all the algorithms, and the proposed AMP algorithm performs better than traditional AMP.

Fig. 3a, Fig. 3b, and Fig. 3c shows the BER performance of the HADMM detection algorithm, LSRS detection algorithm, and MAMP algorithm for the increasing number of base station antennas with 16 users and 16QAM modulation. At BER $=10^{-4}$, for LSRS detection, changing base station antenna from 64 to 128 achieved 3.9dB gain, and an additional gain of 3.5dB is achieved if the number of base station antennas is increased to 256. For HADMM detection at BER$=10^{-3}$, changing base station antenna from 64 to 128 achieved 3dB



(a) (b) (c)

Fig. 3: BER performance comparison with varying number of base station antennas with 16 users (N=16) and using 16-QAM modulation. (a) LSRS detection algorithm. (b) HADMM detection algorithm. (c) MAMP detection algorithm.



(a) (b) (c)

Fig. 4: BER performance comparison with different modulation scheme with 16 users (N=16), and 16 base station antennas (M=16) for different detection algorithms. (a) LSRS detection algorithm (b) HADMM detection algorithm (c) MAMP detection algorithm.

gain, and an additional gain of 3.1dB is achieved if the number of base station antenna is increased to 256. For MAMP detection, the results were similar to that of LSRS and HADMM algorithms.

Fig. 4a, Fig. 4b, and Fig. 4c shows the BER performance of the HADMM detection algorithm, LSRS detection algorithm, and MAMP detection algorithm for different modulation schemes (BPSK, QPSK, 16QAM, and 64QAM) with 16 users and 16 base station antennas. At BER=10^{-1}, for LSRS detection algorithm, changing modulation scheme from 64QAM to 16QAM, we achieved almost 4dB gain, and further 3.9dB gain is achieved if the modulation scheme is changed to QPSK. At BER=10^{-1}, for the HADMM detection algorithm, changing modulation scheme from 64QAM to 16QAM, we achieved almost 8dB gain, and further 10.5dB gain is achieved if the modulation scheme is changed to QPSK. Similar results were obtained to the MAMP algorithm with increased modulation order. Thus, we can conclude that we can improve the BER performance for the proposed detection algorithms by reducing the modulation order. But in the cases where the aim is only to increase the data rate, a higher modulation order would be beneficial. Thus, the selection of modulation order suitable for a particular application is an essential factor to consider.

V. Hardware Architectures and Computational Complexity Analysis

In this section, we propose hardware architectures for the proposed algorithms summarized in Algorithm 1, Algorithm 2, and Algorithm 3.

The proposed hardware architecture for the LSRS detection algorithm is shown in Fig. 5. The proposed hardware architecture consists of three basic units, a preprocessing unit, a signal detection unit, and an SINR computation unit. The architecture takes H, y as inputs, and all the required variables such as augmented Lagrangian parameter(ρ), dual variable (λ), vector z and step size (γ) are initialized and also fed as inputs to the preprocessing unit. The preprocessing unit performs inversion of gram matrix (A) using Cholesky decomposition, which is required during user signal estimation. The Cholesky decomposition expresses the gram matrix as the product of a triangular matrix and its transpose.

$$\begin{aligned} A &= R * R^T \\ A^{-1} &= (R^{-1})^T * (R^{-1}) \end{aligned} \tag{9}$$

Where, R is a lower triangular matrix with positive diagonal elements, and R is called the Cholesky factor of A, and R^T is the conjugate transpose of R. The squeeze function in the algorithm returns the array R or Q with the same elements but with all singleton dimensions removed. The matrix-vector multiplication to calculate $\hat{H}$ in the preprocessing unit reduces the overall computational complexity of the algorithm. The signal detection unit computes the update of x, z, and λ. The x-update is performed using forward-backward solves keeping variable z and λ constant, z update is performed keeping x and λ constant, which is a projection on to a non-convex set. In simple words, the z update is like intermediate sorting. Finally, the λ update is performed, keeping x and z constant, which ensures the convergence of the algorithm by suitably choosing the step size γ. Finally, the SINR computation unit calculates the SINR for the detected user signal.

Similarly, the proposed hardware architecture for the HADMM detection algorithm is presented in Fig. 6. Compared to the hardware architecture of the LSRS algorithm, the proposed hardware architecture for the HADMM algorithm

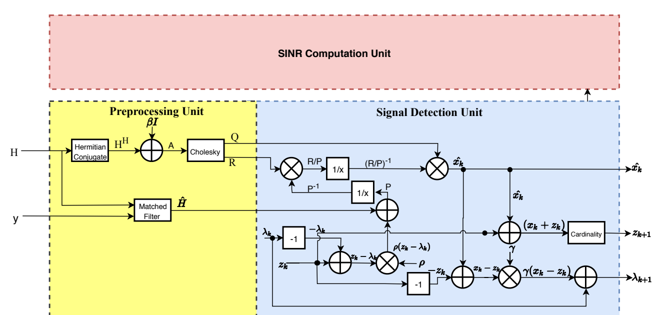


Fig. 5: Proposed hardware architecture for the LSRS uplink signal detection algorithm for Massive MIMO systems.

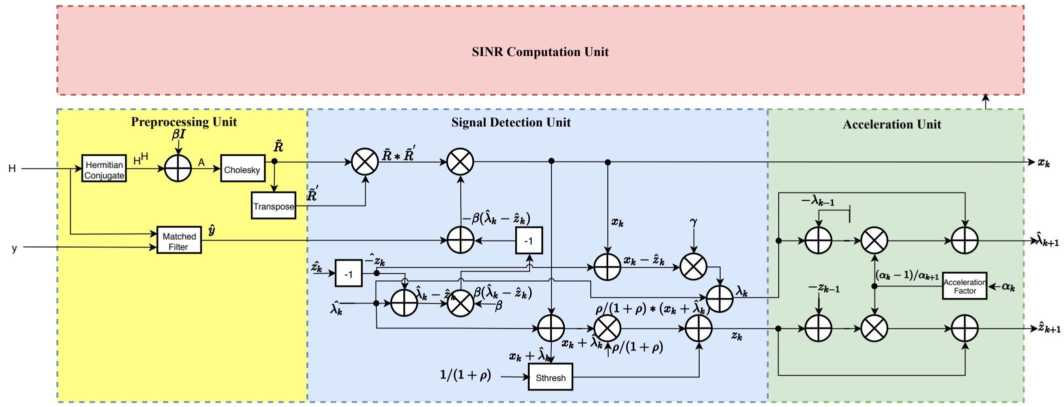


Fig. 6: Proposed hardware architecture for the HADMM uplink signal detection algorithm for Massive MIMO systems.

contains an additional acceleration unit along with a preprocessing unit, a signal detection unit, and an SINR computation unit. The proposed architecture takes H, y, and as inputs, and all the required variables such as augmented Lagrangian parameter (ρ), over-relaxation parameter (α), dual variable (λ), vector z and step size (γ) are initialized and fed as inputs to the preprocessing unit. The preprocessing unit performs inversion of gram matrix (A) using Cholesky decomposition, which is required during user signal estimation. The Cholesky decomposition expresses the gram matrix as the product of a triangular matrix and its transpose. The matrix-vector multiplication to calculate $\hat{y}$ is completed during preprocessing to reduce the computational complexity of the algorithm. The signal detection unit computes the update of x, z, and λ. The x-minimization is performed keeping z, and λ constant, z-update is performed using Huber fitting keeping x and λ constant. Finally, the λ update is performed keeping x and z constant, which ensures the convergence of the algorithm by suitably choosing the step size γ. Finally, the acceleration unit increases the convergence of the algorithm by extrapolating the λ-update during each iteration and using the updated λ to update z. Finally, the SINR computation unit calculates the SINR for the detected user signal.

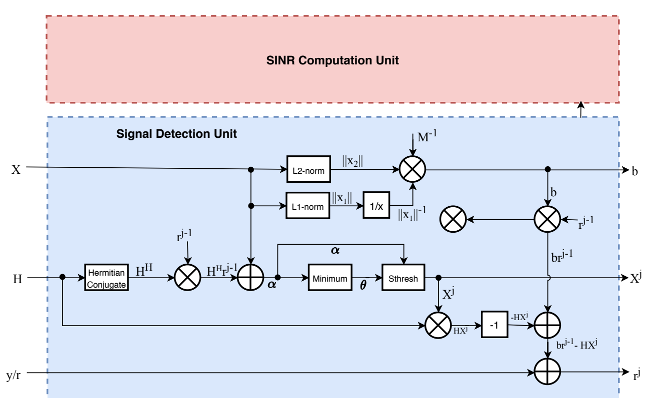


Fig. 7: Proposed hardware architecture for the MAMP uplink signal detection algorithm for Massive MIMO systems.

The proposed hardware architecture for the MAMP detection algorithm is shown in Fig. 7. Compared to the hardware architectures for the LSRS and HADMM algorithms, the proposed hardware architecture for the MAMP algorithm consists of a signal detection unit and an SINR computation unit. The architecture takes H, y, x^0 as inputs. The signal detection unit for the proposed architecture calculates α and threshold value θ by performing matrix-vector multiplication and addition. To detect the signal X, soft thresholding is performed with vector α and threshold value θ. The Proportionality factor is estimated with the ratio of l_1-norm and l_2-norm, and the estimated value is used to minimize the residual value of r^j. Finally, the residual value is minimized until the maximum iteration is reached, and the user signal is recovered as x_k.

Based on the proposed hardware architectures, we computed the computational complexity of the proposed algorithms and compared it with conventional massive MIMO detection algorithms such as ZF and MMSE. We have used big $\mathcal{O}$ notation to measure the complexity, which simply describes the performance or complexity of the algorithm, and specifically, $\mathcal{O}$ represents the worst-case scenario [31]. The complexity of ZF and MMSE is in order of $\mathcal{O}(MN^2)$ [32]. The computational complexity of LSRS detection algorithm, HADMM detection algorithm, and MAMP detection algorithms are found to be in in order of $\mathcal{O}(KN^2)$, $\mathcal{O}(KN^2)$, and $\mathcal{O}(KMN)$ respectively, where K is the number of iterations. The computational com-

TABLE II: Comparison of Computational Complexity

Algorithm	**Complexity**
ZF	$\mathcal{O}(MN^2)$
MMSE	$\mathcal{O}(MN^2)$
Lease Square Regression Selection Detector	$\mathcal{O}(KN^2)$
Huber Fitting based ADMM Detector	$\mathcal{O}(KN^2)$
AMP Detector	$\mathcal{O}(KMN)$

plexity comparison of all the detection algorithms is presented in Table II. Since the number of base station antenna is much more than the number of iterations (i.e., M $\gg$ K), we can conclude that the proposed massive MIMO uplink detection algorithms are computationally less complex than the conventional ZF and MMSE detection algorithms. Also, the complexity of all three algorithms is the same if the number of the antenna at the base station and the number of users is equal (i.e., M=N).

VI. Conclusion

In this paper, we proposed iterative algorithms for uplink signal detection in massive MIMO systems. The simulation results show that the proposed algorithms can achieve near-optimal BER performance with lower computational complexity. The simulation results also show that the BER performance of the proposed detection algorithms improves with the higher number of base station antennas and reduces with higher modulation order. We also proposed novel hardware architectures for the proposed algorithms to identify the required physical components and their interrelationship. Thus, the proposed algorithms provide an efficient trade-off between BER performance and the computational complexity and are suitable for uplink signal detection in a massive MIMO uplink system. In the future, it would be interesting to test the experiment by adding several realistic network parameters and with more number of antennas at the base station and the user terminal.

References

[1] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, *2nd ed.*, vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15-64.

[2] A closer look at Massive MIMO, accessed on Sept. 2019. [Online]. Available: https://business.sprint.com/blog/massive-mimo

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and Challenges With Very Large Arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[4] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[5] T. L. Marzetta, "Massive MIMO: An Introduction," in *Bell Labs Technical Journal*, vol. 20, pp. 11-22, 2015.

[6] X. Wu, N. C. Beaulieu, and D. Liu, "On Favorable Propagation in Massive MIMO Systems and Different Antenna Configurations," in *IEEE Access,* vol. 5, pp. 5578-5593, 2017.

[7] Green Car Congress, accessed on Sept. 2019. [Online]. Available: https://www.greencarcongress.com/2011/07/fordv2v-20110719.html

[8] A. Elghariani and M. Zoltowski, "Successive interference cancellation for large-scale MIMO OFDM," *2015 IEEE International Conference on Electro/Information Technology (EIT),* Dekalb, IL, 2015, pp. 657-661.

[9] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?, " *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[10] Y.C. Liang, G. M. Pan, and Z. D. Bai, "Asymptotic performance of MMSE receivers for large systems using random matrix theory, " *IEEE Trans. Inf. Theory,* vol. 53, no. 11, pp. 4173-4190, Nov. 2007.

[11] S. Ghacham, M. Benjillali, and Z. Guennoun, "Low-complexity detection for massive MIMO systems over correlated Rician fading," *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC),* Valencia, 2017, pp. 1677-1682.

[12] X. Gao, L. Dai, Y. Ma, and Z. Wang, "Low-complexity near-optimal signal detection for uplink large-scale MIMO systems," in *Electronics Letters, vol.* 50, no. 18, pp. 1326-1328, 28 August 2014.

[13] Y. Kim, Jae Hyun Seo, Heung Mook Kim, and S. Kim, "Soft linear MMSE detection for coded MIMO systems," *2013 19th Asia-Pacific Conference on Communications (APCC)*, Denpasar, 2013, pp. 657-660.

[14] R. Bohnke, D. Wubben, V. Kuhn, K.D. Kammeyer," Reduced complexity mmse detection for blast architectures," *2003 Global Telecommunications Conference (GLOBECOM)*, 2003, Vol.4, pp. 2258-2262.

[15] O. Gustafsson, E. Bertilsson, J. Klasson and C. Ingemarsson, "Approximate Neumann Series or Exact Matrix Inversion for Massive MIMO?" *2017 IEEE 24th Symposium on Computer Arithmetic (ARITH)*, London, 2017, pp. 62-63.

[16] Byunggi Kang, Ji-Hwan Yoon, Jongsun Park, "Low-complexity massive MIMO detectors based on Richardson method," *ETRI Journal*, vol. 39, no. 3, pp. 326-335, June 2013.

[17] X. Gao et al., "Matrix Inversion-Less Signal Detection Using SOR Method for Uplink Large-Scale MIMO Systems," *IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 8-12, 2014, pp. 3291-3295.

[18] J. Zhou, Y. Ye, and J. Hu, "Biased MMSE Soft-Output Detection Based on the Jacobi Method in Massive MIMO," *IEEE Int. Conf. Commun. Problem-Solving*, Beijing, China, Dec. 5-7, 2014, pp. 442-445.

[19] Zhizhen Wu, Chuan Zhang, Ye Xue, Shugong Xu, and Xiaohu You, "Efficient architecture for soft-output massive MIMO detection with Gauss-Seidel method," In Proc. *IEEE Circuits and Systems (ISCAS)*, pages 1886- 1889, 2016.

[20] Bei Yin, Michael Wu, Joseph R Cavallaro, and Christoph Studer, "VLSI design of large-scale soft-output mimo detection using conjugate gradients, " In Proc. *IEEE Circuits and Systems (ISCAS),* pages 1498-1501, 2015

[21] R. Chataut, R. Akl and U. K. Dey, "Least Square Regressor Selection Based Detection for Uplink 5G Massive MIMO Systems," *2019 IEEE 20th Wireless and Microwave Technology Conference (WAMICON),* Cocoa Beach, FL, USA, 2019, pp. 1-6.

[22] R. Chataut and R. Akl, "Huber Fitting based ADMM Detection for Uplink 5G Massive MIMO Systems," *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York City, NY, USA, 2018, pp. 33-37.

[23] R. Chataut and R. Akl, "Efficient and low complex uplink detection for 5G massive MIMO systems," *2018 IEEE 19th Wireless and Microwave Technology Conference (WAMICON)*, Sand Key, FL, 2018, pp. 1-6.

[24] Jinsu Kim, Sungwoo Park, Jae Hong Lee, Joonho Lee, and Hanwook Jung, "A scheduling algorithm combined with zero-forcing beamforming for a multi-user MIMO wireless system," *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference*, 2005., Dallas, TX, USA, 2005, pp. 211-215.

[25] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations, and Trends® in Machine Learning*, v.3 n.1, p.1-122, January 2011.

[26] D. Gabay B. Mercier "A dual algorithm for the solution of non-linear variational problems via finite element approximation," *Comput. Math. Appl.* vol. 2 no. 1 pp. 17-40 1976.

[27] A. Elgabli, A. Elghariani, A. O. Al-Abbasi and M. Bell, "Two-stage LASSO ADMM signal detection algorithm for large-scale MIMO," *2017 51st Asilomar Conference on Signals, Systems, and Computers,* Pacific Grove, CA, 2017, pp. 1660-1664.

[28] W. Tao, Z. Pan, G. Wu, and Q. Tao, "The Strength of Nesterov's Extrapolation in the Individual Convergence of Nonsmooth Optimization," in IEEE Transactions on Neural Networks and Learning Systems.

[29] D. L. Donoho, A. Maleki, and A. Montanari, "Message Passing Algorithms for Compressed Sensing," Proc. *Natl. Acad. Sci.* 106 (2009) 18914-18919.

[30] N. Hurley and S. Rickard, "Comparing Measures of Sparsity," In *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723-4741, Oct. 2009.

[31] A beginner's guide to Big O notation accessed on Sept. 2019. [Online]. Available: https://rob-bell.net/2009/06/a-beginners-guide-to-big-o-notation/

[32] S. Lyu and C. Ling, "Hybrid Vector Perturbation Precoding: The Blessing of Approximate Message Passing," in IEEE *Transactions on Signal Processing*, vol. 67, no. 1, pp. 178-193, 1 Jan.1, 2019.