

# Uplink Multi-user MIMO Detection via Parallel Access

Keke Zu

Research and Development  
KK Innovation AB  
Stockholm, Sweden  
zukeke@gmail.com

Jun Zhu

Dept. of Elec. & Comp. Engr.  
University of Victoria  
Victoria, Canada  
zhujun@ece.ubc.ca

Martin Haardt

Communications Research Laboratory  
Ilmenau University of Technology  
Ilmenau, Germany  
martin.haardt@tu-ilmenau.de

**Abstract**—In this paper, we develop simultaneous detection techniques of signals from multiple users for uplink multi-user MIMO (UL MU-MIMO) systems. Conventional detectors do not take the detection delay into account. Two parallelizing access methods are proposed for UL MU-MIMO systems. The multiple uplink users can be detected in parallel after the parallelizing process. Moreover, the multiple uplink users can be scheduled on the same radio frequency resource as if all the other users did not exist. Therefore, the proposed detection methods can scale up with the system dimensions by keeping the bit error rate (BER) and the detection delay at an acceptable level. Simulation results show that the proposed detection methods via parallel access achieve considerable BER gains with much less detection delay as compared to their conventional counterparts.

**Index Terms**—Multi-user MIMO (MU-MIMO), uplink transmission, parallelizing process

## I. INTRODUCTION

A limiting factor for capacity of a radio network is the amount of available radio resources, e.g., in terms of time and frequency bandwidth, and the capacity of a radio network can be improved by more efficient usage of such radio resources. When multiuser MIMO (MU-MIMO) is used, the utilization of radio resources can be improved considerably. For the MU-MIMO uplink, the receiving node such as the *base station* (BS) or the *access point* (AP) receives information from several users simultaneously on the same time-frequency resources. This requires the receiving node to decode all uplink users jointly from the received signal, i.e., *multi-user detection* (MUD) [1]. To achieve a proper and timely detection of all signals in a MU-MIMO scenario is thus a challenge.

### A. Prior Art

Linear receivers, particularly *zero forcing* (ZF) receivers, are largely applied in practice because of their low-complexity [2]. Such a receiver employs a ZF filter to eliminate the interference between the users' signals, at the expense of amplifying the noise power, which results in a degraded system performance. This becomes more severe when the number of antennas and data streams increases. Thus, the conventional ZF receiver cannot provide a good performance when a large number of users is active in uplink MU-MIMO systems.

To enhance the performance for MUD, *successive interference cancellation* (SIC) technologies can be used [3]. The

core ideas of SIC based receivers is layer peeling which can be efficiently implemented by utilizing an QR decomposition [4]. When a detected symbol is assumed to be correct, the SIC receiver successively subtracts the interference contributed by the previous streams from the current stream. This process proceeds to the next symbol until all symbols are detected. One drawback of this technique is that to decode the  $k$ -th user, the streams from all the previous  $k - 1$  users must have been correctly decoded to perform the present decoding. Any decoding error would propagate user by user if one or more previous streams are decoded incorrectly. The successive MMSE based SIC (MMSE-SIC) detection has been proposed in [5] where users are equipped with an arbitrary number of antennas.

For detection techniques based on successive decoding, not only the users but also multiple streams of each user are detected successively. This would result in a considerable detection delay when the number of active users is high and each user transmits multiple streams.

In [6], uplink block diagonalization (UL-BD) based detection techniques have been proposed to decompose the received multi-user signal into independent single-user signals. Then, the users can be detected in parallel by detection techniques like ZF, SIC, etc. The UL-BD is an application of BD on the uplink. If the number of receive antennas is very large, this scheme has a fairly high computational complexity due to the singular value decomposition (SVD) of BD type algorithms as pointed out in [7].

### B. Contributions

In this work, two parallelizing access methods for UL MU-MIMO systems are developed. The multi-user interference is completely removed or efficiently suppressed by the proposed parallelizing process. Then, each user is detected individually in parallel as if the other users did not exist. Therefore, the proposed detection methods can scale up with the system dimensions both in terms of the bit error rate (BER) and the detection delay. It is particularly useful for systems with a large number of active users scheduled in the same resource block where each user can transmit multiple streams. The main contributions of the work can be summarized as follows

- 1) Two parallelizing access methods for UL MU-MIMO systems are proposed, and the utilization of radio resources can be improved by allowing multiple uplink users to access the same radio resource simultaneously.
- 2) The proposed parallelizing methods bring extra flexibility to UL MU-MIMO detection. The received signals from multiple uplink users are decoupled. Then, the uplink users can be detected individually. Therefore, a customized detection strategy can be implemented per user.

## II. SYSTEM MODEL AND DETECTION ALGORITHMS

Assume that there are  $K$  users each equipped with  $N_k$  antennas, and the receiving node is equipped with  $N_R$  antennas. For the  $k$ -th user, its uplink flat fading channel is denoted by  $\mathbf{H}_k \in \mathbb{C}^{N_R \times r_k}$ , where  $r_k$  is the number of transmitted data streams from the  $k$ -th user where  $\mathbb{C}$  refers to the complex field, and the total number of transmitted data streams is  $N_S = \sum_{k=1}^K r_k$ . The received uplink signal  $\mathbf{y}$  is

$$\mathbf{y} = \mathbf{H}_1 \mathbf{s}_1 + \cdots + \mathbf{H}_k \mathbf{s}_k + \cdots + \mathbf{H}_K \mathbf{s}_K + \mathbf{n}, \quad (1)$$

where the quantity  $\mathbf{s}_k \in \mathbb{C}^{r_k \times 1}$  is the  $k$ -th transmitted data vector, and  $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$  is the noise term. We take  $\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_k, \cdots, \mathbf{H}_K] \in \mathbb{C}^{N_R \times N_S}$  as the combined channel matrix, and  $\mathbf{s} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_K^T]^T \in \mathbb{C}^{N_S \times 1}$  as the combined transmit data vector. Thus, the received uplink signal  $\mathbf{y}$  can be rewritten as

$$\mathbf{y} = \mathbf{H} \mathbf{s} + \mathbf{n}. \quad (2)$$

### A. Conventional Detection Algorithms

The ZF detection filter is designed to completely eliminate the inter-user interference, that is

$$\mathbf{G}_{\text{ZF}} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H. \quad (3)$$

As shown in (3), the ZF detection filter at the receiver side is actually the left pseudo-inverse of the channel matrix  $\mathbf{H}$ . Therefore, the ZF detector is also called *channel inversion*. The joint estimation of  $\mathbf{s}$  can be obtained by

$$\hat{\mathbf{s}} = \mathbf{G}_{\text{ZF}} \mathbf{y} = \mathbf{s} + \mathbf{G}_{\text{ZF}} \mathbf{n}. \quad (4)$$

For Gaussian noise with independent identically distributed (i.i.d.) entries of zero mean and variance  $\sigma_n^2$ , the error covariance matrix is

$$\Phi_{\text{ZF}} = E\{(\hat{\mathbf{s}} - \mathbf{s})(\hat{\mathbf{s}} - \mathbf{s})^H\} = \sigma_n^2 \mathbf{G}_{\text{ZF}} \mathbf{G}_{\text{ZF}}^H. \quad (5)$$

From (5), we can see that the noise power could be significantly increased by  $\mathbf{G}_{\text{ZF}}$  especially when the channel is ill conditioned. Thus, the BER performance of the ZF detector will be greatly degraded. This phenomenon is called *noise enhancement*.

In order to reduce the effects of *noise enhancement* caused by the ZF filter  $\mathbf{G}_{\text{ZF}}$ , the *minimum mean square error* (MMSE) criterion takes the noise term into account and is designed according to

$$\mathbf{G}_{\text{MMSE}} = \arg \min_{\mathbf{G}} E\{\|\mathbf{G} \mathbf{y} - \mathbf{s}\|^2\}. \quad (6)$$

By utilizing the orthogonality principle  $E\{\|\mathbf{G} \mathbf{y} - \mathbf{s}\| \mathbf{y}^H\} = \mathbf{0}$ , the MMSE detection filter  $\mathbf{G}_{\text{MMSE}}$  is derived as

$$\mathbf{G}_{\text{MMSE}} = (\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_{N_S})^{-1} \mathbf{H}^H, \quad (7)$$

where the quantity  $\alpha$  is the regularization factor and  $\mathbf{I}_{N_S}$  is an  $N_S \times N_S$  identity matrix. As shown in [8], MMSE detection is equivalent to ZF with respect to an extended system model. The extended channel matrix  $\underline{\mathbf{H}}$  and the extended received signal  $\underline{\mathbf{y}}$  are given by

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \sqrt{\alpha} \mathbf{I}_{N_S} \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{N_S \times 1} \end{bmatrix}, \quad (8)$$

Using these definitions, the MMSE detection filter can have a similar format as the ZF filter, i.e.,

$$\mathbf{G}_{\text{MMSE}} = (\underline{\mathbf{H}}^H \underline{\mathbf{H}})^{-1} \underline{\mathbf{H}}^H. \quad (9)$$

From (9), the MMSE receiver is also called *regularized channel inversion*. Since the multi-user interference mitigation is balanced with *noise enhancement*, a better performance than ZF can be obtained by performing the MMSE detection.

SIC detection provides a good trade-off between the BER performance and the computational complexity. The *zero forcing QR decomposition based detection* (ZF-QRD) scheme can achieve a high spectral efficiency with a reasonable decoding complexity. The procedure of ZF-QRD starts with a QR decomposition, which calculates  $\mathbf{H} = \mathbf{Q} \mathbf{R}$  where the quantity  $\mathbf{Q}$  is a unitary matrix and  $\mathbf{R}$  is an upper triangular matrix. The effective received signal  $\mathbf{y}_{\text{eff}}$  is obtained by multiplying the Hermitian transpose of the unitary matrix  $\mathbf{Q}$  to the received signal  $\mathbf{y}$  as

$$\mathbf{y}_{\text{eff}} = \mathbf{Q}^H \mathbf{y} = \mathbf{R} \mathbf{s} + \tilde{\mathbf{n}}, \quad (10)$$

where the quantity  $\tilde{\mathbf{n}}$  is the effective noise term. By assuming that the detected symbol is correct, the ZF-QRD receiver successively subtracts the interference contributed by previous streams from the current stream. This process proceeds to the next symbol until all symbols are detected. Starting the detection process from the bottom, the ZF-QRD procedure is implemented successively as

$$\hat{s}_{N_S} = Q \left( \frac{y_{(\text{eff}, N_S)}}{r_{(N_S, N_S)}} \right),$$

$$\hat{s}_i = Q \left( \frac{y_{(\text{eff}, i)} - \sum_{j=i+1}^{N_S} r_{(i, j)} \hat{s}_j}{r_{(i, i)}} \right), \quad (11)$$

where the quantity  $r_{(i, i)}$  is the  $i$ -th diagonal element of the matrix  $\mathbf{R}$ , and  $Q(\cdot)$  denotes the slicing function.

## III. PROPOSED PARALLEL DETECTION ALGORITHMS

For uplink MU-MIMO, the conventional channel inversion based linear detection methods can efficiently null multi-user and multi-stream interference, but may result in a performance loss when the number of users increases. Furthermore, the received streams are detected sequentially in SIC-type receivers as described in (11). When the number of users is large, it may cause *error propagation* and also a considerable detection

delay. Therefore, there is a need to develop a method to detect the uplink users scheduled on the same radio resources separately.

#### A. Decoupling Processing for Uplink Users

The system architecture of the proposed parallelizing access is illustrated in Fig. 1. At the receiver side, an estimate of the channel matrix is usually available. With the channel information known at the receiver side, two parallelizing access methods are proposed in this work. The signals from multiple users can be separated from each other and individual users can be detected in parallel as if the other users did not exist.

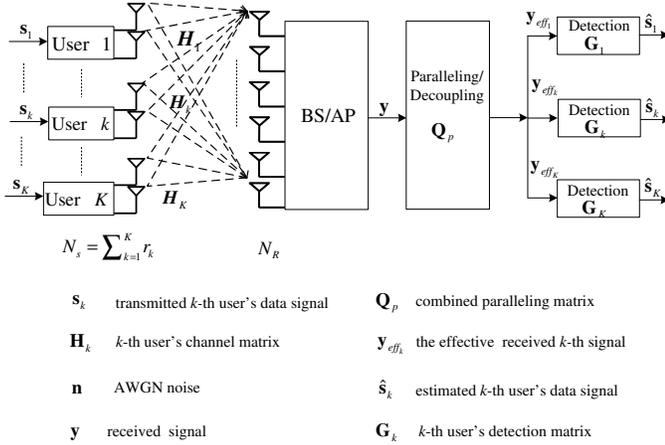


Fig. 1. The model of uplink parallelizing access for MU-MIMO systems.

1) *ZF Decoupling*: We calculate the pseudo-inverse of the combined channel matrix  $\mathbf{H}$  as

$$\mathbf{H}^+ = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H = \begin{bmatrix} \mathbf{H}_1^+ \\ \vdots \\ \mathbf{H}_k^+ \\ \vdots \\ \mathbf{H}_K^+ \end{bmatrix}, \quad (12)$$

where the quantity  $\mathbf{H}_k^+$  is the  $k$ -th sub-matrix of  $\mathbf{H}^+$ . Given that  $\mathbf{H}^+ \mathbf{H} = \mathbf{I}_{N_s}$ , we have,

$$\mathbf{H}_k^+ \mathbf{H}_k = \mathbf{I}_{r_k}, \quad \mathbf{H}_k^+ \mathbf{H}_j = \mathbf{0} \quad \forall j \neq k. \quad (13)$$

Next, we perform the LQ decomposition of  $\mathbf{H}_k^+$  as

$$\mathbf{H}_k^+ = \mathbf{L}_k \mathbf{Q}_k, \quad (14)$$

where the quantity  $\mathbf{L}_k \in \mathbb{C}^{r_k \times r_k}$  is a lower triangular matrix and  $\mathbf{Q}_k \in \mathbb{C}^{r_k \times N_R}$  has unitary rows. Substituting (14) into (13), since  $\mathbf{L}_k$  is invertible, we have

$$\mathbf{Q}_k \mathbf{H}_j = \mathbf{0} \quad \forall j \neq k. \quad (15)$$

From (15), it can be seen that the decoupling matrix  $\mathbf{Q}_k$  forms an orthogonal basis for the common left null space of all

channels  $\mathbf{H}_j, j \neq k$ . The inter-user interferences are cancelled from the  $k$ -th user by  $\mathbf{Q}_k$ . Thus, the effective received  $k$ -th user's signal is

$$\mathbf{y}_{(\text{eff},k)} = \mathbf{Q}_k \mathbf{y} = \mathbf{H}_{(\text{eff},k)} \mathbf{s}_k + \mathbf{n}_{(\text{eff},k)}, \quad (16)$$

where the effective channel matrix for the  $k$ -th user is  $\mathbf{H}_{(\text{eff},k)} = \mathbf{Q}_k \mathbf{H}_k$ , and the effective noise term is  $\mathbf{n}_{(\text{eff},k)} = \mathbf{Q}_k \mathbf{n}$ .

Finally, the combined ZF decoupling matrix  $\tilde{\mathbf{Q}}$  is obtained as

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_k \\ \vdots \\ \mathbf{Q}_K \end{bmatrix}. \quad (17)$$

Moreover, the combined effective channel matrix  $\mathbf{H}_{\text{eff}}$  is

$$\mathbf{H}_{\text{eff}} = \tilde{\mathbf{Q}} \mathbf{H} = \begin{bmatrix} \mathbf{H}_{(\text{eff},1)} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_{(\text{eff},k)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{H}_{(\text{eff},K)} \end{bmatrix}. \quad (18)$$

As shown in (15) and (18), the multiple uplink users are separated from each other by the decoupling matrix  $\tilde{\mathbf{Q}}$ . We call this method *ZF decoupling (ZF)* process.

2) *MMSE Decoupling*: The *ZF decoupling* process only focuses on canceling inter-user interference and neglects the impact of the noise. Thus, a regularized factor is introduced to achieve a tradeoff between inter-user interference and noise. The regularized inversion of the combined channel matrix is

$$\mathbf{H}_{\text{mmse}}^+ = (\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_{N_s})^{-1} \mathbf{H}^H = \begin{bmatrix} \mathbf{H}_{(\text{mmse},1)}^+ \\ \vdots \\ \mathbf{H}_{(\text{mmse},k)}^+ \\ \vdots \\ \mathbf{H}_{(\text{mmse},K)}^+ \end{bmatrix}, \quad (19)$$

where the parameter  $\alpha$  is the regularization factor and the quantity  $\mathbf{H}_{(\text{mmse},k)}^+ \in \mathbb{C}^{r_k \times N_R}$  is the  $k$ -th sub-matrix of  $\mathbf{H}_{\text{mmse}}^+$ . Similar to the ZF decoupling, we have the following relationship

$$\mathbf{H}_{(\text{mmse},k)}^+ \mathbf{H}_k \approx \mathbf{I}_{r_k}, \quad \mathbf{H}_{(\text{mmse},k)}^+ \mathbf{H}_j \approx \mathbf{0} \quad \forall j \neq k, \quad (20)$$

We perform the LQ decomposition of  $\mathbf{H}_{(\text{mmse},k)}^+$  as

$$\mathbf{H}_{(\text{mmse},k)}^+ = \mathbf{L}_{(\text{mmse},k)} \mathbf{Q}_{(\text{mmse},k)}, \quad (21)$$

where the quantity  $\mathbf{L}_{(\text{mmse},k)} \in \mathbb{C}^{r_k \times r_k}$  is a lower triangular matrix and  $\mathbf{Q}_{(\text{mmse},k)} \in \mathbb{C}^{r_k \times N_R}$  has unitary rows. Since  $\mathbf{L}_{(\text{mmse},k)}$  is invertible, we have

$$\mathbf{Q}_{(\text{mmse},k)} \mathbf{H}_j \approx \mathbf{0} \quad \forall j \neq k. \quad (22)$$

Thus, the unitary matrix  $\mathbf{Q}_{(\text{mmse},k)}$  can be used as the *MMSE decoupling* (MD) matrix. The  $k$ -th user's effective received signal is

$$\mathbf{y}_{(\text{eff},k)} = \mathbf{Q}_{(\text{mmse},k)}\mathbf{y} = \mathbf{H}_{(\text{eff},k)}\mathbf{s}_k + \mathbf{n}_{(\text{eff},k)}, \quad (23)$$

where the effective channel is given by  $\mathbf{H}_{(\text{eff},k)} = \mathbf{Q}_{(\text{mmse},k)}\mathbf{H}_k$  and the effective noise is given by  $\mathbf{n}_{(\text{eff},k)} = \mathbf{Q}_{(\text{mmse},k)}\mathbf{n}$ . Finally, the combined MD matrix  $\tilde{\mathbf{Q}}_{\text{mmse}}$  is obtained as

$$\tilde{\mathbf{Q}}_{\text{mmse}} = \begin{bmatrix} \mathbf{Q}_{(\text{mmse},1)} \\ \vdots \\ \mathbf{Q}_{(\text{mmse},k)} \\ \vdots \\ \mathbf{Q}_{(\text{mmse},K)} \end{bmatrix}. \quad (24)$$

After the *ZF/MMSE decoupling* process, the multiple uplink users can be detected individually and separately in parallel. For individual users, the detection delay is largely reduced and more flexibility is introduced by allowing different detection methods for different users, e.g., dependent on their specific channel quality, QoS requirement, and data rate. It can be noted that the implementation of the proposed procedures does not require any feedback nor feedforward information from or to the multiple users.

#### IV. SIMULATION RESULTS

For the UL MU-MIMO system, we assume that a BS or an AP is equipped with  $N_R$  receiving antennas and  $K$  active users are scheduled on the same radio resource block, each user transmitting  $r_k = N_k$  data streams. This configuration is termed as  $(N_R, K, r_k)$ . We also assume that equal power loading between users and streams is performed.

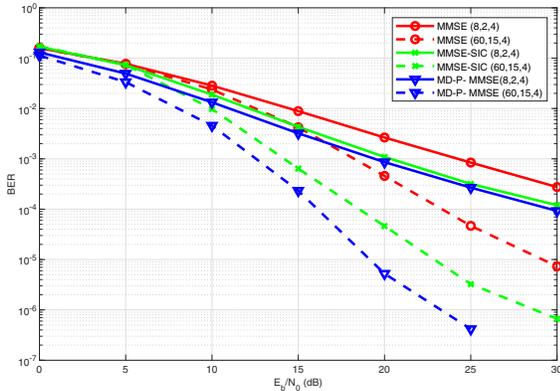


Fig. 2. BER performance comparison of the proposed MD-P-MMSE with the conventional MMSE and MMSE-QRD for the system configuration (8, 4) and (60, 15, 4) respectively.

For the system configuration (60, 15, 4) shown in Fig. 2, the conventional MMSE-QRD detection can achieve a fairly good performance, but the delay among users is a serious problem

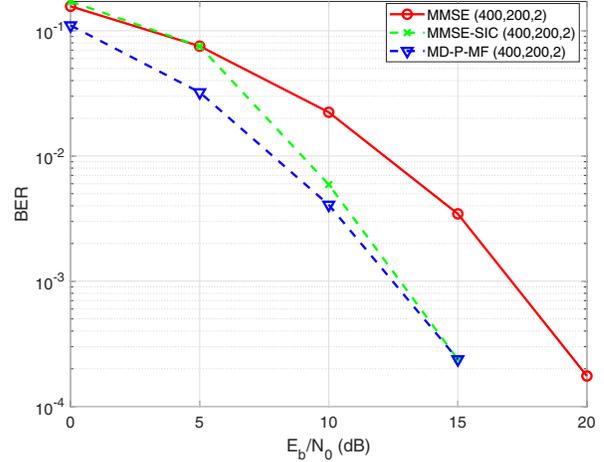


Fig. 3. BER performance comparison of the proposed MD-P-MMSE with MMSE and MMSE-QRD for the system configuration (400, 200, 2).

since all the 60 received streams are detected sequentially. From the delay perspective, therefore, the conventional SIC detectors cannot scale up properly with the system dimensions. The proposed *MMSE decoupling with parallel MMSE detection* (MD-P-MMSE) can achieve 5.2 dB and 1.5 dB gains over the conventional MMSE and MMSE-QRD at a BER of  $10^{-3}$ , respectively. Thus, it can be concluded that the proposed MD-P-MMSE detection can scale up with the system dimensions both in terms of performance and delay.

As pointed out in [9], [10], a simple receiver like maximum ratio combining can achieve the SINR-maximizing performance in the limit of a large number of antennas. However, this observation is only valid for scenarios with a single transmit stream per user. For the system configuration (400, 200, 2) shown in Fig.3, we apply *matched filtering* (MF) detection to each individual user after the proposed *MMSE decoupling* process. The BER performance of the proposed *MMSE decoupling with parallel MF* (MD-P-MMSE) is even better than that of the MMSE-QRD in the medium-low SNR region (more common SNR in practice) with much less detection delay.

#### V. CONCLUSIONS

The proposed UL MU-MIMO *ZF/MMSE decoupling* process provides a new perspective for implementing multi-user detection algorithms. The signals from multiple users can be separated from each other as if the other users did not exist and detected in parallel. The utilization of radio resources can be improved, the detection delay can be reduced, and extra flexibility is introduced. Simulations show that the proposed detection methods can achieve considerable BER gains with much less delay as compared to the conventional detection methods.

## REFERENCES

- [1] D. Tse, and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [2] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge University Press, 2003.
- [3] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. Valenzuela, "V-BLAST: architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI ISSSE*, pp. 295-300, Sep. 1998.
- [4] D. Wübben, J. Rinas, R. Böhnke, V. Kühn, and K. D. Kammeyer, "Efficient algorithm for detecting layered space-time codes," in *Proc. ITG Conf. on Source and Channel Coding (SCC)*, Berlin, Germany, Jan. 2002.
- [5] V. Stankovic, and M. Haardt, "Improved diversity on the uplink of multi-user MIMO systems," in *Proc. European Conference on Wireless Technology (ECWT 2005)*, Paris, France, pp. 113 - 116, Oct. 2005.
- [6] L. Arevalo, R. C. de Lamare, M. Haardt, and R. Sampaio-Neto, "Uplink block diagonalization for massive MIMO-OFDM systems with distributed antennas," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, Dec. 2015.
- [7] K. Zu, R. C. de Lamare and M. Haardt, "Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser MIMO systems", *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4232-4242, Oct. 2013.
- [8] D. Wübben, R. Böhnke, V. Kühn, and K. D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, pp. 508-512, Oct. 2003.
- [9] 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 Proces. Mag.*, vol. 30, no. 1, pp. 40-46, Jan. 2013.
- [10] 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.