Decremental User Selection for Large-Scale Multi-User MIMO Downlink with Zero-Forcing Beamforming

Shengchun Huang, *Student Member, IEEE,* Hao Yin, Haoming Li, *Student Member, IEEE,* and Victor C. M. Leung, *Fellow, IEEE*

***Abstract*—This paper proposes a decremental user selection algorithm based on zero-forcing beamforming when the number of users K in the network is smaller than the number of antennas M at a base station. The algorithm is specifically designed for large-scale multi-user multiple-input multiple-output (MIMO) downlink channels. While previous user selection algorithms are based on incremental search that starts from an empty user set, our proposed delete the minimum lambda (DML) algorithm starts by selecting all users and then deleting one user per iteration. DML substantially reduces the computational complexity as the cardinality of the final user set is close to K. Simulation results indicate that on average DML achieves an equal or higher sum rate performance than previous algorithms with greatly reduced complexity of $O(MK^2)$.**

***Index Terms*—Large-scale MIMO, decremental user selection, multi-user MIMO, zero-forcing beamforming.**

I. Introduction

In a multi-user multiple-input multiple-output (MIMO) communication system where zero-forcing beamforming (ZFBF) is used in downlink transmissions [1], the base station (BS) has to select a subset of users to maximize the sum rate. Previous studies on user selection have mostly focused on the scenario where the number of users K is larger than the number of antennas M at the BS because of two reasons: 1) the BS with M transmit antennas can simultaneously serve K single antenna users without user selection when $K \leq M$; 2) small-scale M was considered in the past such that brute-force exhaustive search can be utilized to obtain the global optimal user subset even if user selection is needed. However, in large-scale MIMO systems where tens or even hundreds of antennas are equipped at the BS, the 'Select All' strategy provides a poor sum rate performance especially when $K \to M$ [2] and the brute-force exhaustive search has a prohibitively high complexity. For instance, when $K = M = 50$ and the transmit signal-to-noise ratio (SNR) is $20\,dB$, the 'Select All' strategy on average achieves less than 33.8% of the sum rate of the globally optimal user subset. On the other hand, exhaustively searching for the globally optimal user subset with brute-force involves the evaluation and comparisons of $\sum_{i=1}^{K} \frac{M!}{i!(M-i)!}$ sum rates, which translates to 10^{15} comparisons for $K = M = 50$ and may be prohibitively costly. Thus, a sub-optimal user selection algorithm with reduced complexity is preferred.

Several user selection algorithms have been proposed for multi-user MIMO downlink transmission with ZFBF in the literature, such as zero-forcing with selection (ZFS) [3] and semi-orthogonal user selection (SUS) [4]. ZFS starts with an empty user set and adds one user in each step to maximize the sum rate increment. SUS searches for a user set with near-orthogonal channel vectors. ZFS provides a higher sum rates performance than SUS [5], while both algorithms have a complexity of $O(MK^3)$. However, all these algorithms are based on incremental search that starts from an empty user set and adds one user in each iteration. In the case of $K \leq M$, the globally optimal user set includes almost all users, which is influenced by the channel matrix and the transmit power constraint. As a result, the decremental user selection algorithm that we propose in this paper is able to greatly reduce the computational complexity of the algorithm.

In this paper, ZFBF rather than linear minimum mean square error (MMSE) precoding is considered because a large-scale MIMO system usually works in the medium to high SNR region. ZFBF achieves the same sum rate as MMSE in this region but with lower computation complexity, and ZFBF does not need the noise covariance matrix that is necessary for MMSE precoding [6].

Main contribution: We propose a decremental user selection algorithm called delete the minimum lambda (DML) for $K \leq M$ that achieves a slightly higher sum rate than ZFS with a greatly reduced complexity of $O(MK^2)$. This algorithm starts by selecting all users and then deletes one user that has the minimum effective-channel-gain in each iteration until the sum rate decreases. DML involves the same level of complexity as the 'Select All' strategy that serves all users without selection, which will be proved in Section III.

II. System Model

Consider a single cell quasi-static flat-fading MIMO downlink channel with M transmit antennas at the BS serving K ($K \leq M$) single antenna users. Let $\mathbf{H} = [\mathbf{h}_1^*, \cdots, \mathbf{h}_K^*]^* \in \mathbb{C}^{K\times M}$ be the channel matrix of all users, where $\mathbf{h}_k = [h_{k,1}, \cdots, h_{k,M}] \in \mathbb{C}^{1\times M}$ is the channel vector of user k and $\mathbf{h}_k^*$ is the complex conjugate of $\mathbf{h}_k$. Assume the BS has full knowledge of $\mathbf{H}$ and the set of indexes of the selected users is $S = \{\pi(1), \cdots, \pi(|S|)\} \subset \{1, \cdots, K\}$, where $\pi(n)$ is the index of the n-th selected user and $|S|$ denotes the size of set S. The transmit signal vector $\mathbf{x}$ is a linear combination

Manuscript received May 30, 2012. The associate editor coordinating the review of this letter and approving it for publication was M. Tao.

S. Huang and H. Yin are with the School of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, P. R. China (e-mail: huangsc@nudt.edu.cn, yinhao1@263.net).

H. Li and V. C. M. Leung are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: {hlih, vleung}@ece.ubc.ca).

Digital Object Identifier 10.1109/WCL.2012.070312.120400

of all selected users' data streams, constructed as

$$\mathbf{x} = \sum_{i\in S} \mathbf{w}_i \sqrt{p_i} s_i \,, \tag{1}$$

where $\mathbf{w}_i \in \mathbb{C}^{M\times 1}$ is the ZFBF weight vector, p_i is the transmit power scaling factor and s_i is the information symbol of user i. Thus, the received signal at user $k \in S$ is given by

$$y_k = (\mathbf{h}_k \mathbf{w}_k \sqrt{p_k}) s_k + \sum_{i\in S, i\neq k} (\mathbf{h}_k \mathbf{w}_i \sqrt{p_i}) s_i + n_k \,, \tag{2}$$

where n_k is white Gaussian noise with zero mean and unit variance. The power constraint for the transmitted signal is $E\{\mathbf{x}^*\mathbf{x}\} \leq P$. Since the noise has unit variance, P also represents the total transmit SNR [7].

A. Zero-Forcing Beamforming

ZFBF inverts the channel matrix at the transmitter in order to create orthogonal channels between BS and users without users' cooperation. The ZFBF precoding matrix $\mathbf{W}$ for S is the Moore-Penrose pseudo-inverse of the channel matrix $\mathbf{H}_S$ of the selected users

$$\mathbf{W} = [\mathbf{w}_{\pi(1)}, \cdots, \mathbf{w}_{\pi(|S|)}] = \mathbf{H}_S^* (\mathbf{H}_S \mathbf{H}_S^*)^{-1} \,. \tag{3}$$

The beamforming vector $\mathbf{w}_i$ can also be obtained through the *effective channel vector* (ECV) $\boldsymbol{\nu}_i$ defined by [5]

$$\mathbf{w}_i = \frac{\boldsymbol{\nu}_i^*}{\|\boldsymbol{\nu}_i\|^2} \tag{4}$$

$$\boldsymbol{\nu}_i = \mathbf{h}_i \mathbf{P}_i^{\perp} \,, \tag{5}$$

where $\mathbf{P}_i^{\perp} = \mathbf{I}_M - \mathbf{H}_{S\setminus\{i\}}^* (\mathbf{H}_{S\setminus\{i\}} \mathbf{H}_{S\setminus\{i\}}^*)^{-1} \mathbf{H}_{S\setminus\{i\}}$ is the orthogonal projector matrix on the subspace $V_i = span\{\mathbf{h}_j | j \in S, j \neq i\}$ [8], $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\mathbf{H}_{S\setminus\{i\}}$ is the row-reduced channel matrix of all the selected users except user i. $S \setminus \{i\}$ denotes the set difference that deletes the element i from the set S.

The sum rate achieved by S is

$$R(S) = \max_{p_i: \sum_{i\in S} \lambda_i^{-1} p_i \leq P} \sum_{i\in S} \log(1 + p_i) \,, \tag{6}$$

where

$$\lambda_i = \frac{1}{\|\mathbf{w}_i\|^2} = \|\boldsymbol{\nu}_i\|^2 \tag{7}$$

is the effective-channel-gain of user i [4], and $\lambda_i^{-1} p_i = \|\mathbf{w}_i\|^2 p_i$ is the transmit power allocated to user i. The optimal p_i in (6) can be found by waterfilling

$$p_i = (\mu \lambda_i - 1)^+ \,, \tag{8}$$

where $(x)^+ = \max\{x, 0\}$, and μ is the water level satisfying

$$\sum_{i\in S} \left(\mu - \lambda_i^{-1}\right)^+ = P \,. \tag{9}$$

The sum rate (6) of ZFBF can be optimized with respect to the selected user set S. Thus, the user selection problem can be formulated as

$$\begin{array}{ll} maximize & R(S) \\ subject\ to & S \subset \{1, \cdots, K\} \,. \end{array} \tag{10}$$

III. DML Algorithm

The DML algorithm works as follows: it starts by selecting all users and then deletes the user with the minimum effective-channel-gain λ_k in each iteration until the sum rate increment $\Delta R = R(S \setminus \{k\}) - R(S) < 0$, and then calculates the precoding matrix with ZFBF and waterfilling power allocation.

A. λ updating scheme for decreased user set

Before constructing the DML algorithm, we provide here an efficient effective-channel-gain updating scheme for all the remaining users in $S\setminus\{k\}$ when user $k \in S$ with the minimum λ_k is deleted from the selected user set S. Denote the updated effective-channel-gain, ECV and the ZFBF weight vector of the remaining user $i \in S\setminus\{k\}$ as λ_i^-, $\boldsymbol{\nu}_i^-$ and $\mathbf{w}_i^-$, respectively. An ECV based λ_i^- updating scheme is provided in [5] as:

$$\lambda_i^- = \frac{\lambda_i^2 \lambda_k}{\lambda_i \lambda_k - |\boldsymbol{\nu}_i \boldsymbol{\nu}_k^*|^2} \tag{11}$$

$$\boldsymbol{\nu}_i^- = \frac{\lambda_i \lambda_k}{\lambda_i \lambda_k - |\boldsymbol{\nu}_i \boldsymbol{\nu}_k^*|^2} \left(\boldsymbol{\nu}_i - \frac{\boldsymbol{\nu}_i \boldsymbol{\nu}_k^*}{\lambda_k} \boldsymbol{\nu}_k \right) . \tag{12}$$

By plugging (4) and (7) into (11) and (12), we can get the updated λ_i^- based on the beamforming vector $\mathbf{w}_i$ as:

$$\lambda_i^- = \frac{\lambda_i}{1 - \lambda_k \lambda_i |\mathbf{w}_k^* \mathbf{w}_i|^2} \tag{13}$$

$$\mathbf{w}_i^- = \mathbf{w}_i - \lambda_k \mathbf{w}_k^* \mathbf{w}_i \mathbf{w}_k \,. \tag{14}$$

B. Construction of DML algorithm

Let S be the index set of selected users. The $\mathbf{w}_i$ and λ_i are respectively the ZFBF weight vector and effective-channel-gain of selected user $i \in S$. By utilizing the $\mathbf{w}_i$ based λ_i updating scheme in (13) and (14), DML is constructed as follows.

Step 1) Initialization:

$$S = \{1, \cdots, K\}$$

$$\lambda_i = \frac{1}{\|\mathbf{w}_i\|^2}, \quad i \in S \,.$$

where $[\mathbf{w}_1, \cdots, \mathbf{w}_K] = \mathbf{H}^*(\mathbf{H}\mathbf{H}^*)^{-1}$.

Step 2) Delete the user with the minimum λ:

$$k = \arg\min_{i\in S} \{\lambda_i\} \tag{15}$$

$$\lambda_i^- = \frac{\lambda_i}{1 - \lambda_k \lambda_i |\mathbf{w}_k^* \mathbf{w}_i|^2}, \quad i \in S \setminus \{k\} \,. \tag{16}$$

If $\Delta R = R(S\setminus\{k\}) - R(S) \geq 0$, update S, $\mathbf{w}_i$, λ_i by deleting user k and then go to Step 2).

$$S \leftarrow S \setminus \{k\}$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \lambda_k \mathbf{w}_k^* \mathbf{w}_i \mathbf{w}_k, \quad i \in S$$

$$\lambda_i = \lambda_i^-, \quad i \in S \,.$$

Otherwise, go to Step 3).

Step 3) Precoding matrix:

$$\mathbf{W} = \left[\sqrt{\mu \lambda_{(1)} - 1}\, \mathbf{w}_{(1)}, \ \cdots, \ \sqrt{\mu \lambda_{(|S|)} - 1}\, \mathbf{w}_{(|S|)} \ \right], \tag{17}$$

where

$$\mu = \frac{1}{|S|} \left(P + \sum_{i\in S} \lambda_i^{-1} \right) \tag{18}$$

is the water level for power allocation, $\mathbf{w}_{(i)}$ and $\lambda_{(i)}$ are the weight vector and effective-channel-gain of the i-th user in S.

Step 1) initializes by serving all users and calculates the beamforming vector $\mathbf{w}_i$ and effective-channel-gain λ_i according to (3) and (7). In Step 2), the user $k \in S$ with the minimum effective-channel-gain is deleted from S if the removal of user k increases the sum rate. However, if the condition $\eta = \left(P + \sum_{i \in S} \lambda_i^{-1}\right)/|S| \leq \lambda_k^{-1}$ is satisfied, the user k is deleted without calculating ΔR. Because $\eta \leq \lambda_k^{-1}$ indicates that the user k is allocated with zero transmit power by waterfilling over $\lambda_i \in S$, deleting users with $p_i = 0$ provides positive sum rate increment [3]. If $\eta > \lambda_k^{-1}$, the sum rate of S is calculated as

$$R(S) = \sum_{i \in S} \log(\eta \lambda_i)\,. \tag{19}$$

The sum rate $R(S \setminus \{k\})$ can also be achieved as in (19) without the iterative waterfilling if the inequality (20) holds:

$$\frac{1}{(|S|-1)}\left(P + \sum_{i \in S \setminus \{k\}} \frac{1}{\lambda_i^-}\right) > \frac{1}{\min_{i \in S \setminus \{k\}} \lambda_i^-}\,. \tag{20}$$

If (20) does not hold, the user $k' = \arg\min_{i \in S \setminus \{k\}} \lambda_i^-$ is deleted and the sum rates are compared between $S \setminus \{k, k'\}$ and S, and then the user set with a larger sum rate is selected for the next iteration.

When DML exits from the user deletion loop in Step 2), $\eta > (\min_{i \in S_{DML}} \lambda_i)^{-1}$ is satisfied for the final user set S_{DML} when entering Step 3). Now every user in S_{DML} is allocated with positive transmit power after waterfilling power allocation. According to (9), the water level is calculated as in (18) and the transmit power scaling factor p_i is

$$p_i = \mu \lambda_i - 1\,. \tag{21}$$

The precoding matrix (17) is obtained by plugging (21) into (2).

C. Complexity analysis

The complexity of DML lies mainly in the initialization step of the Moore-Penrose pseudo-inverse of $\mathbf{H}$, which involves a complexity of $O(MK^2)$ [8]. We now analyze the computations involved in the remaining steps of DML and show that the additional complexity is no larger than $O(MK^2)$.

The corresponding λ_i initialization in Step 1) involves K 2-norms of $1 \times M$ vectors, which include MK complex multiplications. The updating of $\mathbf{w}_i$ and λ_i in Step 2) involves $|S| - 1$ vector-vector multiplications and $|S| - 1$ 2-norms, which include $2M(|S|-1)$ complex multiplications. Suppose the selected user set generated by DML is S_{DML}. The total complexity is $MK + \sum_{n=|S_{DML}|}^{K} 2M(n-1)$, which is smaller than $MK + \sum_{n=1}^{K} 2M(n-1) = MK^2$ and is asymptotically $O(MK)$ since DML selects almost all users that $|S_{DML}| \to K$, especially when the transmit SNR is large and K is small.

We have shown that the complexity of DML is $O(MK^2)$, which is the same as the 'Select All' strategy that also has complexity of $O(MK^2)$ due to Moore-Penrose pseudo-inverse of $\mathbf{H}$. This is a significant improvement over previous user selection algorithms, such as ZFS and SUS, which all have a complexity of $O(MK^3)$ [3], [9].

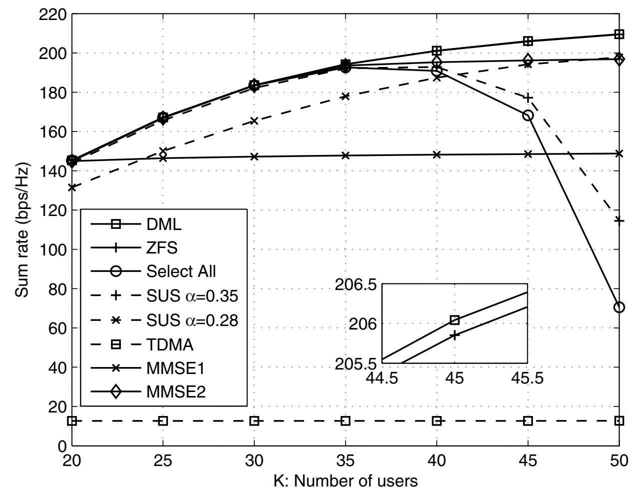

Fig. 1. Sum rate performance comparison of DML, ZFS, SUS, 'Select All' with ZFBF, TDMA and MMSE beamforming serving a fixed number of users, with $M = 50$ and $P = 20\,dB$.

IV. Simulation Results

In this section, we compare the performance of DML, ZFS, SUS, 'Select All' with ZFBF, TDMA and MMSE beamforming serving a fixed number of users. The BS in the simulated multi-user MIMO system is equipped with $M = 50$ antennas. The transmit SNR is $20\,dB$, and the number of users K ranges from 20 to 50. All curves are obtained by averaging over 10^5 independent channel matrices with each entry being a zero-mean unit-variance circular symmetric complex Gaussian random variable.

We compare the sum rate and the number of selected users in Figs. 1 and 2. TDMA serves the best channel user with maximum ratio transmission and full power. MMSE1 and MMSE2 are schemes which serve twenty and $\min(K, 35)$ best channel users with MMSE beamforming, respectively. The sum rate increases with K for DML, ZFS, TDMA, MMSE1 and MMSE2, while DML provides a slightly higher sum rate than ZFS on average. The sum rate of 'Select All' increases with K when $K \leq 35$ and then drops after $K \geq 40$, while the sum rate variation of SUS highly depends on the choice of threshold α. From Fig. 2 we see that the number of selected users increases with K for all the algorithms considered except TDMA, MMSE1 and MMSE2.

TDMA provides the worst sum rate performance that achieves only 6.07% the sum rate of ZFS at $K = 50$, because it does not exploit the multiplexing gain. The 'Select All' strategy provides a good sum rate before K reaches 30. However, it provides worse sum rate performance when $K \geq 40$. According to Fig. 1, it achieves only 33.8% the sum rate of ZFS at $K = M = 50$, and the ratio will decrease further when the transmit SNR P decreases, indicating that user selection is crucial even for the scenario $K \leq M$.

SUS is also not suitable for $K \leq M$ as it is highly sensitive to the choice of threshold α, and for any given α SUS cannot guarantee good sum rate performance over the whole range of K. When $\alpha = 0.28$, which is the optimum threshold for $K =$

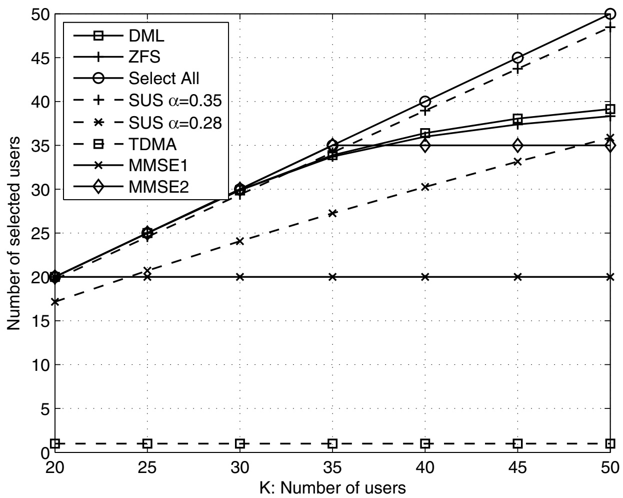

Fig. 2. The comparison of the numbers of selected users under DML, ZFS, SUS, 'Select All' with ZFBF, TDMA and MMSE beamforming serving a fixed number of users, with $M = 50$ and $P = 20\,dB$.

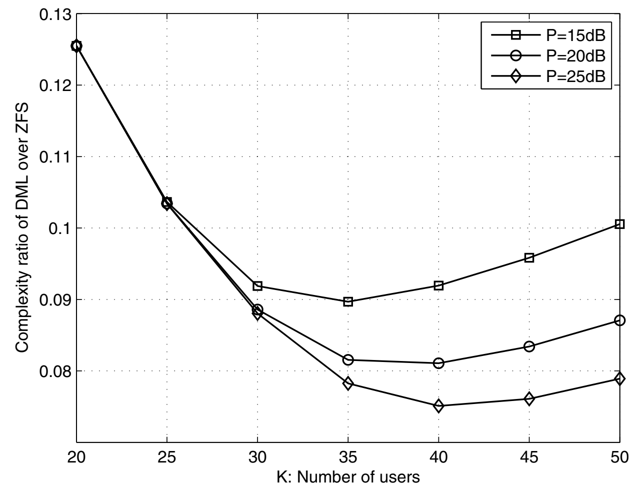

Fig. 3. Complexity ratio of DML over ZFS for $M = 50$ and $P = 15,\ 20$ and $25\,dB$.

50, SUS provides low sum rate when K is small; when α is larger than 0.28, such as $\alpha = 0.35$ in Fig. 1, SUS provides low sum rate when $K \to M$; when α is smaller than 0.28, the sum rate performance of SUS will decrease in the whole K range considered. The number of selected users under SUS increases linearly with K in Fig. 2, where the slope is proportional to the threshold α.

The sum rate of MMSE1 and MMSE2 increase slightly with K for $K \geq 20$ and $K \geq 35$, respectively, because the achieved multi-user diversity gains increase with K when serving a fixed number of users. However, since the best user set, which provide the highest sum rate for a given precoding scheme, is a compromise between high channel gain and the orthogonality of different channels, both MMSE1 and MMSE2 have certain performance loss as they consider only the channel gain factor. When $K \geq 35$, the gaps between MMSE2 and ZFS increase with K for both the sum rates and the number of selected users as shown in Fig. 1 and 2.

DML and ZFS achieve a higher sum rate than all the other schemes considered in Fig. 1. DML achieves a slightly higher sum rate than ZFS on average, and there is a 0.2 bps/Hz sum rate increment over ZFS at $K = 45$. According to Fig. 2, the numbers of selected users of DML and ZFS increases linearly when $K \leq 25$ because the BS can support all users in this range; the increment ratio decreases when $K \to M$ because the multi-user interference increases with K and more users should be deleted to achieve the maximum sum rate.

Fig. 3 shows the complexity ratio of DML over ZFS, which is defined as the number of multiplications involved in DML as a faction of that of ZFS, for $P = 15,\ 20$ and $25\,dB$. DML has a computation complexity that is only 7.5% to 12.5% of that of ZFS, which favors practical implementation greatly. The complexity reduction is attributed to three reasons. First, a smaller number of iterations is involved in DML as most users are selected; second, DML chooses a user with the minimum effective-channel-gain λ in each iteration, while ZFS has to calculate sum rates for all possible user set in each iteration; third, the weight vector $\mathbf{w}$ based λ updating scheme in (13) and (14) reduces the complexity of user selection in DML. The complexity ratio increases when K is small or when $K \to M$, because the complexity of ZFS decreases when K is small and the number of users that need to be deleted increases when $K \to M$. At higher SNRs, the complexity reduction by DML is more significant as fewer users are deleted.

V. Conclusion

In this paper, we have proposed a decremental user selection algorithm, DML, which achieves slightly higher sum rate and much lower complexity than previous user selection algorithms, such as ZFS and SUS, for multi-user MIMO systems when $K \leq M$. DML is useful for large-scale multi-antenna system as it achieves a high sum rate performance while incurring a similar complexity as the 'Select All' strategy that does not perform any user selection.

References

[1] 3GPP R1-093511, NTT DOCOMO, "Investigation on enhanced DL MU-MIMO processing based on channel vector quantization for LTE-Advanced," WG1 Meeting #58, Shenzhen, China, Aug. 2009.

[2] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, pp. 195–202, Jan. 2005.

[3] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, pp. 3857–3868, Oct. 2005.

[4] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 528–541, Mar. 2006.

[5] S. Huang, H. Yin, J. Wu, and V. C. M. Leung, "User selection for multi-user MIMO downlink with zero-forcing beamforming," submitted to *IEEE Trans. Veh. Technol.*

[6] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[7] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, July 2003.

[8] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd edition. The Johns Hopkins University Press, 1996.

[9] J. Q. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate," *IEEE Trans. Signal Process.*, vol. 56, pp. 3713–3726, Aug. 2008.