PROCEEDINGS OF SCIENCE

Efficient and Flexible VLSI Architecture for Soft-Output Massive MIMO Detector

Qiushi Wei[1a], Leibo Liu[2b]
Department of Microelectornics and Nanoelectronics, Tsinghua University
Beijing, 100084, China
E-mail: [a]`weiqs15@mails.tsinghua.edu.cn`; [b]`liulb@tsinghua.edu.cn`

Guiqiang Peng
Department of Microelectornics and Nanoelectronics, Tsinghua University
Beijing, 100084, China
E-mail: `pgq13@mails.tsinghua.edu.cn`

Shouyi Yin
Department of Microelectornics and Nanoelectronics, Tsinghua University
Beijing, 100084, China
E-mail: `yinsy@tsinghua.edu.cn`

Shaojun Wei
Department of Microelectornics and Nanoelectronics, Tsinghua University
Beijing, 100084, China
E-mail: `wsj@mail.tsinghua.edu.cn`

The Mean-Square-Error (MMSE) detection achieves near-optimal performance in signal detection for massive Multiple-Input-Multiple-Output (MIMO) systems. But MMSE detection still suffers from high complexity of matrix inversion. In this paper, an efficient and flexible architecture is proposed based on the modified version of the Symmetric Successive Over Relaxation (SSOR) method. A Reconfigurable Computing Array (RCA) is used to implement the SSOR Method. In order to speed up the iteration, an initial solution is adopted. Approximated LLRs computational method is used to scale down the computing load of Log-Likelihood Rate computations. FPGA implementation results show a superior performance over the state-of-the-art designs.

ISCC2017
16-17 December 2017
Guangzhou, China

[1]Speaker
This work was supported by the National Natural Science Foundation of China (Grant No. 61672317)
[2]Corresponding Auther

PoS(ISCC 2017)055

1. Introduction

MIMO is a key technology in most modern wireless communication standards; however, the traditional MIMO systems can not satisfy the increasing requirements for data rates, spectral, link reliability and energy efficiency in the future wireless systems. Massive MIMO is a very promising technique for the 5G wireless communications and it has been proved that the massive MIMO provides opportunities to achieve the ever growing demand in the future wireless systems.

Bringing the amazing benefits of massive MIMO faces a few challenges, one of which is significantly increasing computational complexity by orders of magnitude in the base station. Some optimal detection methods like Maximum Likelihood (ML) [1], K-Best [2] are able to achieve high performance in data detection. Unfortunately, the problem of computational complexity is nonnegligible when the number of antennas is large. Zero-Forcing (ZF) and MMSE can achieve a tradeoff between the performance and complexity; however, the complex matrix complexity is involved in MMSE detection. Recently, Neumann series (NS) method [3], Conjugate Gradient (CG) method [4], [5], and Gauss-Seidel (GS) method [6], [7] were proposed to achieve matrix inversion indirectly, but the reduction in complexity is not obvious because of large iterative number.

In this paper, we describe a low-complexity data detection algorithm based on SSOR for massive MIMO system in the uplink. Firstly, we focus on linear soft out detection in combination with an optimized matrix inversion method based on SSOR algorithm. Then, we propose a speed-up method in the SSOR method considering the initial solution of the iteration. Finally, an approximated Log-Likelihood Rates (LLRs) computational method is proposed to scale down the complexity. The simulation results show that the proposed algorithm achieves higher detection accuracy when compared with the algorithm as recently proposed. Based on the proposed algorithm, we develop an efficient and flexible VLSI architecture for signal detection in massive MIMO systems. In particular, we propose a reconfigurable computing array (RCA). Furthermore, different antenna configurations in massive MIMO system can be achieved based on this flexible architecture of various configurations. The experimental results of our design on a Xilinx Virtex-7 FPGA show that our design performs 3.43×, 2.84×, 1.71× throughput per slice compared with the NS-based detector [3], CG-based detectors [5] and GS-based detectors [7].

2. System Model

The massive MIMO (usually $N >> M$ [6]) uplink system has N antennas at the base-station (BS) to simultaneously communicate with M single-antenna users. The parallel transmit bit streams of M users are encoded by utilizing channel encoders, and then, the results are mapped to constellation symbols to get a sequent of transmit vectors $\mathbf{s}$. Let $\mathbf{s}$ denote the $M \times 1$ transmitted signal vector of all M users, and vector $\mathbf{y}$ stand for $N \times 1$ that received signal at the BS. We have:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad , \tag{2.1}$$

where $\mathbf{H} \in \mathbb{C}^{N \times M}$ stands for flat Rayleigh fading channel matrix whose elements are independent and identically distributed (i.i.d.) and follow $\mathcal{N}(0,1)$, and all elements of $\mathbf{n}$ denote $N \times 1$ i.i.d. zero-mean complex additive white Gaussian noise (AWGN) whose power spectral density are $\mathbb{E}(\mathbf{n}^H\mathbf{n}) = N_0\mathbf{I}$. Furthermore, we assume the power of transmitted vector is $\mathbb{E}(\mathbf{s}^H\mathbf{s}) = E_s\mathbf{I}$.

According to **H** and **y**, the base station detector can compute soft-estimates in the form of LLRs for **s**. The estimation of **s** in MMSE which is the most common can be computed as:

$$\hat{\mathbf{s}} = \left(\mathbf{H}^H\mathbf{H} + N_0 E_s^{-1}\mathbf{I}_U\right)^{-1}\mathbf{H}^H\mathbf{y} = \mathbf{A}^{-1}\mathbf{y}^{MF} \quad , \tag{2.2}$$

where $\mathbf{A}=\mathbf{G}+N_0E_s^{-1}\mathbf{I}_U$ denotes the MMSE filtering matrix, and $\mathbf{y}^{MF}=\mathbf{H}^H\mathbf{y}$, $\mathbf{G}=\mathbf{H}^H\mathbf{H}$ are the matched-filter vector and the Gram matrix. Let $\mathbf{U}=\mathbf{A}^{-1}\mathbf{G}$ denotes the equivalent channel gain and $\mathbf{v}=\mathbf{A}^{-1}\mathbf{H}^H\mathbf{n}$ denotes the equivalent noise, combing (2.1) and (2.2), the MMSE estimation of the transmitted vector can be rewritten as $\hat{\mathbf{s}}=\mathbf{U}\mathbf{s}+\mathbf{v}$. As the matrix **U** is a non-diagonal matrix, the estimation of transmitted symbol by MMSE for the ith users $\hat{s}_i$ not only contains the information of s_i but also includes the interferences of another s_j $(j\neq i)$. In order to distinguish the useful information and interference, each element of $\hat{\mathbf{s}}$ can be written as:

$$\hat{s}_i = U_{i,i}s_i + \sum_{j=1,\, j\neq i}^{M} U_{i,j}s_j + v_i = U_{i,i}s_i + \eta_i\ , \tag{2.3}$$

where U_{ij} presents the elements of matrix **U** in the ith and jth column, U_{ii} is the effective channel gain and $\eta_i=\sum_{j=1,\, j\neq i}^{M} U_{i,j}s_j+v_i$ denotes the post-equalization Noise-Plus-Interference (NPI) variance including interferences and noise. It is obviously that si and η_i are independent when the streams are independent. Hence, the expectation of η_i is 0. Let σ_{eq}^2 denote the variance of η_i and b be the bit index of the LLR of ith user. Then the max-log LLR can be expressed as:

$$L_{i,b}(\hat{s}_i) = \varsigma_i^2\left(\min_{s\in S_b^0}\left|\frac{\hat{s}_i}{\mu_i}-s\right|^2 - \min_{s\in S_b^1}\left|\frac{\hat{s}_i}{\mu_i}-s\right|^2\right), \tag{2.4}$$

where $\varsigma_i^2 = U_{ii}^2/\sigma_{eq}^2$ is the signal-to-noise-plus-noise ratio (SINR) for ith user, S_b^0 and S_b^1 denote to the sets of modulation constellation symbols, where the ith bit is 0 and 1, respectively. So the soft-output of MMSE detector could be exported. Unfortunately, the MMSE detection consists of the inversion of **A**, which results in an excessive complexity of $\mathcal{O}(M^3)$. It is obviously that the detector needs a larger number of multiplications when M is large. Hence, the practical solutions for uplink massive MIMO detection demand low complexity for matrix inversion.

3. SSOR-Based Signal Detection Method for Massive MIMO Systems

In this section, firstly, an optimized SSOR iterative method is used to achieve matrix inversion. Then, we propose a speed-up method of the SSOR method by using properties of massive MIMO channel. Finally, an approximated Log-Likelihood Rates (LLRs) computational method is proposed to scale down complexity.

3.1. Proposed SSOR-based Signal Detection Method

3.1.1. Optimized SSOR Method

In the massive MIMO systems, the matrix **H** is asymptotically orthogonal; hence, the matrix **G** and matrix **A** are Hermitian positive definite [6]. The SSOR-based signal detection method is used to solve the linear equation, as shown in (2.2). According to the SSOR iteration, we decompose the matrix **A** into three parts: $\mathbf{A}=\mathbf{L}+\mathbf{D}+\mathbf{L}^H$, in which **D**, **L** and $\mathbf{L}^H$ denote the diagonal component, the strictly lower triangular component, and the strictly upper triangular component of **A** respectively. Here the SSOR-based can be expressed as:

$$(\mathbf{D}+\omega\mathbf{L})\hat{\mathbf{s}}^{(k+\frac{1}{2})}=[(1-\omega)\mathbf{D}-\omega\mathbf{L}^H]\hat{\mathbf{s}}^{(k)}+\omega\mathbf{y}^{MF}$$
$$(\mathbf{D}+\omega\mathbf{L}^H)\hat{\mathbf{s}}^{(k+1)}=[(1-\omega)\mathbf{D}-\omega\mathbf{L}^H]\hat{\mathbf{s}}^{(k+\frac{1}{2})}+\omega\mathbf{y}^{MF} \quad , \tag{3.1}$$

where k = 0,1, ... is the number of iterations, $\hat{\mathbf{s}}^{(0)}$ is the initial solution (discussed later in the paper), ω is the relaxation parameter. To realize the iteration method in hardware efficiently,we change the computing rule considering the definition of **D** and **L**, the iteration can be presented as:

$$\hat{s}_i^{(k+\frac{1}{2})}=\frac{\omega}{D_{i,i}}(y_i^{MF}-\sum_{j<i}L_{i,j}\hat{s}_j^{(k+\frac{1}{2})}+\frac{1-\omega}{\omega}D_{i,i}\hat{s}_i^{(k)}-\sum_{j>i}L_{i,j}^T\hat{s}_j^{(k)})$$
$$\hat{s}_i^{(k+1)}=\frac{\omega}{D_{i,i}}(y_i^{MF}-\sum_{j<i}L_{i,j}\hat{s}_j^{(k+1)}+\frac{1-\omega}{\omega}D_{i,i}\hat{s}_i^{(k+\frac{1}{2})}-\sum_{j>i}L_{i,j}^T\hat{s}_j^{(k+\frac{1}{2})}) \quad , \tag{3.2}$$

where $D_{i,j}$ and $L_{i,j}$ denote the ith row and jth column of matrix **D** and matrix **L**, and $\hat{s}_i^{(k+1)}$, $\hat{s}_i^{(k+1/2)}$, $\hat{s}_i^{(k)}$, and y_i^{MF} denotes the ith element of $\hat{\mathbf{s}}^{(k+1)}$, $\hat{\mathbf{s}}^{(k+1/2)}$, $\hat{\mathbf{s}}^{(k)}$, and and $\mathbf{y}^{MF}$, respectively. According to (3.1.1.2), the optimized SSOR method takes full advantage of information of the whole matrix **A**. Also, all the computations (vector multiplications) are similar, indicating that the method can be implemented in hardware easily with high efficiency and flexibility.

3.1.2. Speed-up

In order to improve the speed of iteration, we consider the initial solution. If the initial solution is nearby the exact final solution, the iteration number could be small. Hence, the next task is to determine the initial solution, the traditional set as a zero-vector. For massive MIMO systems, the Gram matrix **G** and matrix **A** are diagonally dominant, indicating that we have:

$$\mathbf{h}_i^H\frac{\mathbf{h}_j}{N}\rightarrow 0,\quad i\neq j,\quad i,j=1,2,\cdots,M \quad , \tag{3.3}$$

where $\mathbf{h}_i$ denotes the ith column vector of the matrix **H**. The domination of the diagonal elements of matrix **A** is more and more obvious when the number of N/M is increasing. By analyzing the special properties of massive MIMO systems, a low complexity initial solution is proposed as:

$$\hat{\mathbf{s}}^{(0)}=\frac{1}{N}\mathbf{H}^H\mathbf{y} \quad . \tag{3.4}$$

The proposed initial method can speed-up the iteration obviously. The complexity of the initial matrix is very low, so the computation can be calculated in parallel.

3.1.3. Approximated LLRs Computational Method

The equivalent channel matrix can be expressed as $\mathbf{U} = \mathbf{A}^{-1}\mathbf{G}$, the NPI variance σ_{eq}^2 can be computed like [3]:

$$\sigma_{eq}^2 = E_s U_{ii} - E_s U_{ii}^2 \quad , \tag{3.5}$$

which indicates that we need to compute the matrix **U** before the NPI and SINR computations. According to (2.4) and (3.1.3.1), the SINR can be computed as:

$$\varsigma_i^2=\frac{U_{ii}^2}{\sigma_{eq}^2}=\frac{1}{E_s\left(1/U_{ii}-1\right)} \, . \tag{3.6}$$

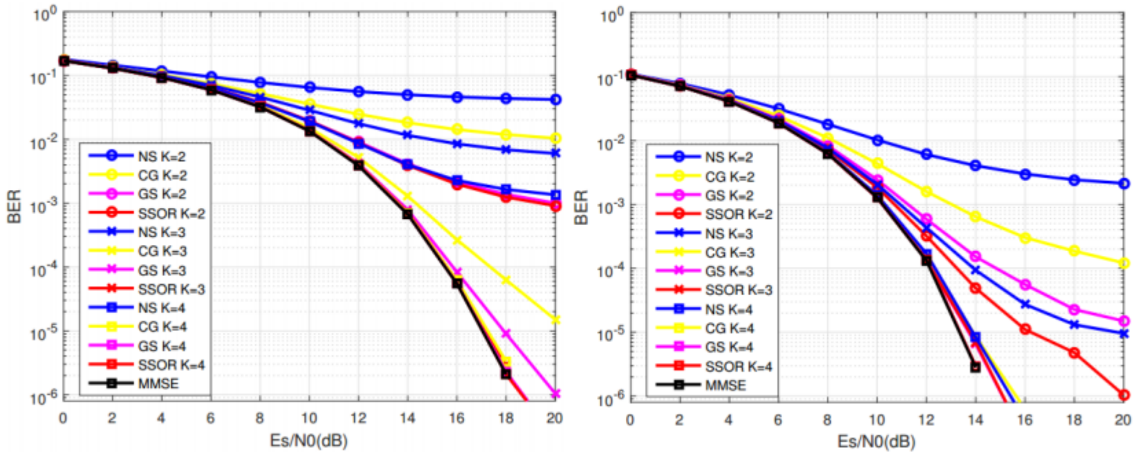
However, according to the definition of matrix **U**, we need to compute the inversion of matrix **A**, which suffers from the complexity of $\mathcal{O}(M^3)$. Since matrix **A** is diagonal dominant in massive MIMO systems. Inspired by (3.1.2.1) and (3.1.2.2), we can approximate the channel gain matrix **U**, as:

$$\hat{\mathbf{U}} = \hat{\mathbf{A}}^{-1}\mathbf{G} = \frac{1}{N}\mathbf{G} \quad . \tag{3.7}$$

Combining (2.4), (3.1.3.2) and (3.1.3.3), the soft-output LLRs of the detector can be approximately computed. The approximated computation of the LLRs avoids the large-scale matrix multiplications and inversions, which can reduce the complexity.

3.2. Simulation Results

To evaluate the performance of the proposed SSOR-based algorithm, we simulate the BER performance when compared with the NS, CG and GS methods, as shown in Figure 1. Nothing that the exact MMSE algorithm with Cholesky decomposition [7] is also provided in this figure to be the reference of these approximate methods. The Rayleigh fading channel model is provided. These simulation results show the proposed SSOR-based method can achieve much more near-optimal performance in different antenna configurations when compared with other algorithms. Hence, to achieve the same performance, the SSOR-based algorithm needs smaller iterative number. For example, in Figure 1-(b), the simulation result of K=2 in SSOR-based algorithm has almost the same BER performance of K=3 in NS-based algorithm.



(a) N=64, M=8 (b) N=128, M=8

Figure 1: BER Performance Comparison between SSOR and Other Methods.

4. Reconfigurable VLSI Architecture

We propose a low complexity VLSI architecture based on the proposed SSOR detection method. The overall architecture consists of a Reconfigurable Computing Array (RCA), which is shown in Figure 2. The array can be reconfigured in order to compute the Gram matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H}$, matched-filter $\mathbf{y}^{MF}=\mathbf{H}^H\mathbf{y}$, initial and iteration computations of vector and LLR computation.

4.1. Reconfigurable Computing Array

In the RCA, a Finite State-Machine (FSM) controls the date memory, configurable memory, interconnection and the RCA. The main blocks of the RCA can be reconfigured according to the SSOR method. In order to achieve high parallelism, there are 16 reconfigurable computing unit (RCU) in the RCA. The input data can be stored into the data memory and the configuration can be stored into the configure memory. The RCA can achieve the whole SSOR algorithm. In addition, for different antenna configurations in massive MIMO systems, the RCA can be reconfigured to achieve the signal detection.

4.2. Reconfigurable Computing Unit

The reconfigurable computing unit includes four real-real multipliers, four adders, two accumulations, one subtracter, and three multiplexers, as shown in Figure 3. Each RCU supports one complex multiplication accumulation, one complex multiplication or two conjugate complex multiplication accumulations. For different steps of the SSOR algorithm, the RCU can be reconfigured in order to achieve different functions. According to different configurations, the outputs are selected from three kinds of values, including the real-real multiplication results, the accumulated results and the addition results with parameters. In order to support the SSOR method, each RCU supports the following elementary operations: matrix-matrix multiplication, matrix-vector multiplication, initial solution and iteration and LLR computation.

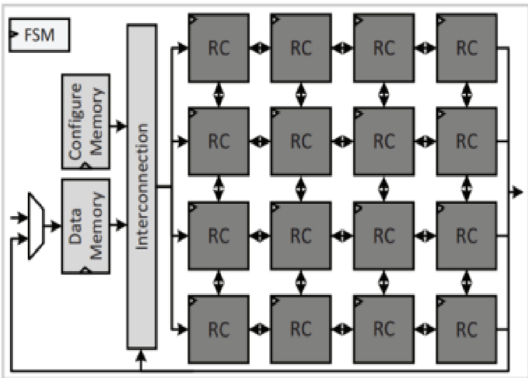


Figure 2: Reconfigurable Computing Array for SSOR-based Soft-output Signal Detection.

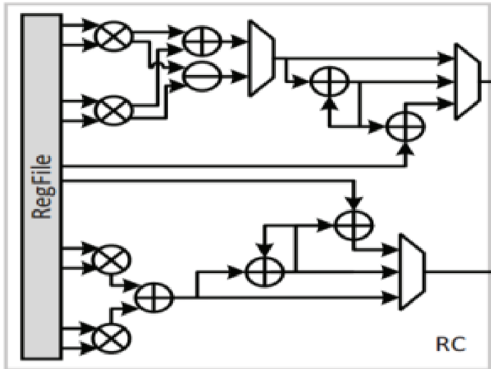


Figure 3: Reconfigurable Computing Unit for SSOR-based Soft-output Signal Detection.

For the Gram matrix computation, all the RCUs are reconfigured to achieve the matrix multiplication. In each RCU, there are three main steps. Firstly, the real-real multiplications are achieved. Secondly, the results are transmitted to accumulations. Finally, in order to get Matrix **A**, the outputs of each accumulation are added with a parameter $N_0E_s^{-1}$. The matched-filter also can be computed based on the RCU. The matched-filter computation is a matrix-vector multiplication, and the accumulation results are exported. In the initial solution and iteration computations, each element of the vector $\hat{\mathbf{s}}$ is computed in each RCU. The times of iterations can be controlled by the FSM according to the configurable memory. The results are exported from accumulations. Considering the LLR computation, the results are related to the SINR and a piecewise linear function for Gray mappings. In each RCU, only two real-real multipliers and some other lookup tables (LUTs) can be used to achieve LLR computation.

5. Experimentation Results and Conclusion

We implemented our SSOR-based massive MIMO detector for a 128×8 system on a Xilinx Virtex-7 XC7VX690T FPGA to achieve a fair comparison with NS detector [3], CG detector [5] and GS detector [7]. Table 1 compares the FPGA implementation results of the proposed SSOR-based detector with other detectors. From Table 1, the SSOR-based implementation has a lower throughput but consumes much smaller hardware resources. Hence, the ratio of throughput/slices in the proposed detector is 3.43× [3], and 2.84× [7]. In addition, compared with CG-based detector [5], this detector can achieve a better throughput/slices (1.71×).

Inversion method	NS	GS	CG	SSOR
LUT slices	168125	18976	3324	3292
FF slices	193451	15864	3878	3456
DSP48	1059	232	33	36
Frequency [MHz]	317	309	412	314
Throughput [Mb/s]	603	48	20	32
Thoughput/slices [Mbps/K slices]	**1.67**	**1.38**	**2.78**	**4.74**

Table 1: FPGA Implementation Results

In this paper, we proposed an efficient and flexible VLSI architecture for SSOR-based soft-output massive MIMO detection. An initial solution method is proposed to speed up the iteration and a low complexity LLR computation method is proposed. It has been demonstrated that the proposed VLSI architecture is suitable for different SSOR iterations and antenna configurations. The FPGA implementation results show advantages on throughput per slice. Future work will focus on the development of the reconfigurable coarse-grain hardware architecture to achieve high area and energy efficiencies in uplink massive MIMO systems.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "*What will 5G be*" IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] M. O. Damen, H. El Gamal, and G. Caire, "*On maximum-likelihood detection and the search for the closest lattice point,*" IEEE Trans. Inf. Theory., vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[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] D. Auras, R. Leupers, and G. H. Ascheid, "*A novel reduced-complexity soft-input soft-output MMSE MIMO detector: Algorithm and efficient VLSI architecture*," in Proc. IEEE Int. Conf. Commun. (ICC), Sydney, Australia, Jun. 2014, pp. 4722–4728.

[5] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "*VLSI design of largescale soft-output MIMO detection using conjugate gradients*," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2015, pp. 1498–1501.

[6] L. Dai, X. Gao, X. Su, S. Han, Z. Wang, "*Low-Complexity Soft-Output Signal Detection Based on Gauss-Seidel Method for Uplink Multi-User Large-Scale MIMO Systems,*" IEEE Trans. Veh. Technol., vol. 64, no. 10, pp. 4839–4845, Oct. 2015.

[7] Z. Wu, C. Zhang, Y. Xue, S. Xu, and X. You, "*Efficient architecture for soft-output massive MIMO detection with Gauss-Seidel method*," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2016, pp. 1886–1889.