APSIPA ASC 2011 Xi'an

# Cross Entropy Algorithm for Joint Precoding and Transmit Antenna Selection in Multiuser MIMO Systems with Limited Feedback

Min-Han Chiu[1], Yi-Syun Yang[2], Jung-Chieh Chen[3], and Chih-Peng Li[4]
[1,4] National Sun Yat-Sen University, Kaohsiung
E-mail: ridm352@hotmail.com, cpli@faculty.nsysu.edu.tw Tel: +886-7-5254475
[2] National Taiwan University, Taipei
E-mail: d97942018@ntu.edu.tw Tel: +886-2-33663075
[3] National Kaohsiung Normal University, Kaohsiung
E-mail: jcchen@nknucc.nknu.edu.tw Tel: +886-7-7172930

*Abstract*—**This paper considers the joint precoding and transmit antenna selection to reduce the hardware cost such as the radio-frequency (RF) chains associated with antennas in the downlink of multi-user multi-input multi-output (MU-MIMO) systems with limited feedback. The joint precoding and transmit antenna selection algorithm requires an exhaustive search (ES) of all possible combinations and permutations to find the optimum solution at transmitter, thus resulting in extremely high computational complexity. In order to reduce the computational load while still maximizing channel capacity, the cross entropy (CE) method is addressed to determine sub-optimum solution. Compared with the conventional genetic algorithm (GA) and random search method, simulation results show that the CE method provides not only a better performance but also a lower computational complexity.**

## I. INTRODUCTION

To cope with the growing demand for transmission rate, multi-input multi-output (MIMO) has been regarded as an essential technique because of its higher capacity over conventional single-input single-output (SISO) systems [1]. However, some practical issues need to be addressed, such as the interference induced by multiple data stream, which may result in performance degradation [2], and the cost and complexity of multiple radio-frequency (RF) chains, including converters and amplifiers.

To deal with the interference caused by multiple data stream, the linear multi-user MIMO (MU-MIMO) precoding technique has been proposed in [2] and [3]. In [3], a block-diagonalization method is presented, which is capable of canceling the multiple data stream interference under specific antenna-number constraints. However, this approach requires perfect channel state information (CSI) in the transmitter, which is not practical because it takes enormous feedback bits. As a result, the concept of a codebook, discussed in [4] and [5], is adopted in the current research. The receiver first selects the optimal precoder from the codebook, which is a predefined set of precoding vectors and is known in both transmitter and receiver sides, and then feedbacks the index of the precoding vector with a limited number of bits to the transmitter.

As regards the number of RF chain cost, a straightforward and effective solution is given by selecting a part of the available antennas to transmit instead of simultaneously employing all the antennas [6]. This approach not only reduces the hardware cost but also maintains the full spatial diversity of the original systems. However, the degradation of array gain is inevitable [7]. For this reason, the authors in [8] propose a joint selection to promote the system performance further because this degradation of array gain can be compensated by precoding. Nevertheless, due to the enormous combination of joint selections, the exhaustive search (ES) of this problem is hard to implement. Thus, the authors in [8] provide a genetic algorithm (GA) as an alternative, which is a well-known metaheuristic algorithm.

In the current research, we investigate this problem by adopting the cross entropy (CE) method, which was first proposed by Rubinstein in 1997, to solve the rare event probability estimation [9]. Afterward, this powerful method is successfully extended to solve the combinatorial optimization problems, e.g., NP-hard problems. In contrast to the existing metaheuristic optimization methods, such as GA or particle swarm optimization (PSO), the main concept of the CE method is to invoke a probability distribution function of possible solutions, instead of directly operating on the population samples, and attempt to find an optimal sampling distribution that leads to an optimal solution with a probability of one. Finally, we compare the performance of the CE method with the conventional GA and random search methods by computer simulation. The result shows that the proposed CE method provides not only better performance but also enjoys a lower computational complexity.

The rest of this paper is organized as follows. Section II describes the system model. Section III briefly illustrates the CE method. Simulation results and discussions are provided in Section IV. Finally, Section V presents some concluding remarks.

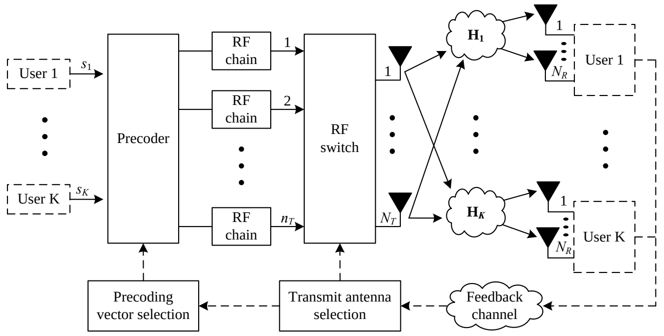


Figure 1. Transmitter structure of proposed joint scheme.

## II. SYSTEM MODEL

In the current research, we consider a downlink MU-MIMO system with limited feedback. As shown in Fig. 1, the base station (BS) is equipped with $N_T$ transmit antennas and transmits over flat Rayleigh fading channels with $K$ active users, where every user has $N_R$ receive antennas. The BS has $n_T$ RF chains, which are demultiplexed into $N_T$ transmit antennas, where $n_T \leqq N_T$. The $L$ bit codebook, which includes $2^L$ precoding vectors, is designed and known at both the transmitter and receiver sides in advance. Thus, the $N_R \times 1$ received signal for the $k$th user can be denoted as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{C}\mathbf{s} + \mathbf{n}_k, \quad k = 1, 2, ..., K \tag{1}$$

where $\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_K \end{bmatrix}^T$ is the information vector, in which each element is an independent binary phase-shift keying (BPSK) modulation symbol of the $k$th user, and $\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_K \end{bmatrix}$ is the $N_T \times K$ precoding matrix, in which each column is a precoding vector for the $k$th user designed according to the Grassmannian line-packing criterion [10]. $\mathbf{H}_k$ is the $N_R \times N_T$ channel matrix from the BS to user $k$, and $\mathbf{n}_k$ denotes the $N_R \times 1$ noise vector, where each element is an independent and identically distributed (i.i.d.) complex Gaussian random variable with zero mean and a variance $N_0$. In addition, the total transmit energy is assumed $\sum_{k=1}^{K} p_k = p_{total}$, where $p_k$ is the transmit energy of the $k$th user.

In the current work, the linear receiver is designed based on the minimum mean square error (MMSE) criterion. Thus, the linear MMSE decoding matrix of the $k$th user is given by [5]

$$\mathbf{F}_k = \mathbf{u}_{k,k}^H \left( \mathbf{U}_k \mathbf{U}_k^H + \frac{K N_0}{p_{total}} \mathbf{I}_{N_R} \right)^{-1} \tag{2}$$

where $(\cdot)^H$ represents the conjugation transpose operation, $\mathbf{I}_{N_R}$ is the $N_R \times N_R$ identity matrix, and $\mathbf{U}_k \equiv \mathbf{H}_k \mathbf{C} = \begin{bmatrix} \mathbf{u}_{k,1} & \mathbf{u}_{k,2} & \cdots & \mathbf{u}_{k,K} \end{bmatrix}$, where $\mathbf{u}_{k,k} = \mathbf{H}_k \mathbf{c}_k$ is the $k$th column of $\mathbf{U}_k$. Accordingly, the signal to interference plus noise ratio (SINR) for the $k$th user after the linear MMSE decoding can be expressed as [5]

$$SINR_k = \frac{p_k \left\| \mathbf{F}_k \mathbf{H}_k \mathbf{c}_k \right\|^2}{\sum_{i=1, i \neq k}^{K} \left( p_i \left\| \mathbf{F}_k \mathbf{H}_k \mathbf{c}_i \right\|^2 + \left\| \mathbf{F}_k \right\|^2 N_0 \right)} \tag{3}$$

As previously mentioned, due to the interference of multiple streams and hardware cost, the current study investigates a joint selection of precoding and transmit antenna to maximize the system capacity. Hence, this problem can be formulated as [8]

$$\arg \max_{\{\mathbf{c}_k\}_{k=1}^{K} \in \mathcal{C}, \{\mathbf{H}'_k\}_{k=1}^{K} \in \mathcal{H}} \sum_{k=1}^{K} \log_2 \left( 1 + SINR_k \right) \tag{4}$$

where $\mathcal{C}$ represents the set that contains all the permutations of precoding vectors in the codebook, $\mathbf{H}'_k$ is the $N_R \times n_T$ sub-matrix of $\mathbf{H}_k$, and $\mathcal{H}$ denotes the union of all possible channel sub-matrices of $K$ users. The searching space becomes large because the precoding vector and antenna subset are jointly considered. ES yields optimal performance but is too complicated to implement. Therefore, we invoke a highly efficient metaheuristic algorithm, the CE method, to solve this problem sub-optimally.

## III. THE PROPOSED CE APPROACH

In this section, the CE method, which is an efficient metaheuristic algorithm, is adopted to solve (4). The aim is to determine sub-optimally $K$ out of the $2^L$ precoding vectors and $n_T$ out of the $N_T$ transmit antennas that maximize the system capacity as given in (4).

### *A. The CE Method*

As previously mentioned, the main concept of the CE method is to invoke a probability distribution function of possible solutions. Generally, the CE method is an iterative procedure that consists of two stages. First, the random samples are generated according to a probability distribution using the stochastic properties of these samples. Then, the probability distribution is updated to produce better samples in the next iteration. Readers interested in the CE method are referred to [9].

To apply the CE method to our joint selection concern, we first transform (4) into a stochastic sampling problem by defining each sample as a series of two binary bit strings $\boldsymbol{\omega}_{1 \times N_T}$ and $\boldsymbol{\omega}_{1 \times \left\lceil \log_2 \left( P_K^{2^L} \right) \right\rceil}$. Each bit of $\boldsymbol{\omega}_{1 \times N_T}$ represents each transmit antenna being selected or not (one is for positive and zero is for negative), and $\boldsymbol{\omega}_{1 \times \left\lceil \log_2 \left( P_K^{2^L} \right) \right\rceil}$ denotes the selection of permutation of $2^L$ precoding vectors. In other words,

$$\boldsymbol{\omega} = \left[\omega_0 \quad \omega_2 \quad \cdots \quad \omega_{Q-1}\right] = \left[\boldsymbol{\omega}_{1\times N_T} \quad \boldsymbol{\omega}_{1\times\left\lceil \log_2\left(P_K^{2^L}\right)\right\rceil}\right] \tag{5}$$

where $Q = N_T + \left\lceil \log_2\left(P_K^{2^L}\right)\right\rceil$, $\lceil \cdot \rceil$ is the ceiling operation, and $\omega_q \in \{0,1\}\ \forall q$. Each element of $\boldsymbol{\omega}$ is modeled as an independent Bernoulli random variable with a probability mass function $\Pr\{\omega_q = 1\} = v_q$, $\Pr\{\omega_q = 0\} = 1-v_q$, for $q = 0, 1, ..., Q-1$. The problem is now converted to a discrete case; thus, a family of Bernoulli pdfs associated with $\boldsymbol{\omega}$ can be written as

$$f(\boldsymbol{\omega};\mathbf{v}) = \prod_{q=0}^{Q-1} v_q^{1_q(\boldsymbol{\omega})}(1-v_q)^{1-1_q(\boldsymbol{\omega})} \tag{6}$$

where $1_q(\boldsymbol{\omega}) \in \{0,1\}$ is an indicator function that denotes whether the $q$th element of $\boldsymbol{\omega}$ is selected, with "1" for positive and "0" for negative.

In addition, because $\boldsymbol{\omega}_{1\times N_T}$ is a binary vector subject to certain constraint, i.e., the number of 1s in the selected antenna is restricted, an additional operation is necessary to guarantee that each sample falls into the feasible solution region by fixing the number of 1s. This operation is known as the restricted search operation [11], in which the 1s are randomly added or removed to meet the selection constraint. For $\boldsymbol{\omega}_{1\times\left\lceil \log_2\left(P_K^{2^L}\right)\right\rceil}$, a re-sampling is required to determine if its value is larger than $P_K^{2^L}$. As a result, the CE method is forced to search in a reduced searching space.

With the definition of the sampling distribution $f(\boldsymbol{\omega};\mathbf{v})$, the CE method can be performed by repeating three major steps until the stop condition is satisfied. First, we exploit the pdf $f\left(\cdot,\mathbf{v}^{(t-1)}\right)$ to produce $N_s$ samples $\{\boldsymbol{\omega}^{(i,t)}\}_{i=1}^{N_s}$, where $t$ denotes the iteration index of the CE method, and $N_s$ is the number of samples. In step 2, the elite set $\Xi^{(t)} = \left\{\boldsymbol{\omega}^{(i,t)} : \mathcal{G}\left(\boldsymbol{\omega}^{(i,t)}\right) \ge T^{(t)}\right\}$ is generated by setting a predefined threshold $T^{(t)}$ at the $t$th iteration, where $\mathcal{G}\left(\boldsymbol{\omega}^{(i,t)}\right)$ is the performance of sample $\boldsymbol{\omega}^{(i,t)}$, i.e., the system capacity given in (4). The general approach to generate the elite set $\Xi^{(t)}$ is by collecting the $\lceil \rho N_s \rceil$ best samples, where $0 < \rho \le 1$ is the ratio of the elite samples to $N_s$. Therefore, $T^{(t)}$ can be expressed as

$$T^{(t)} = \tilde{\mathcal{G}}^{(\lceil \rho N_s \rceil)} \tag{7}$$

where $\tilde{\mathcal{G}}^{(i)}$ denotes the $i$th order statistic of the performances $\mathcal{G}\left(\boldsymbol{\omega}^{(i,t)}\right), 1 \le i \le N_s$. In step 3, we update the distribution function $f\left(\cdot,\mathbf{v}^{(t)}\right)$ to approach the goal distribution $f\left(\cdot,\mathbf{v}^{*}\right)$ by minimizing the Kullback–Leibler divergence, which is equivalent to solving [9]

$$\mathbf{v}^{(t)} = \arg\max_{\mathbf{v}} \sum_{i=1}^{N_s} I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}} \ln f\left(\boldsymbol{\omega}^{(i,t)};\mathbf{v}\right) \tag{8}$$

where $I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}}$ is an indicator function that denotes whether the $i$th sample $\boldsymbol{\omega}^{(i,t)}$ is included in the elite set, i.e., $I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}} = 1$ if $\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}$; otherwise $I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}} = 0$. By taking the first derivative with respect to $v_q$ as zero for $q = 0, 1, ..., Q-1$, the updated rule can be rewritten as

$$v_q^{(t)} = \frac{\sum_{i=1}^{N_s} I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}} 1_q\left(\boldsymbol{\omega}^{(i,t)}\right)}{\sum_{i=1}^{N_s} I_{\left\{\boldsymbol{\omega}^{(i,t)} \in \Xi^{(t)}\right\}}},\ q = 0, 1, ..., Q-1. \tag{9}$$

Instead of directly updating parameter $\mathbf{v}^{(t-1)}$ to $\mathbf{v}^{(t)}$, a smoothing process [9] is adopted to avoid the local optimum, which is denoted by

$$\mathbf{v}^{(t)} = \alpha \times \mathbf{v}^{(t)} + (1-\alpha) \times \mathbf{v}^{(t-1)} \tag{10}$$

Obviously, when $\alpha = 1$, the smoothing process degenerates to the original updating formulation (9).

The aim of the CE method is to find an optimal sampling distribution $f\left(\cdot,\mathbf{v}^{*}\right)$ that leads to the optimal selection of the transmit-antenna subset and precoding vectors with a probability of one. Thus, the update operation is iteratively processed. The convergence proof of the CE method can be found in the Appendix of [12].

To end this subsection, we summarize the procedure of the proposed CE method-based joint precoding and transmit-antenna-selection algorithm as follows:

1) Initialize the iteration counter $t = 1$ and the probability vector $\mathbf{v}^{(0)} = \left\{v_q^{(0)}\right\}_{q=0}^{Q-1}$, where $v_q^{(0)} = 0.5$, $\forall q$.
2) Exploit the density function $f\left(\cdot,\mathbf{v}^{(t-1)}\right)$ to generate randomly $N_s$ samples $\left\{\boldsymbol{\omega}^{(i,t)}\right\}_{i=1}^{N_s}$, Ensuring that each sample falls into the feasible set by employing the restricted search operator or re-sampling.
3) Examine the objective values $\mathcal{G}\left(\boldsymbol{\omega}^{(i,t)}\right)$, $i = 1, ..., N_s$, i.e., the system capacity, for each sample.
4) Determine the threshold according to (7) and obtain the elite set $\Xi^{(t)}$.
5) Calculate the parameter $\mathbf{v}^{(t)}$ using (9) and smoothly update it as (10).
6) With $t = t + 1$, repeat Steps 2–6 until the maximum number of iterations is met.

### *B. Computational Complexity*

As the CE method is essentially a population-based search method similar to other metaheuristic algorithms, the complexity can be evaluated by counting the number of total

TABLE I. COMPARISON OF COMPLEXITIES OF VARIOUS ALGORITHMS

| **Algorithms** | **Number of Capacity Calculation** |
|---|---|
| ES | $C_{n_T}^{N_T} \times P_K^{2^L}$ |
| GA | $N_s \times Iter$ |
| CE Method | $N_s \times Iter$ |

samples, i.e., the calculation times of the system capacity in (4). In this subsection, we consider the complexities of the ES, GA, and CE methods. Complexity can be discussed exhaustively in two aspects, namely, transmit-antenna selection and precoding vector selection. In the first part, because of the selection of $n_T$ out of the $N_T$ transmit antenna, the number of the calculated capacity is $C_{n_T}^{N_T}$ ; for the precoding selection, the complexity is $P_K^{2^L}$ , as different permutations yield different output. Thus, the ES of a joint scheme requires $C_{n_T}^{N_T} \times P_K^{2^L}$ times to calculate (4), which is too large to implement. If we adopt the GA or CE methods to solve this problem, they both require $N_s \times Iter$ times, where $N_s$ is the number of samples, and $Iter$ is the number of iterations. However, the GA and CE methods only guarantee suboptimum performance, which is at the expense of lower complexity. The performance evaluation will be provided in the next section. Finally, we summarize this subsection by showing the results in Table I.

## IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed CE method and conventional GA by computer simulations, in which the downlink MU-MIMO system is considered. We also assume the number of user $K$ = 2, the number of receive antennas $N_R$ = 2, and the number of transmit antenna $N_T$ = 16; the number of selected antenna is $n_T$ = 4 only. Furthermore, the data symbols are BPSK modulated and delivered through the i.i.d. Rayleigh fading channels to each user. The codebook adopted here is the six-bit Grassmannian codebook, which contains 64 precoding vectors [13]. Thus, $Q = 16 + \lceil \log_2 4032 \rceil = 28$ .In the current study, we compare four approaches to implement the joint precoding and transmit-antenna selection, including the random search, ES, conventional GA, and the proposed CE methods, in terms of system capacity and bit error rate (BER). With regard to the parameters used in the conventional GA method, the crossover and mutation probabilities are set to 50% and 1%, respectively; for the proposed CE method, the smoothing parameter $\alpha$ is 0.8, and the ratio of the elite samples to population size $\rho$ is 0.1. The size of the population $N_s$ in both methods is 840. The algorithms are stopped when the iteration index exceeds the predefined threshold.

Fig. 2 shows the average capacity versus the number of iterations at SNR = 6 dB. The proposed CE method provides a better average capacity than the conventional GA and random search methods under the same complexity. Conversely, the

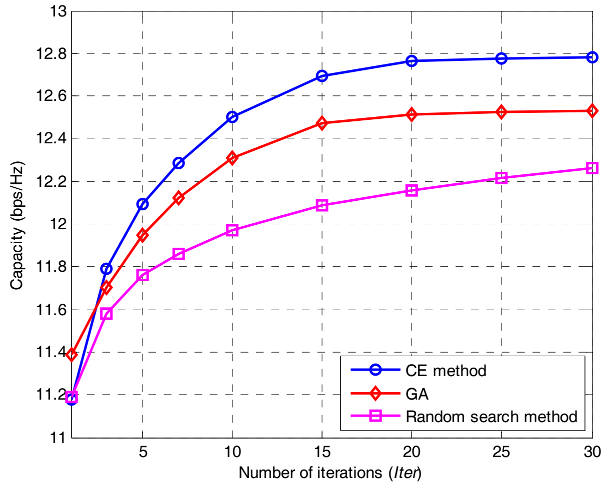


Figure 2. Capacity comparison with different number of iterations.

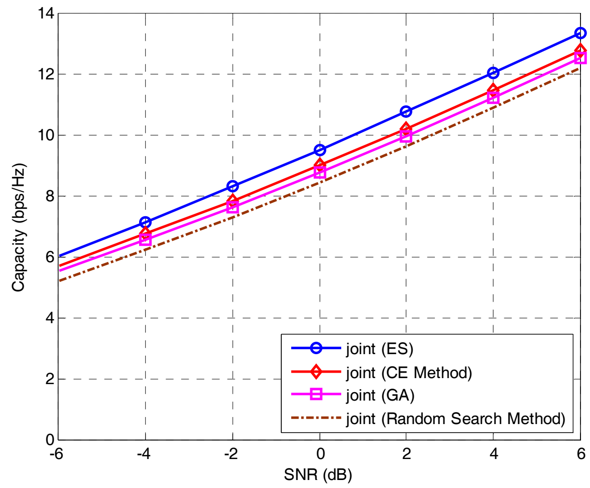


Figure 3. Capacity versus SNR for different methods.

proposed CE method obtains a lower complexity under the same average capacity performance. The capacity improves as $Iter$ increases. However, the improvement becomes negligible when $Iter \geq 20$ in both CE and GA methods. Thus, we choose $N_s$ = 840 and $Iter$ = 20 in the following simulation.

The capacity performance versus SNR of the various schemes is shown in Fig. 3. The receiver with ES yields optimum performance as expected, and the proposed CE method outperforms the conventional GA and random search methods in all SNRs. We compare the BER performance in Fig. 4, which again proves that the CE method shows better performance than the conventional GA and random search methods. The computational complexity, based on the definition presented in Section III-B, is 7,338,240 in ES and 16,800 in both CE and GA algorithms. Hence, complexity can be significantly decreased using the CE method in solving this problem. However, the price is the performance gap compare with ES, as shown in Figs. 3 and 4.

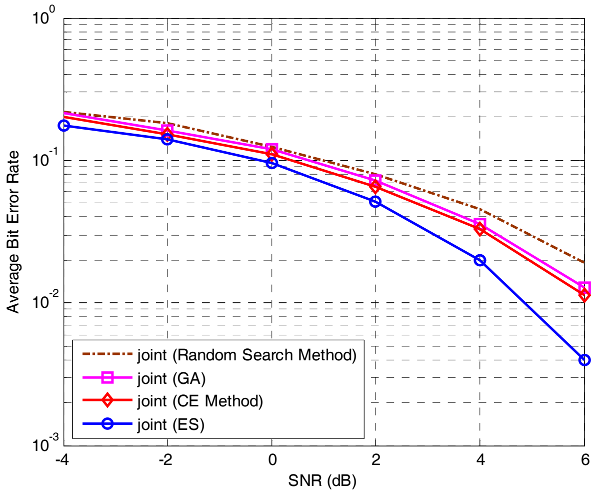


Figure 4. Average BER versus SNR for different methods.

## V. CONCLUSION

The current paper presents a CE-based scheme that realizes the joint precoding and transmit-antenna selection in the downlink of MU-MIMO systems with limited feedback to reduce the interference effectively and lower the required RF chains. With the aid of the CE method, the large amount of search required can be successfully reduced. The simulations demonstrate that the proposed CE method not only provides better capacity performance but also enjoys complexity advantages compared with the conventional GA method.

## ACKNOWLEDGMENT


The work of J.-C. Chen was supported in part by the National Science Council (NSC) of Taiwan under Grant NSC 100-2221-E-017-006-MY3. The work of C.-P. Li was supported in part by the NSC of Taiwan under Grants NSC 100-2219-E-007-010 and NSC 100-2628-E-110-004.


## REFERENCES


[1] G. J. Foschini and M. J. Gans, “On limites of wireless communications in a fadinf environment when using multiple antennas,” *Wirel. Pers. Commun.*, vol. 6, pp. 311–335, Feb. 1998.

[2] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, “An introduction to the multi-user MIMO downlink,” *IEEE Commun. Mag.*, pp. 60–67, Oct. 2004.

[3] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Sig. Proc.,* vol. 52, no. 2, pp. 461–471, Feb. 2004.

[4] D. J. Love, R. W. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Trans. Inform. Theory,* vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[5] S. Fang, L. H. Li, and P. Zhang, “Grassmannian precoding MU-MIMO scheme,” *Journal of China Universities of Posts and Telecommunications*, vol. 15, no. 3, pp. 105–109, Sept. 2008.

[6] S. Sanayei and A. Nosratinia, “Antenna selection in MIMO systems,” *IEEE Commun. Mag.* vol. 42, pp. 68–73, Oct. 2004.

[7] R. W. Heath, S. Sandhu, and A. Paulraj, “Antenna selection for spatial multiplexing systems with linear receivers,” *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142–144, Apr. 2001.

[8] S.-C. Huang, W.-H. Fang, H.-S. Chen, and Y.-T. Chen, “Hybrid genetic algorithm for joint precoding and transmit antenna selection in multiuser MIMO systems with limited feedback,” *in Proc. IEEE VTC spring*, May 2010, pp. 1–5.

[9] R. Y. Rubinstein and D. P. Kroese, The Cross-Entropy Method. Berlin, Germany: Springer, 2004.

[10] D. J. Love, R. W. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Trans. Inform. Theory,* vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[11] S. Salcedo-Sanz, G. Camps-Valls, F. Perez-Cruz, J. Sepulveda-Sanchis, and C. Bousono-Calzon, “Enhancing genetic feature selextion through restricted search and Walsh analysis,” *IEEE Trans. Syst. Man Cybern. C Appl. Rev*., vol. .4, no. 4, pp.398–406, Nov. 2004.

[12] Y. Zhang, C. Ji, W. Malik, D. O’Brien, and D. Edwards, “Receive antenna selection for MIMO systems over correlated fading channels,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4393–4399, Sep. 2009.

[13] D. J. Love. Grassmannian Subspace Packing. [Online] Available: http://cobweb.ecn.purdue.edu/□djlove/packings/one_d