IMPLEMENTATION OF THE PRECODER MATRIX INDICATOR SELECTION USING MMSE TRACE CRITERION FOR THE DOWNLINK TRANSMISSION IN LTE

Chang-Han Chung, Tsung-Hsien Liu, and Yuan-Sun Chu

National Chung Cheng University
168 University Rd., Min-Shiung, Chia-Yi, Taiwan

ABSTRACT

Based on the minimum mean squared error (MMSE) trace criterion, selection of the precoding matrix indicator (PMI) from the downlink codebook of the LTE system is considered in this paper. Due to the codebook of 16 member precoding matrices and up to 1200 sub-carriers in this MIMO-OFDM system, the PMI selection needs to compute matrix inversion of up to 19,200 matrices. To satisfy this required workload, LDL^H decomposition is applied at the algorithmic level, and, pipeline matrix multiplication and backward substitution modules are used at the architectural level. The VLSI implementation results of our architecture under the TSMC 90 nm CMOS technology reveals that our architecture requires 124.6K gates at operating frequency 120 MHz. Thus, our designed architecture can finish the 19,200 matrix inversions in about 1.60 ms and meets the requirement of 2 ms period of periodic channel state report.

Index Terms— multiple input multiple output (MIMO) system, precoding matrix, LTE, PMI report

1. INTRODUCTION

Precoding is one of the practical and effective approaches for the multiple input multiple output orthogonal frequency division multiplexing (MIMO-OFDM) system to fight against fading channels [1,2]. The precoded MIMO-OFDM communication system provides better error rate performance than the space-time coded or spatial division multiplexing MIMO-OFDM system [3, Chap. 13]. However, the major disadvantage of the precoded MIMO-OFDM system is that it requires the channel state information (CSI) estimated at the receiver side to be feedback to the transmitter side. To mitigate the time-varying channel effect, the time difference between the transmission of training pilots and the transmission of precoded data has to be short. One solution to achieve this goal is to reduce the amount of CSI information feedback to the transmitter side.

Two different ways have been adopted by the IEEE 802.11n/ac WLAN [3] and LTE [4–8] to reduce the amount of CSI information. In IEEE WLAN, the precoding matrix associated with each sub-carrier is determined from the singular value decomposition of the channel matrix [9]. Angles representing the precoding matrix of each sub-carrier are quantized [10] and sent back to the transmitter side. This approach is referred to as the compressed beamforming weights feedback. The transmitter constructs each precoding matrix based on the received quantized angles and starts the transmission of the precoded data.

In LTE, a codebook of finite precoding matrices is known to both the transmitter and receiver sides. The receiver estimates the CSI of each sub-carrier based on the pilots sent by the transmitter. The receiver then determines the rank indicator (RI) and precoding matrix indicator (PMI) from the codebook. Only the indices are sent back to the transmitter side through the channel state report. Once the transmitter receives the indices, it finds the corresponding precoding matrix from the codebook and starts the transmission of the precoded data. One major difference here from that in IEEE WLAN is that the values of RI and PMI are determined from the CSI of 12 to 1200 sub-carriers [4].

Practical criteria for determining RI and PMI in LTE include [11–15] minimum mean square error (MMSE) geometric mean, MMSE trace, maximum post processing SNR, post MMSE, maximum channel capacity, and maximum mutual information. Note that most of these criteria are related to the MMSE detector and rely on the computation of the error covariance matrix. The computation tricks applied to one selection criterion can therefore be applied to another selection criterion. Among these selection criteria, the MMSE trace criterion leads to the best bit error rate (BER) performance and complexity trade-off [11, 13]. Additionally, for the downlink of the LTE transmission, the report of both RI and PMI is requested aperiodically; the receiver needs to yield the report in 4 ms [7]. The PMI is reported periodically by the receiver with period 2 ms [4]. Thus, in this paper, we consider only the PMI selection based on the MMSE trace criterion.

Several fast algorithms or low-cost hardware architectures have been reported on this PMI selection problem in LTE. The nested property of the precoding matrices based on the Householder matrix was proposed to reduce the complexity of the selection algorithm [16–18]. Tree search algorithm to utilize the correlation between the precoding matrices was reported [8]. Cholesky factorization, selected FFT, and nested cofactor expansion schemes were used in [7] to reduce both the algorithm and hardware architecture.

To our knowledge, no hardware architecture has been reported for the periodic PMI report over 1200 sub-carriers in the downlink of LTE. As mentioned earlier, MMSE trace is the criterion for determining the PMI. The main difficulty of implementing this PMI selection is to compute the error covariance matrices of up to 1200 sub-carriers for each of the precoding matrix in the codebook. In totality, 19,200 matrix inversions have to be finished in 2 ms. Since this problem is related to the MMSE detection, we will utilize the tricks in [19, 20] to propose a hardware architecture to achieve this goal.

2. REPORT OF PRECODER MATRIX INDICATOR FOR THE DOWNLINK TRANSMISSION

The LTE downlink transmission [4, 6] is an example of the precoded MIMO-OFDM system illustrated in Fig. 1. At the network side (eN-

This paper was supported in part by the Ministry of Science and Technology, Taiwan, under Contract MOST 104-2221-E-194-018.

odeB), a stream of digitally modulated symbols is passed through a spatial stream parser to produce N_s streams of symbols. A precoder with $N_t \times N_s$ precoding matrix $\mathbf{P}_k$ for the k-th sub-carrier transforms the N_s spatial streams of data to produce N_t, $N_t \geq N_s$, spatial streams of data. The precoding matrix $\mathbf{P}_k$ is a member of the finite-sized downlink codebook $\mathcal{F}$. The resulting streams of data are then inverse discrete Fourier transformed (DFT) to produce baseband signals for transmission through the N_t transmit antennas. Assume that the channel is slowly fading, indicating that channel does not change over the duration that the precoded data are being transmitted. Also, assume that there are N_r, $N_r \geq N_s$, receive antennas at the receiver side (terminal). At a particular time instant (OFDM symbol), the received equivalent baseband $N_r \times 1$ signal at sub-carrier k can be expressed as

$$\mathbf{x}_k = \mathbf{H}_k \mathbf{P}_k \, \mathbf{s}_k + \mathbf{z}_k, \tag{1}$$

where $N_r \times N_t$ $\mathbf{H}_k$ is the equivalent channel matrix at sub-carrier k, $N_s \times 1$ $\mathbf{s}_k$ is the transmitted vector at sub-carrier k, and $N_r \times 1$ $\mathbf{z}_k$ is the additive noise with covariance matrix $\sigma^2\mathbf{I}$ at sub-carrier k.

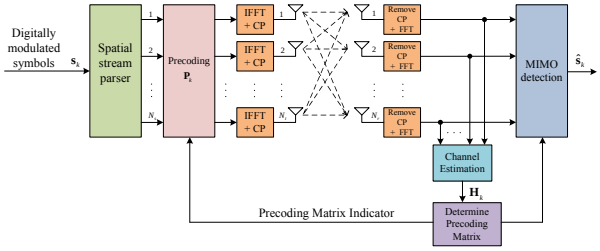

Fig. 1. A precoded MIMO-OFDM system with precoding weights feedback.

Assume that N_s and $\mathbf{H}_k$, $\forall k$, are known at the receiver side. The LMMSE detector [19] applies the filtering weight $\mathbf{G}_{\text{MMSE}} = \mathbf{H}_k\mathbf{P}_k \left(\mathbf{P}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{P}_k + \sigma^2\mathbf{I}\right)^{-1}$ to the received $\mathbf{x}_k$ and produces output $\mathbf{G}_{\text{MMSE}}^H\mathbf{x}_k$ with error covariance matrix equal to $\left(\mathbf{P}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{P}_k + \sigma^2\mathbf{I}\right)^{-1}$. Note that this error depends on the precoding matrix $\mathbf{P}_k$. To reduce this error, the receiver has to determine a $\mathbf{P}_k$ from the downlink codebook $\mathcal{F}$, i.e., $\mathbf{P}_k \in \mathcal{F}$, and send its index back to the transmitter side. Furthermore, there are multiple sub-carriers used for transmission in the considered LTE MIMO-OFDM system. To feedback the index of PMI for each sub-carrier may occupy large resource, especially when the number of sub-carriers is large. Also, the channel matrices and associated precoding matrices for sub-carriers located at neighboring frequencies are approximately the same. Thus, in the LTE design, a same precoding matrix is used for transmitting signal at multiple sub-carriers. For this reason, a good precoding matrix is determined by

$$\mathbf{W}_{\text{opt}} = \arg \min_{\mathbf{W}_i \in \mathcal{F}} \sum_{k=1}^{K} \text{trace} \left\{ \left(\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I} \right)^{-1} \right\}. \tag{2}$$

The $\mathbf{W}_i$, $i = 1, \cdots, 16$, denote all the 16 member precoding matrices in the codebook $\mathcal{F}$. The trace$\{\cdot\}$ operation is to sum all the errors associated all N_s streams. The summation over K sub-carriers indicates that the same precoding is applied to K sub-carriers, where K may range from 12 to 1200. The criterion used in (2) for determining the precoding matrix $\mathbf{W}_{\text{opt}}$ is referred to as the minimum MMSE trace criterion [11]. To report the index of $\mathbf{W}_{\text{opt}}$ in $\mathcal{F}$ is called the report of PMI. In LTE, the terminal has to periodically report PMI for every 2 ms.

In summary, given $\mathbf{H}_k$, $k = 1, \cdots, K$, and, $\mathbf{W}_i$, $i = 1, \cdots, 16$, our following algorithm and architecture will be designed to output the $\mathbf{W}_{\text{opt}}$ or equivalent PMI in 2 ms.

3. PROPOSED ALGORITHM

Our fast algorithm to compute the $\mathbf{W}_{\text{opt}}$ in (2) is related to our fast algorithm [20] for determining the MMSE detector with interference cancellation. We therefore apply the computation tricks in [20] to obtain the solution in (2). The complete steps of our algorithm are listed below. The inputs to the algorithm include $\mathbf{H}_k$, $k = 1, \cdots, K$, σ^2, and $\mathbf{W}_i$, $i = 1, \cdots, 16$. The algorithm generates $\mathbf{W}_{\text{opt}}$ as output, the index of which is to be feedback to the transmitter.

1) Repeat the following steps 2-9 for $i = 1$ to 16.
2) Repeat the following steps 3-8 for $k = 1$ to K.
3) Compute $\mathbf{H}_k\mathbf{W}_i$ through direct multiplications.
4) Compute the upper triangular part of $\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I}$ through direct multiplications.
5) Perform LDL^H decomposition of $\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I}$ to obtain $\mathbf{R}_i$ and $\mathbf{D}_i$, where $\mathbf{R}_i$ and $\mathbf{D}_i$ satisfy $\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I} = \mathbf{R}_i^H\mathbf{D}_i\mathbf{R}_i$.
6) Compute the inversed matrices $\mathbf{R}_i^{-1}$ and $\mathbf{D}_i^{-1}$.
7) Compute the diagonal entries of $\mathbf{R}_i^{-1}\mathbf{D}_i^{-1}(\mathbf{R}_i^{-1})^H$ through direct multiplications.
8) Sum the diagonal entries in step 7 to produce $\eta_{i,k}$.
9) Calculate the MMSE $\eta_i = \sum_{k=1}^{K} \eta_{i,k}$ associated with $\mathbf{W}_i$.
10) Find the minimum of $\eta_1, \cdots, \eta_{16}$, the index of which is determined to be the PMI. The associated $\mathbf{W}_i$ is the determined $\mathbf{W}_{\text{opt}}$.

The tricks used by this algorithm include: (i) Only the entries in the upper triangular part of $\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I}$ are computed, because the matrix is conjugate symmetric; (ii) LDL^H decomposition is applied to a conjugate symmetric matrix; (iii) The complicated matrix inversion is applied only to the triangular matrix $\mathbf{R}_i$; and, (iv) Only the diagonal entries of $\left(\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i + \sigma^2\mathbf{I}\right)^{-1}$ are computed.

4. PROPOSED ARCHITECTURE

Based on the algorithm given at the last paragraph of the preceding section, we propose the VLSI architecture in Fig. 2 for computing the PMI to be feedback. This architecture is for receive antennas $N_r = 4$, transmit antennas $N_t = 4$, and spatial streams $N_s = 4$. This architecture includes mainly three blocks, the complex matrix multiplication #1, the complex matrix multiplication #2, and matrix inversion. The last part of this architecture is a set of 16 parallel accumulators to accumulate the K errors revealed by step 9 of the proposed algorithm.

The architecture of the complex matrix multiplication #1 block in Fig. 2 is a systolic array [21] and is illustrated in Fig. 3. This block is for the computations in step 3 of our algorithm. Entries of complex channel matrices $\mathbf{H}_k$ and $\mathbf{W}_i$ are serial inputs to this block. Because of the parallel and pipeline structure of this systolic array, different matrix pairs of $\mathbf{H}_k$ and $\mathbf{W}_i$ can be inputs to this block. To compute

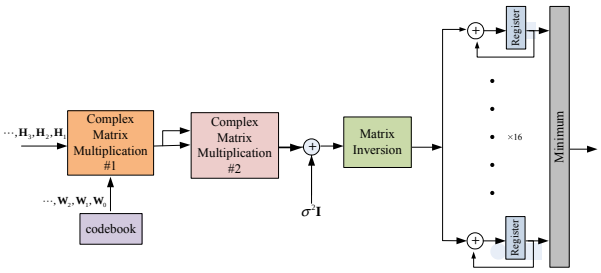

Fig. 2. The proposed hardware architecture.

the product of two complex numbers, a newly designed complex multiplication with accumulation (CMULACC), illustrated in Fig. 4, is used as a processing element in the systolic array. The left half of the CMULACC architecture includes 3 real multipliers and 5 adders to form a complex multiplier, whereas the right half includes two accumulators to accumulate a complex number. After every 4 clock cycles, the two registers in each CMULACC store a complex number for each entry of $\mathbf{H}_k\mathbf{W}_i$. A total of 16 CMULACCs are used to compute the 16 entries of $\mathbf{H}_k\mathbf{W}_i$.

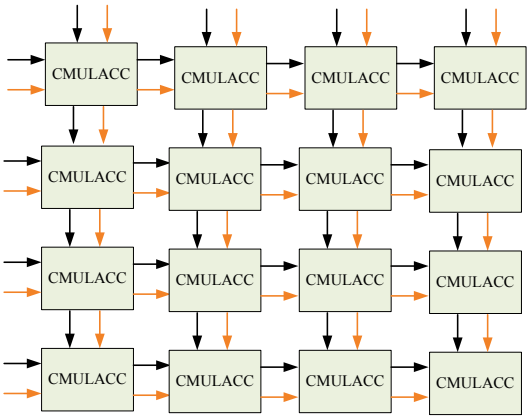

Fig. 3. The block of complex matrix multiplication #1 in Fig. 2.

A systolic array is used again for the complex matrix multiplication #2 block in Fig. 2 and is illustrated in Fig. 5. This systolic array is for computing the conjugate symmetric matrix $\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i$ as shown by step 4 of the algorithm. Because the symmetry of the computed matrix, this triangular systolic array only includes modules of CMULACCs at the left-bottom half of the processing elements.

The block of matrix inversion in Fig. 2 is further illustrated in Fig. 6. This architecture is for the computations in steps 5-9 of our algorithm. According to [20], the LDL^H decomposition can be expressed as a sequence of backward substitutions. We use the block in [20, Fig. 3], which is now illustrated in Fig. 7, iteratively here for LDL^H decomposition. This block consists of three cascaded processing elements (PEs). Each PE includes 3 real-valued multipliers and 5 adders to perform a complex-valued multiplication. Two additional real-valued multipliers in each PE are required for performing backward substitution. The two real-valued two-stage pipelined dividers on the right part of Fig. 7 are associated with the divisions by the entries of $\mathbf{D}_i$. This block requires 10 clock cycles to output

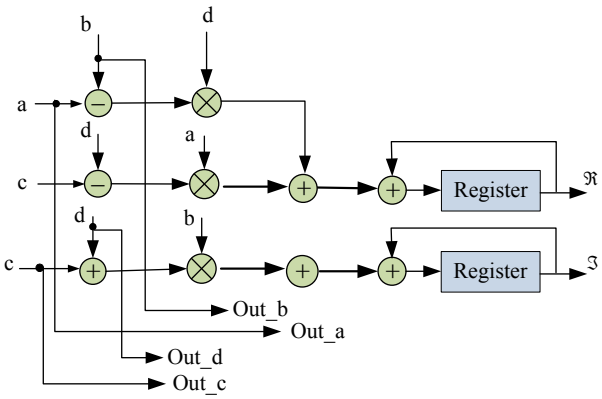

Fig. 4. The complex multiplication and accumulation (CMULACC) in Fig. 3.

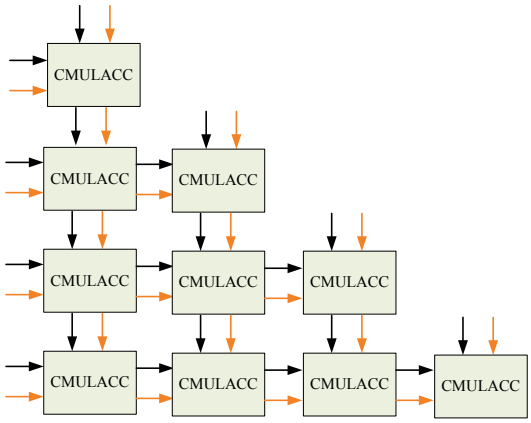

Fig. 5. The block of complex matrix multiplication #2 in Fig. 2.

every pair of $\mathbf{R}_i$ and $\mathbf{D}_i$.

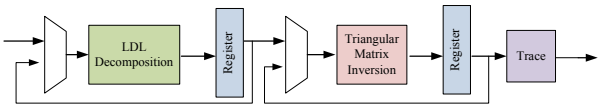

Fig. 6. The block of matrix inversion in Fig. 2.

The block of triangular matrix inversion in Fig. 6 is for the matrix inversion $\mathbf{R}^{-1}$ and $\mathbf{D}^{-1}$ in step 6 of our algorithm. Because this matrix inversion can also be expressed as a sequence of backward substitutions [20], this block is exactly the same as the one in Fig. 7, except that the multiplexer at the very right-hand-side is removed. The block of trace in Fig. 6 is associated with the steps 7 and 8 of our algorithm and is now illustrated in Fig. 8. It accumulate the squares of the entries of $\mathbf{R}_i^{-1}$, the results of which are multiplied by the entries of $\mathbf{D}_i^{-1}$. The resulting 4 outputs are summed to produce $\eta_{i,k}$. Note that $\mathbf{R}_i^{-1}$ has 1's on its diagonal. The $(4,4)$-th entry of $\mathbf{R}_i^{-1}\mathbf{D}_i^{-1}(\mathbf{R}_i^{-1})^H$ is equal to the $(4,4)$-th entry of $\mathbf{D}_i^{-1}$. Computation of the $(3,3)$-th entry of $\mathbf{R}_i^{-1}\mathbf{D}_i^{-1}(\mathbf{R}_i^{-1})^H$ does not involve accumulation. The block of Fig. 8 take advantage of these properties

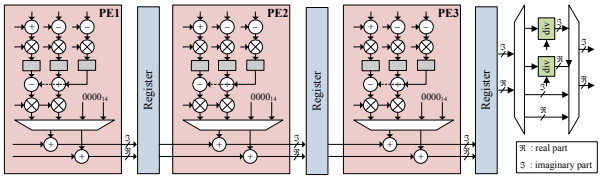

Fig. 7. The block of LDL^H decomposition in Fig. 6.

to reduce complexity.

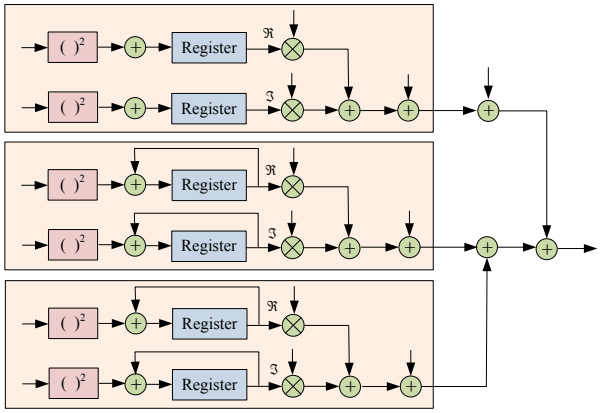

Fig. 8. The block of trace in Fig. 6.

4.1. Implementation Results

The architecture in Fig. 2 was described in Verilog HDL and synthesized by the Synopsys Design Compiler, the results of which are given in Table 1. The pipeline structure of our architecture can compute the MMSE trace, trace $\left\{\left(\mathbf{W}_i^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{W}_i+\sigma^2\mathbf{I}\right)^{-1}\right\}$, for each values of i and k in every 10 clock cycles. Thus, to compute the solution of (2) with the maximum $K = 1200$, our architecture requires 1200*16*10/120 MHz=1.60 ms, which is smaller than the 2 ms period for periodic channel state report. For comparison, the architecture of [7] can compute both PMI and RI of only 110 subcarriers in 3.3 ms. Note that such comparison may be unfair for the architecture in [7], because the architecture is designed for computing both PMI and RI of 4-by-4 and 8-by-8 channel matrices. We cannot find other related architecture beyond the one in [7]. Nevertheless, the use of systolic arrays and pipeline backward substitution block allows us to design a hardware to achieve the goal.

Table 1. Comparisons of hardware architectures.

Architecture	[7]	This work
Feedback mode	Aperiodic and periodic	periodic
Selection Criterion	Mutual Information	MMSE trace
Matrix size	4×4, 8×8	4×4
Technology	90 nm	90 nm
Gate count	547.6K	124.6K
Frequency	125 MHz	120 MHz
Power	58 mW	14.78 mW

5. CONCLUSIONS

A hardware architecture for computing the PMI to be feedback by the terminal to the eNodeB in the downlink of LTE has been proposed. Because of the MMSE trace criterion, the difficulty of this design is to compute the metrics associated with 1200 channel matrices and 16 precoding matrices in a short time of 2 ms. Our architecture to compute the 19,200 matrix inversions has been demonstrated to achieve this goal. Thus, our architecture can be used for periodic channel report by the terminal of the LTE downlink transmission.

6. ACKNOWLEDGMENT

The authors greatly appreciate the National Chip Implementation Center of Taiwan for providing technical support over the duration of this work.

7. REFERENCES

[1] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory,* vol. 51, no. 8, pp. 2967-2976, Aug. 2005.

[2] D. Love, R. Heath Jr., V. Lau, D. Gesbert, and B. Rao, "An overview of limited feedback in wireless communications systems", *IEEE J. Selected Areas Commun.,* vol. 26, no. 8, pp. 1341-1365, Aug. 2008.

[3] E. Perahia and R. Stacey, *Next Generation Wireless LANs 802.11n and 802.11ac.* 2nd Ed., Cambridge University Press, 2013.

[4] E. Dahlman, S. Parkvall, and J. Skold, *4G LTE/LTE-Advanced for Mobile Broadband.* Academic Press, 2011.

[5] S. Sesia, I. Toufik, and M. Baker, *LTE-The UMTS Long Term Evolution.* Wiley and Sons Ltd, 2011.

[6] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-Advanced," *EURASIP Journal Wireless Commun. Networking,* 2009.

[7] Y.-H. Lin, Y.-H. Chen, C.-Y. Chu, C.-Z. Zhan, and A.-Y. Wu, "Dual-mode low-complexity codebook searching algorithm and architecture for LTE/LTE-advanced systems," *IEEE Trans. Signal Proc.,* vol. 61, no. 14, pp. 3545-3562, July 2013.

[8] T.-H. Liu, C.-N. Chiu, and C.-Y. Lin, "Fast codeword selection for limited feedback beamforming MIMO systems using breadth-first tree search," *IET-Commun.,* vol. 7, no. 6, pp. 531-537, June 2013.

[9] Y.-C. Lin, T.-H. Liu, C.-P. Chou, and Y.-S. Chu, "Implementation of the SVD-based precoding sub-system for the compressed beamforming weights feedback in IEEE 802.11n/ac WLAN," in Proc. *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP),* Brisbane, Australia, 2015, pp. 1081-1085.

[10] C.-H. Tseng, T.-H. Liu, Y.-C. Lin, and Y.-S. Chu, "Hardware implementation of the joint precoding sub-system and MIMO detection preprocessor in IEEE 802.11n/ac WLAN," *in preparation.*

[11] Z. Bai, C. Spiegel, G. H. Bruck, P. Jung, M. Horvat, J. Berkmann, C. Drewes, and B. Gunzelmann, "Closed loop transmission with precoding selection in LTE/LTE-advanced system," in Proc. *Int. Symp. Applied Sciences in Biomedical and Communication Technologies (ISABEL),* Nov. 2010, pp. 1-5.

[12] Z. Bai et al., "On the physical layer performance with rank indicator selection in LTE/LTE-advanced system," in Proc. *IEEE Personal, Indoor and Mobile Radio Communications Workshops*, Sep. 2010, pp. 393-398.

[13] A. Wagdy, T. Khattab, and E. A. Sourour, "Modified QR-D and MMSE PMI selection technique for MIMO closed loop spatial multiplexing in LTE/LTE-advanced," in Proc. *IEEE GCC Conf. and Exhibition,* Nov. 2013, pp. 93-97.

[14] S. Schwarz , C. Mehlfuhrer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE" in Proc. *IEEE Conf. Wireless Adv.* 2010, pp. 1-6.

[15] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Process.,* vol. 54, pp. 1653-1665, 2006.

[16] B. Varadarajan, E. Onggosanusi, A. Dabak, and R. Chen, "Nested codebook design for MIMO precoders, in Proc. *IEEE Asilomar Conf. Signals, Systems and Comp. (ACSSC),* Oct. 2008, pp. 723-727.

[17] G. Berardinelli, T. B. Srensen, P. Morgensen, K. Pajukoski, "SVD-based vs. release 8 codebooks for single user MIMO LTE-A uplink," in Proc. *IEEE Vehicular Technol. Conf. (VTC),* May 2010, pp. 1-5.

[18] J. Wang, M. Wu, and F. Zheng, "The codebook design for MIMO precoding systems in LTE and LTE-A," in Proc. *Int. Conf. Wireless Communications Networking and Mobile Computing (WiCOM),* Sep. 2010, pp. 1-4.

[19] T.-H. Liu, "Some results for the fast MMSE-SIC detection in spatially multiplexed MIMO systems," *IEEE Trans. Wireless Commun.,* vol. 8, no. 11, pp. 5443-5448, Nov. 2009.

[20] T.-H. Liu, J.-Y. Jiang, and Y.-S. Chu, "A low-cost MMSE-SIC detector for the MIMO system: algorithm and hardware implementation," *IEEE Trans. Circuits Syst. II: Express Brief,* vol. 58, no. 1, pp. 56V61, Jan. 2011.

[21] H. Elhuni, A. Vergis. and L. Kinney, "C-testability of two-dimensional iterative arrays," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* vol. 5, no. 4, pp. 573-581, Oct. 1986.