

User Selection and Precoding Schemes Based on Partial Channel Information for Broadcast MIMO Systems

Hamid Reza Bahrami¹, Tho Le-Ngoc²

¹Department of ECE, University of Akron, Ohio, USA

²Department of ECE, McGill University, Montreal, Canada

Email: hrb@uakron.edu, Tho.le-NGOC@mcgill.ca

Received December 14, 2009; revised January 16, 2010; accepted February 20, 2010

Abstract

Due to the large number of users and the time-varying characteristics of wireless channels, it is very tough to inform the transmitter of full channel information in real multi-user MIMO broadcast systems. On the other hand, the capacity of multi-user systems greatly depends on the knowledge of the channel at the transmitter while this is not always the case in single-user MIMO systems. In this paper, we investigate combined user selection and zero-forcing precoding schemes that use partial channel information, *i.e.*, very low amount of channel information at the base station. We show that while greatly reducing the complexity and channel knowledge feedback load, the proposed schemes preserve the optimality of zero-forcing scheme in term of achievable ergodic sum capacity in limit of large number of active users.

Keywords: Broadcast Channel, Multi-User System, Partial Channel Information, Precoding

1. Introduction

Linear precoding is an effective tool to achieve better performance and higher throughput in multiple-input-multiple-output (MIMO) communication systems. MIMO precoding techniques for single-user systems have been extensively studied. It was shown that the achievable capacity of a MIMO point-to-point link is limited by the minimum number of transmit and receive antennas [1,2], and linear precoding scheme can be applied to achieve this capacity [3-5].

On the other hand, evaluation of the achievable rate and capacity of multi-user MIMO broadcast (BC) channels is still an open problem. It was the first time in [6] that the achievable region of MIMO BC channels was evaluated using an approach known as dirty paper coding (DPC) [7]. The rate region was obtained for two single-antenna users analytically. Later [8] and [9] suggested a duality between multi-access (MA) and MIMO BC channels. Using this duality, the rate region and capacity of a general MIMO BC system with arbitrary number of (multi-antenna) users can be calculated. It is shown that the DPC approach can achieve the capacity of a general BC channel [8].

The duality between precoder design problem and calculating the capacity of the system, however, does not exist in MIMO BC systems. In other words, there is no lin-

ear scheme that achieves the capacity of MIMO BC channels. Instead, one has to use the DPC approach, which is a nonlinear scheme to achieve the so-called sum-rate capacity of a multi-user BC system. DPC can be implemented using a successive interference pre-subtraction (cancellation) technique at the transmitter [10] based on the structure of generalized decision-feedback equalizer (GDFE) which is a nonlinear approach.

Although the results presented in these papers are interesting from a theoretic point of view, due to the complex nature of optimization processes used in these methods and the assumptions necessary for applying DPC approach, their application in real communication systems is limited. It is mainly because of two important factors; first, the number of users in real multi-user communication systems can be very high and second, the fast changing characteristics of the MIMO wireless channels prohibits time-consuming optimization at the transmitter.

On the other hand, the presence of channel knowledge is more important in multi-user systems [11,12]. One can say that the absence of channel knowledge at the transmit side of multi-user systems will result in higher degradation in terms of performance and capacity as compared to the single-user systems. These facts draw the attention to the use of partial channel information in multi-user communication systems. Compared to the full channel knowledge, partial channel information can be obtained easily

while it is outdated more slowly. Therefore, it is more logical to assume that the transmitter of a multi-user MIMO system has partial channel knowledge. Our emphasis is that the value of partial channel information is far crucial in multi-user as the volume of full channel knowledge in a multi-user system is much larger than that in single-user systems. For example, [12] shows the effect of different levels of channel information over achievable rate in a MIMO broadcast system.

While considering the use of partial channel information in MIMO BC channels of great importance, there are few papers that address this issue in the design of practical transmission schemes. Recently, [13] and [14] suggested a random beamforming scheme based on partial channel information with a rate that can achieve the same scaling factor as the capacity obtained with perfect CSI using dirty paper coding, *i.e.*,

$$\lim_{n \rightarrow \infty} \frac{R}{C_{sum}} = \lim_{n \rightarrow \infty} \frac{R}{R_{DPC}} = 1 \quad (1)$$

where n is the number of the users, R is the achievable rate of the proposed scheme and C_{sum} is the sum capacity of the system assuming full channel knowledge at the transmitter which is the same as the rate achieved by DPC (R_{DPC}).

Our main focus in this paper is to attain the capacity of a multi-user MIMO broadcast channel assuming that the transmitter just has a partial knowledge of users' channels. In other words, we investigate downlink precoding schemes that is capable of achieving the capacity of a MIMO broadcast channel in which a multiple-antenna transmitter communicates with a number of mobile units. This approach is based on the assumption of partial channel side information available at the transmitter while each receiver has perfect channel knowledge.

In [15], using full channel knowledge at the transmit side, a zero-forcing precoding scheme has been introduced and its performance has been evaluated. It has also been proved that when the number of users tends to infinity, the zero-forcing scheme is optimal in term of capacity, *i.e.*, it can achieve the ergodic sum capacity of BC channels when the number of users is large. The problem with this scheme lies, however, in the need for full channel knowledge of the channel at the transmitter which is difficult to obtain in fast-varying wireless channels. Another difficulty with the scheme proposed in [15] is the complexity of the transmit operations. The transmitter has to find a set of best channels in each transmission interval which have the most-orthogonal channel vectors. The authors have recently generalized this scheme to the case of partial channel information by using a quantized feedback scheme in [16]. Note that, the optimality of the above approaches in [13] and [15] is asymptotical, *i.e.* the optimality decreases when the number of users reduces. The problem of zero-forcing in differ-

ent interference scenarios and SNR regimes has been also addressed in [17].

Our study shows that careful selection of channel side information is very important in the sense that it can reduce the feedback cost and transmit complexity while still provides a considerable performance. We propose zero-forcing transmission schemes that use only partial channel information with low feedback load and also facilitate the algorithm of selecting the best users at the transmitter. Although it reduces the feedback load and algorithm complexity, it can be shown that its performance can be comparable to that of the scheme proposed in [15] in term of achievable rate.

Therefore, four important issues in transmission in multi-user systems are addressed in our proposed scheme:

1) Our scheme significantly reduces the feedback load needed to calculate the optimal precoding vectors at the transmitter. The difference in feedback overhead compared to other schemes grows by increasing the number of users and/or transmits antennas.

2) It greatly relaxes the complexity of selecting the best user algorithm at the transmitter by distributing the load of processing among the users communicating in the network. The complexity issue becomes more important when the number of users and/or transmit antennas grows. Therefore, this scheme can be considered as a candidate for distributed network management schemes.

3) Our scheme achieves the same ergodic sum capacity as the scheme in [15] (and hence DPC) and the optimality preserves better especially in the systems with large number of users. However, one important issue in [15] is the fairness of user selection. The original selection scheme in [15] is not fair as the users with high channel gains (users closer to the base station) always have a better chance to be selected. Therefore, a scheduling procedure becomes inevitable to satisfy fairness requirements. We also discuss the issue of fairness for different schemes. We show that using these schemes, the chance of selecting a user does not depend on the channel gain (power) and the distance to the base station and hence all the users have equal chance to be selected as the best (winning) users.

4) We propose four different strategies for user selection, power allocation and precoding. Each of these strategies is suitable for specific propagation scenario and channel condition. However, as the structure and algorithms of all of these schemes are very similar, it is possible to implement all of them at the base station and switch amongst them when necessary. This gives a degree of robustness to the system that can cope with channel impairments and changes.

2. System Model

We consider a broadcast system using a transmitter with

M transmit antennas to serve n users, each with N_i antennas ($i = 1, \dots, n$). The channel for user i can be represented by an $N_i \times M$ matrix \mathbf{H}_i whose entries are assumed to be Gaussian with zero mean and unit variance. At time instant k , the $M \times 1$ transmit vector \mathbf{s} to different users has a constrained power, i.e., $E\{\mathbf{s}\mathbf{s}^H\} \leq P$ and the $N_i \times 1$ received signal vector for user i is

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{s} + \mathbf{n}_i \quad i = 1, 2, \dots, n \quad (2)$$

where \mathbf{n}_i is the $N_i \times 1$ additive white Gaussian noise sample vector for user i . Using precoding, the transmitted vector \mathbf{s} is a linear combination of transmitted symbols for each user \mathbf{s}_i ,

$$\mathbf{s} = \sum_{i=1}^n \sqrt{P_i} \mathbf{W}_i \mathbf{s}_i, \quad \sum_{i=1}^n P_i = P$$

where \mathbf{W}_i , \mathbf{s}_i and P_i are the $M \times N_i$ precoding matrix, transmitted symbol, and allocated power for user i , respectively. For simplicity, in the following discussions, we assume $N_i = 1$, $i = 1, \dots, n$. The received signal for the i th user can be represented as

$$y_i = \sqrt{P_i} \mathbf{h}_i \mathbf{w}_i s_i + \sum_{j \neq i} \sqrt{P_j} \mathbf{h}_i \mathbf{w}_j s_j + n_i \quad i = 1, 2, \dots, n \quad (3)$$

where the second term is due to the interference from other users. The bold-faced, lower-case symbols in the equation denote vectors corresponding to the case of $N_i = 1$. Our objective is to find the set of precoding vectors to maximize the achievable sum rate, i.e.,

$$R = \max_{\mathbf{w}_i, \sum \|\mathbf{w}_i\|^2 \leq P} \sum_{i=1}^n \log \left(1 + \frac{P_i |\mathbf{h}_i \mathbf{w}_i|^2}{1 + \sum_{j \neq i} P_j |\mathbf{h}_i \mathbf{w}_j|^2} \right) \quad (4)$$

3. BC Capacity and Zero-Forcing Precoding

The sum rate capacity of a BC channel achieved by DPC approach has been shown to be [6,8]:

$$C_{sum} = R_{DPC} = \max_{\sigma_i, \sum_{i=1}^n \sigma_i \leq P} \log \det \left(\mathbf{I} + \sum_{i=1}^n \sigma_i \mathbf{h}_i^H \mathbf{h}_i \right) \quad (5)$$

which is in fact the capacity of the dual MAC channel.

Note that σ_i 's are not the same as P_i 's in (3). In [13], it has been shown that when the number of users (n) is large, the ergodic capacity asymptotically scales like $M \log \log n$. In other words, the ergodic capacity can be written as:

¹The easiest selection of \mathbf{w}_i is such that $\mathbf{W}(S)$ is the pseudo-inverse of $\mathbf{H}(S)$ where $\mathbf{W}(S)$ is a matrix whose columns are composed of the precoding vectors in S and $\mathbf{H}(S)$ is the channel matrix whose columns are composed of channel vectors of the users in S , i.e. $\mathbf{W}(S) = \mathbf{H}^\dagger(S)$.

$$E\{C_{sum}\} = E\{R_{DPC}\} \sim M \log \left(1 + \frac{P}{M} \log n \right) \quad (6)$$

i.e., in the limits, it can provide a diversity order of $\log n$ compare to a single user system. Also, a linear increase in the capacity is achieved by increasing the number of transmit antenna (M). It can be seen from (6) that a scheme allocating an average power of $(P/M) \log n$ to each of M best users via M independent paths (subchannels), can provide the same ergodic capacity as the DPC and hence, is asymptotically optimal.

On the other hand, although suboptimum, zero-forcing scheme can provide M parallel (independent) subchannels from transmitter to M users by selecting the precoding vectors, \mathbf{w}_i 's, such that $\mathbf{h}_i \mathbf{w}_j = 0$ ($i \neq j$) (i.e., no interference from other users). Since the size of \mathbf{h}_i 's is $1 \times M$, there will be at most M precoding vectors that can satisfy the above equations. Therefore, at most M users should be selected among n available users. Let $S \subset \{1, \dots, n\}$ $|S| \leq M$ be a subset of user selected for transmission. By suitable selection of precoding matrices¹, there will be M independent subchannels and hence the achievable rate can be written as:

$$R_{ZF}(S) = \max_{P_i, \sum_{i \in S} P_i \leq P} \sum_{i \in S} \log(1 + \lambda_i P_i) \quad (7)$$

where λ_i is the i th subchannel gain. Furthermore, P_i 's can be found via a waterpouring process as $P_i = [v - \lambda_i^{-1}]^+$ where $[x]^+ = \max\{0, x\}$ for a scalar x and the constant v is such that the power constraint P is satisfied.

Note that there are different possible selections of user subsets. Achievable sum rate of optimum zero-forcing is defined as [15]:

$$R_{ZF} = \max_S R_{ZF}(S) \quad (8)$$

where $S \subset \{1, \dots, n\}$ $|S| \leq M$. The optimal solution may need a lengthy and complicated exhaustive search of all possible user subsets for zero-forcing schemes. Furthermore, required knowledge of instantaneous channel vectors (full channel information) at the transmitter may need a large amount of channel feedback load. This limits the application of optimal zero-forcing method to simple cases with small number of users. In [15], it has been shown indirectly that in the limit of large number of users (n), the zero-forcing beamforming scheme can provide a sum rate equal to that of DPC in (6),

$$E\{R_{ZF}\} \sim M \log \left(1 + \frac{P}{M} \log n \right) \sim E\{R_{DPC}\} \quad (9)$$

An appealing sub-optimum zero-forcing scheme with simple selection of M best users was discussed in [5]. However, it still needs full channel knowledge at the transmitter with potentially high complexity and feedback

load in real wireless applications.

By increasing the number of users in the network, it is more likely that there exists a subset of M users such that a linear precoding scheme (like zero-forcing) can achieve the sum rate capacity of the system. In other words, for a large user population, different suboptimum zero-forcing algorithms have similar asymptotic performance, and hence, complexity becomes the key selection factor. This paper mainly focuses on how to reduce the processing and feedback load in selecting the best M users with as much as possible orthogonal channels to maintain the asymptotic gain of $\log \log n$ over ergodic capacity in (9).

4. Partial Knowledge Zero-Forcing Schemes

Assume that transmitter and all users have a predetermined known orthonormal basis² $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M)$ of size M (e.g., standard basis for \mathbb{C}^M). We will discuss the selection of this orthonormal basis in Section 6.

Now, we consider two different strategies: *power* user selection (PUS) and *normalized* user selection (NUS). In the first strategy, at the start of each transmission period, each user calculates the projections of its channel vector on each of the vectors in orthonormal basis (\mathbf{u}_i), i.e., for j^{th} user, the projection on i^{th} vector is:

$$\gamma_{ij} = \langle \mathbf{h}_j, \mathbf{u}_i \rangle^2 = |\mathbf{h}_j \mathbf{u}_i^H|^2 ; i = 1, \dots, M; j = 1, \dots, n \quad (10)$$

In the second strategy, each user calculates the norm of the projections of its channel vector on each of the vectors in orthonormal basis, i.e.,

$$\gamma_{ij} = |\mathbf{h}_j|^{-2} \langle \mathbf{h}_j, \mathbf{u}_i \rangle^2 = |\mathbf{h}_j|^{-2} |\mathbf{h}_j \mathbf{u}_i^H|^2 ; i = 1, \dots, M; j = 1, \dots, n \quad (11)$$

Next, in both strategies, each user sends its maximum γ_{ij} along with its index to the transmitter,

$$\gamma_j = \max_i \gamma_{ij}; \alpha_j = \arg \max_i \gamma_{ij} \quad (12)$$

Transmitter then easily selects the best user for each orthonormal basis by finding the maximum γ_j over those users. Assume that the indices of users for which $\alpha_j = i$ are saved in a set $S(i)$. Therefore,

$$\gamma_{i,\max} = \max_{j \in S(i)} \gamma_j \quad i = 1, \dots, M \quad (13)$$

The transmitter selects these M users as the winning users and asks them to send back their channel vectors (totally M vectors of size $M \times 1$). It then selects the precoding vectors. For selection of precoding vectors, we consider two different precoding schemes: opportunistic precoding (OP) and *channel-aware* precoding (CAP). In *opportun-*

²A set \mathbf{U} is orthonormal if $\forall i, j \quad u_i^H u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

istic precoding, as its name implies, the transmitter does not know the channel vector of selected users and therefore transmit in the direction of orthonormal basis, i.e.

$$\mathbf{w}_i = \mathbf{u}_i; i = 1, \dots, M \quad (14)$$

where \mathbf{w}_i 's are the M precoding vectors associated to each of the selected users. Sending users' signals in the direction of orthonormal basis clearly minimizes the interference. Therefore, opportunistic precoding can also be referred as *interference-minimizer* precoding. Maximum sum-rate of the users can then be written as:

$$R_{OP} = \sum_{i=1}^M \left(1 + \frac{P}{M} \frac{|\mathbf{h}_i \mathbf{u}_i^H|^2}{1 + \sum_{j=1, j \neq i}^M |\mathbf{h}_i \mathbf{u}_j^H|^2} \right) \quad (15)$$

In the limit of large number of users, the channel vectors of M selected users are in the direction of orthonormal basis with high probability. Therefore, the sum rate can be rewritten as:

$$R_{OP} = \sum_{i=1}^M \log \left(1 + \frac{P}{M} |\mathbf{h}_i \mathbf{u}_i^H|^2 \right) = \sum_{i=1}^M \log \left(1 + \frac{P}{M} |\mathbf{h}_i|^2 \right) \quad (16)$$

(16) is the direct result of (15) and the fact that, in opportunistic precoding, at the limit of large number of users there is no interference between selected users. In channel-aware precoding, transmitter does know the selected users' channel vectors and selects the precoding vectors as the normalized vector of users' channel vectors:

$$\mathbf{w}_i = |\mathbf{h}_i|^{-1} \mathbf{h}_i \quad \forall i : 1, \dots, M \quad (17)$$

where the best M users in (15) are indexed from 1 to M . In this scheme, we target maximizing receive SNR. It is easy to see that by sending in the direction of selected users' channel vectors, we maximize the receive SNR of each user regardless of interference introducing to other users. Therefore, channel-aware precoding can also be called *SNR-maximizer* precoding. Again, the idea behind this scheme is that each channel vector of M selected users is almost in the direction of one of the basis vectors; hence their channel vectors are near-orthogonal to each other. This orthogonality increases with increased number of users in the network. In the best case when the selected channels compose an orthogonal set, the following rate is achievable:

$$R_{CAP} = \max_{P_i} \sum_{i=1}^M \log \left(1 + P_i |\mathbf{h}_i|^2 \right) \text{ s.t. } \sum_i P_i \leq P \quad (18)$$

The set of M power loading values P_i 's are selected based on waterpouring as $P_i = [v - |\mathbf{h}_i|^2]_+$. Obviously, other users have zero power.

We can therefore, distinguish between four different user selection and precoding strategies: PUS-OP, PUS-CAP, NUS-OP, NUS-CAP. The features of each of these schemes are summarized in **Table 1**. Regarding fairness, as PUS selects users based on the maximization of projection of their channel vectors over orthonormal bases, it is more likely that users with strong power (*i.e.* users near base station) are finally selected. Therefore, the issue of fairness is not considered in PUS while it is not the case in NUS scheme. We will elaborate the performance and specifications of each scheme more in Section 7 when we present numerical results.

5. Asymptotic Performance of Zero-Forcing Schemes

In this section, we investigate the asymptotic performance of NUS user selection schemes. Since the analysis of PUS scheme is also very similar to CAP, for the sake of brevity we do not discuss it here.

For any two arbitrary vectors \mathbf{a} and \mathbf{b} , we define the following orthogonality measure:

$$\varepsilon(\mathbf{a}, \mathbf{b}) \triangleq 1 - |\mathbf{a}^H \mathbf{b}|^2 / |\mathbf{a}|^2 |\mathbf{b}|^2 \quad (19)$$

Clearly, as \mathbf{a} tends to \mathbf{b} , $\varepsilon(\mathbf{a}, \mathbf{b})$ tends to zero. For the best user channel \mathbf{h}_i in the direction of basis vector \mathbf{u}_i , we have:

$$\varepsilon_i = \varepsilon(\mathbf{h}_i, \mathbf{u}_i) = \max_j \varepsilon(\mathbf{h}_j, \mathbf{u}_i) = 1 - \max_j |\mathbf{h}_j \mathbf{u}_i^H|^2 / |\mathbf{h}_j|^2 \quad (20)$$

Lemma 1: For large number of users, the cumulative distribution function (cdf) function of ε_i ($i : 1..M$) is linearly increased with n , *i.e.*, $\forall i : F_{\varepsilon_i}(x) = \Pr(\varepsilon_i < x) \sim O(nx)$.

Proof: Since \mathbf{U} is an orthonormal set, the random vari-

ables $|\mathbf{h}_j \mathbf{u}_i^H|^2$ are i.i.d. over i and j with $\chi^2(2)$ -distribution. As $|\mathbf{h}_j|^2$ is the sum of square of M Gaussian random variables, it is $\chi^2(2M)$ -distributed. It follows that the probability density function (pdf) and cdf of $t_{ij} = |\mathbf{h}_j \mathbf{u}_i^H|^2 / |\mathbf{h}_j|^2$ are, respectively,

$$f_t(x) = \int_0^\infty y e^{-yx} \frac{y^{M-1} e^{-y}}{(M-1)!} dy = M(1+x)^{-(M+1)} \quad (21)$$

and

$$F_t(x) = 1 - (1+x)^{-M} \quad (22)$$

Hence, the cdf of ε_i is

$$F_{\varepsilon_i}(x) = P(\varepsilon_i < x) = P(\max_j t_{ij} > 1-x) = 1 - F_t^n(1-x) = 1 - \left[1 - (2-x)^{-M} \right]^n \quad (23)$$

For small x , the right hand side of (22) can be approximated as

$$F_{\varepsilon_i}(x) \approx 1 - \left[1 - 2^{-M} (1 + 2^{1-M} Mx) \right]^n \approx 2^{1-2M} Mnx \sim O(nx) \quad (24)$$

for any ε_i ($i : 1..M$).

Lemma 1 indicates that, the probability of ε_i smaller than a specific small value, increases linearly with the user population, n , and the best users' channel vectors become more and more orthogonal to the basis directions (vectors). One can also say that by increasing the number of users, it is more probable that there would be M users with each channel vector very close to a basis. The following Lemma sheds some light on this fact.

Lemma 2: For large number of users, the probability that a user has the largest channel vector projection among all

Table 1. Features of user selection precoding strategies.

Scheme	channel knowledge of all users at TX	channel knowledge of selected users at TX	Power allocation and waterpouring	Fairness	Interference between users
PUS-OP	No	No	No	No	No
PUS-CAP	No	Yes	Yes	No	Yes
NUS-OP	No	No	No	Yes	No
NUS-CAP	No	Yes	Yes	Yes	Yes

users on two or more directions is linearly proportional to $1/n$.

Proof: From Lemma 1, the squared projections, $|\mathbf{h}_j \mathbf{u}_i^H|^2$, are i.i.d $\chi^2(2)$ -distributed variables over i and j . Among them, let Y_n and X_{M-1} denote the *largest* of the whole population of n users, and the *second largest* of the selected set of M users, respectively. A user has the best channel for two basis vectors if its *second largest* squared projection is larger than the squared projections of all other users for this specific basis vector. The probability of this event is:

$$\begin{aligned} & P(X_{M-1} > Y_n) \\ &= 1 - \int_0^\infty (1 - F_X(x)) F_X^{M-1}(x) n f_X(x) F_X^{n-1}(x) dx \end{aligned} \quad (25)$$

where $F_X(x) = 1 - e^{-x}$ and $f_X(x) = e^{-x}$ are cdf and pdf of a $\chi^2(2)$ random variable. Therefore,

$$\begin{aligned} & P(X_{M-1} > Y_n) \\ &= 1 - \int_0^\infty n(1 - e^{-x})^{n+M-2} e^{-2x} dx \\ &\approx \left[1 + \frac{n}{M} \right]^{-1} \end{aligned} \quad (26)$$

The right hand side of (23), behaves like $O(1/n)$ when n is larger than M .

The following Lemma guarantees that the proposed NUS-CAP method performs always better than opportunistic zero-forcing precoding (OP) in which M orthonormal vectors are selected randomly at the transmitter and the best users are selected for those random vectors. Transmitter then sends the information to the best M users on those random orthonormal vectors.

Lemma 3: The proposed NUS-CAP zero-forcing method always performs better than NUS-OP zero-forcing precoding.

Proof: Consider ε_i ($i : 1..M$) in (20). Assuming equal power allocation, from (16) the average sum rate for the defined NUS-OP zero-forcing precoding can be written as:

$$\begin{aligned} R_{OP} &= E\left\{\sum_{i=1}^M \left(1 + \frac{P}{M} \frac{|\mathbf{h}_i \mathbf{u}_i^H|^2}{1 + \sum_{j=1, j \neq i}^M |\mathbf{h}_i \mathbf{u}_j^H|^2}\right)\right\} \\ &= E\left\{\sum_{i=1}^M \left(1 + \frac{P}{M} \frac{1 - \varepsilon_i}{1 + \varepsilon_i}\right)\right\} \end{aligned} \quad (27)$$

The last equality results from the fact that \mathbf{u}_i 's are orthonormal. Our proposed NUS-CAP scheme, however, is

based on the distance of the best users' channel vectors, *i.e.* $|\mathbf{h}_j \mathbf{u}_i^H|$ ($i = 1..M; j = 1..M$) and can be found in terms of ε_i ($i = 1..M$). It is easy to check that on average:

$$|\mathbf{h}_i \mathbf{h}_j|^2 \leq M^{-1} \varepsilon_i (1 - \varepsilon_i) \quad i = 1..M \quad (28)$$

Therefore, considering equal power allocation (no waterpouring), the average sum rate of NUS-CAP scheme is:

$$\begin{aligned} R_{CAP} &= E\left\{\sum_{i=1}^M \left(1 + \frac{P}{M} \frac{|\mathbf{h}_i \mathbf{h}_i^H|^2}{1 + \sum_{j=1, j \neq i}^M |\mathbf{h}_j \mathbf{h}_i^H|^2}\right)\right\} \\ &= E\left\{\sum_{i=1}^M \left(1 + \frac{P}{M} \frac{1}{1 + \varepsilon_i (1 - \varepsilon_i)}\right)\right\} \end{aligned} \quad (29)$$

From (27) and (29), as $1 - \varepsilon_i / 1 + \varepsilon_i < [1 + \varepsilon_i (1 - \varepsilon_i)]^{-1}$, the average sum rate provided by the NUS-CAP scheme is always greater than that of NUS-OP zero-forcing.

Note that in the proof of Lemma 3, we considered equal power allocation on basis vectors, while if we apply the waterpouring in (18), NUS-CAP scheme performs much better than random zero-forcing scheme. Now, we prove that our scheme is optimum for large number of users.

Theorem 1: In the limit of large number of users, the NUS-CAP partial knowledge zero-forcing precoding approach proposed in Section 4 can achieve an average sum rate equal to that of DPC strategy in (6), *i.e.*,

$$E\{R_P\} \sim M \log \left(1 + \frac{P}{M} \log n\right) \quad (30)$$

Proof: Based on Lemma 3, it is sufficient to show that the discussed NUS-OP zero-forcing precoding scheme can achieve the rate in (30). As the proposed NUS-CAP precoding scheme always outperforms NUS-OP zero-forcing scheme, it turns out that it is capable of achieving the rate in (30) in the limit of large n .

Based on the results from extreme value theory in [10],

$$\max_i t_{ij} = \max_i |\mathbf{h}_j \mathbf{u}_i^H|^2 |\mathbf{h}_j|^2$$

behaves like $\log n + O(\log \log n)$. Therefore, from (19) and (26), the average sum rate of the NUS-OP zero-forcing precoding for large n

$$\begin{aligned} & \lim_{n \rightarrow \infty} R_{rand} \\ &= \lim_{n \rightarrow \infty} E\left\{\sum_{i=1}^M \left(1 + \frac{P}{M} \left(\frac{1}{\varepsilon_i} - 1\right)\right)\right\} \sim M \log \frac{P}{M} (1 + \log n) \end{aligned} \quad (31)$$

As the NUS-CAP scheme always outperforms NUS-OP zero-forcing precoding, from Lemma 3, it can achieve the sum rate in (31).

Note that in the proof of Theorem 1, we assumed that the available power is distributed equally among M precoder vectors (sometimes called directions). In practice, as mentioned in Section 4, we apply a standard water-pouring to calculate the optimum power allocation. However, as the proof shows, using equal powers on all precoder vectors is also asymptotically optimal (when the number of users is large).

6. Feedback Load, Complexity and Design Issues

In this section, we briefly discuss the practical considerations and design issues for our proposed schemes. We show that the proposed schemes can be applied with very low amount of feedback load and transmitter complexity as compared to that of full channel knowledge schemes. At the end of this section we briefly point out the selection of basis vectors set \mathbf{U} .

6.1. Feedback Load

Consider a BC system using a transmitter equipped with M antennas to serve n single-antenna active users. For schemes based on full channel knowledge such as DPC, optimal zero-forcing described in Section 3, and the zero-forcing scheme proposed in [5], at any transmission period, $2Mn$ real values should be fed back to the transmitter. On the other hand, the proposed CAP scheme based on partial knowledge scheme initially needs only n integer indices of the best basis vectors and n real best projections of the users' channel vectors and only $2M^2$ for the best M users channel vectors after initialization. Therefore, in total, it needs $n + 2M^2$ real and n integer values. Clearly, the proposed scheme requires much lower feedback load, especially at large number of users.

It is also possible to exclude bad users (those are not near-orthogonal to any of the basis vectors) from channel information feedback. Similar to the idea in [10], we can define a threshold value and if a user has its maximum channel vector projection below this threshold, it is considered as a *bad* user.

Lemma 4: At the limit of large number of users, for each of the basis vectors, there exists at least one user j with its channel vector \mathbf{h}_j satisfying

$$\max_{i=1..M} \left| \mathbf{h}_j \mathbf{u}_i^H \right|^2 > \log n .$$

Proof. Based on Lemma 2, for large number of users, n , it is not likely that a particular user can be the best one for two or more basis vectors. Hence, maximization over a set of users that have the maximum projection on a specific basis vector is equivalent to maximization over all users for that particular basis vector, *i.e.*,

$$\max_{j=1..n} \left| \mathbf{h}_j \mathbf{u}_i^H \right|^2 = \max_{j \in S(i)} \left| \mathbf{h}_j \mathbf{u}_i^H \right|^2 \quad i = 1 .. M \quad (32)$$

As $\left| \mathbf{h}_j \mathbf{u}_i^H \right|^2$ are $\chi^2(2)$ -distributed random variables, by applying the extreme value theory [13], for any i , $\max_{j \in S(i)} \left| \mathbf{h}_j \mathbf{u}_i^H \right|^2$ and its average behave like $\log n$ at large n . Therefore, there is always at least one user with channel satisfying

$$\max_{j \in S(i)} \left| \mathbf{h}_j \mathbf{u}_i^H \right|^2 > \log n = \lambda_{\text{tr}} \quad i = 1 .. M \quad (33)$$

for all basis vectors.

Lemma 4 indicates that a good threshold value is $\lambda_{\text{tr}} = \log n$. Using this threshold, the feedback load can be expressed as $n \Pr\{\lambda_i > \lambda_{\text{tr}}\}$ ($i : 1 .. M$) real and integer numbers. Note that, in practice $\Pr\{\lambda_i > \lambda_{\text{tr}}\}$ is much smaller than 1 and therefore, inserting this threshold value decrease feedback overhead dramatically.

6.2. Complexity

The proposed NUS-CAP scheme which is the most complex amongst four schemes just needs n comparisons at the transmitter and M projections and M comparisons at each user while DPC, optimal and suboptimal zero-forcing schemes are much more complex. For example, the zero-forcing algorithm proposed in [15], needs almost $\sum_{i=1}^M (n-1)(i+1) > nM(M+1)/2$ projections and nM comparisons at the transmitter to find a near-orthogonal channel vector set. In other words, the proposed NUS-CAP method has a much lower complexity than the other optimal and sub-optimal schemes, and also distributes low processing among users in the system.

6.3. Selection of Basis

As mentioned at the beginning of Section 4, the set of orthonormal basis \mathbf{U} , can be generally any arbitrary orthonormal basis in \mathbb{C}^M and as shown in Section 5, the proposed schemes can still work fine. However, orthonormal basis can be optimized by a very minor increase in the feedback load from active users to the transmitter.

Let assume that instead of just one predetermined orthonormal basis, we have a number of k bases, $\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^k$. Both the transmitter and users know these bases. Each user calculates its channel vector projections onto each of these bases vectors and reports the best value along with the indices of this basis and its own basis. This requires one real and two integer numbers (instead of one real and one integer). From the reported information, the transmitter selects the best orthonormal basis, corresponding to the best M users with the best orthogonality measures. As an illustrative example, consider $M = 2$ and two orthonormal bases to select. If we select the

first one as standard basis for \mathbb{C}^2 , i.e., $\mathbf{U}^1 = \{(1,0), (0,1)\}$, the other one should be the one whose vectors are as much as non-orthogonal to the standard basis, i.e., $\mathbf{U}^2 = \{1/\sqrt{2}(1,1), 1/\sqrt{2}(1,-1)\}$. Similarly, if we want to select three orthonormal bases, a reasonable choice is $\mathbf{U}^1 = \{(1,0), (0,1)\}$, $\mathbf{U}^2 = \{(a,b), (b,-a)\}$ and $\mathbf{U}^3 = \{(b,a), (a,-b)\}$ where $a=0.5, b=\sqrt{3}/4$.

The same approach can be applied for larger M and larger number of orthonormal bases. Increasing the number of orthonormal bases can greatly improve the performance of the proposed schemes at the cost of a slight increase in the feedback overhead. More precisely, using k different orthonormal bases, the probability in (24) is linearly increasing with almost nk (instead of n) and the probability in (26) is linearly proportional to $1/(nk)$ (instead of $1/n$). It means faster orthogonality rate by increasing the number of users and hence, the proposed schemes become optimal at lower number of users.

7. Numerical Results

In this section we examine the ability of zero-forcing precoder schemes in term of ergodic capacity by means of simulation. We also compare the complexity and feedback overhead of the proposed precoding schemes to that of DPC and full knowledge based zero-forcing precodings.

We consider a four antenna transmitter ($M = 4$), for all simulations. Besides, for our schemes we just consider one basis vector set. First, let compare four different schemes in two different scenarios, i.e. in low and high average users' channel power. **Figure 1** shows the sum rates of different user selection and precoding schemes in the low and high SNR regimes. By SNR regimes we mean the channel SNRs averaged over all available users' channels. As shown, while PUS selection strategy works better at low SNRs, NUS strategy outperforms it at high SNR regime. Therefore, one can conclude that when the average users' SNR is high, it is more reasonable to use NUS selection scheme while in low average SNR, PUS is more beneficial. Also, consider that at very fast fading

environment, it would be more desirable to use OP rather than CAP schemes. Based on this discussion, the diagram in **Figure 2**, gives us an idea how to select between different schemes. Note that because of similarity of structure and algorithms, it is possible to switch between them in a real time system. In other words, when fading is very fast the system uses OP schemes and in slow fading environment switches into CAP schemes. On the other hand, in high quality channel, it works in NUS regime while in poor-conditioned channels it switches into PUS regimes where users' channel power becomes a very important and significant factor.

Next, we simulate the feedback load, complexity and performance of our scheme and compare it with that of optimum and suboptimum schemes. In **Figure 3**, the feedback load of the schemes for different numbers of users (n) is illustrated. As shown, the CAP scheme needs much lower amount of feedback than the full channel knowledge based scheme and this difference in feedback overhead increases with the number of users in the system. At large n , the CAP scheme needs the same amount of channel feedback as the scheme in [15].

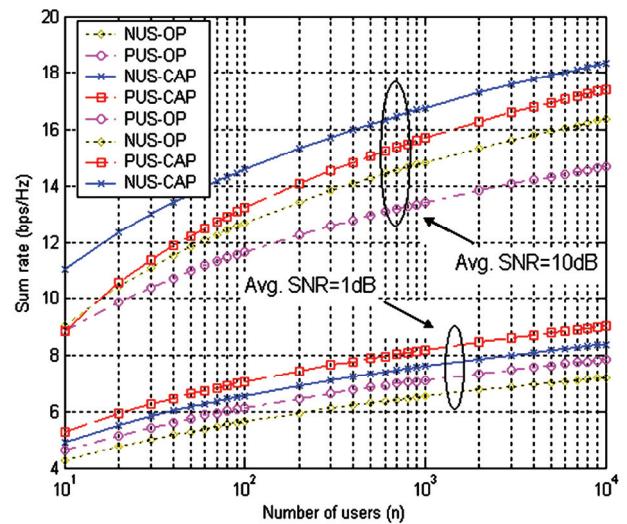


Figure 1. Comparison of sum rates of different precoder schemes in different SNR regimes.

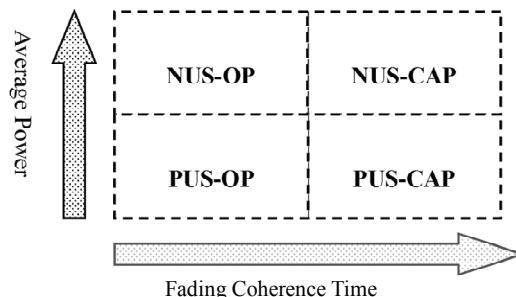


Figure 2. Selection of different schemes in various scenarios.

Figure 4 shows the complexity of different schemes in term of the number of CPU operation needed to finish the optimization tasks. Complexity of NUS-CAP scheme grows linearly with the number of users while the complexity of the optimal zero-forcing and DPC is exponentially increasing with the number of users. Besides, NUS-CAP scheme needs less CPU clock than the full channel knowledge based zero-forcing in [15].

In **Figure 5**, achievable sum rates of different precoding schemes are compared for different number of users. We assume a fixed available power of $P = 10$ dB at the transmitter. Different schemes are considered, including random user selection (M users are selected randomly each time), scheduling zero-forcing discussed in [15], random zero-forcing (OP) scheme discussed in Section 5, proposed NUS-CAP scheme, full channel zero-forcing

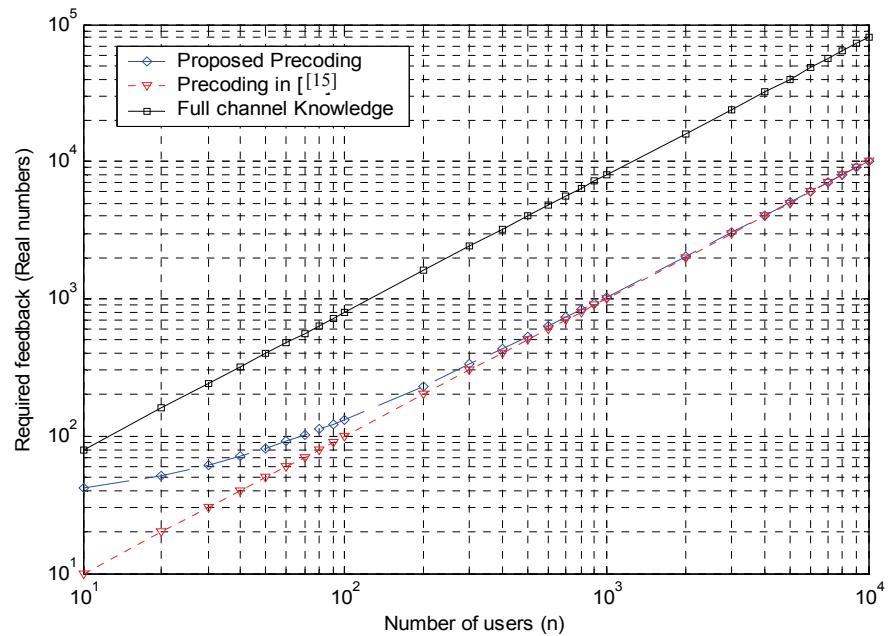


Figure 3. Feedback overhead comparison of different schemes with $M = 4$.

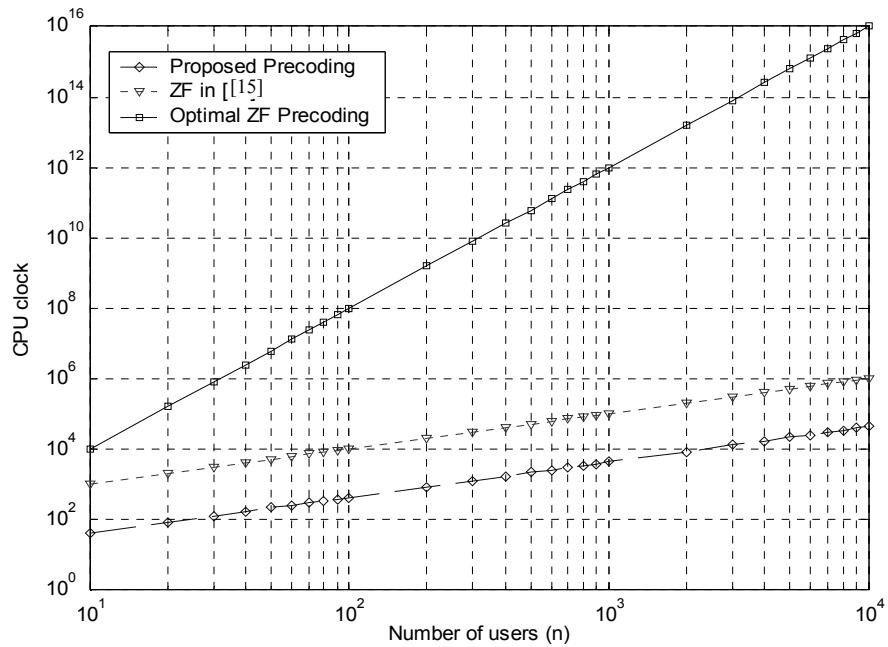


Figure 4. Complexity comparison of different schemes with $M = 4$.

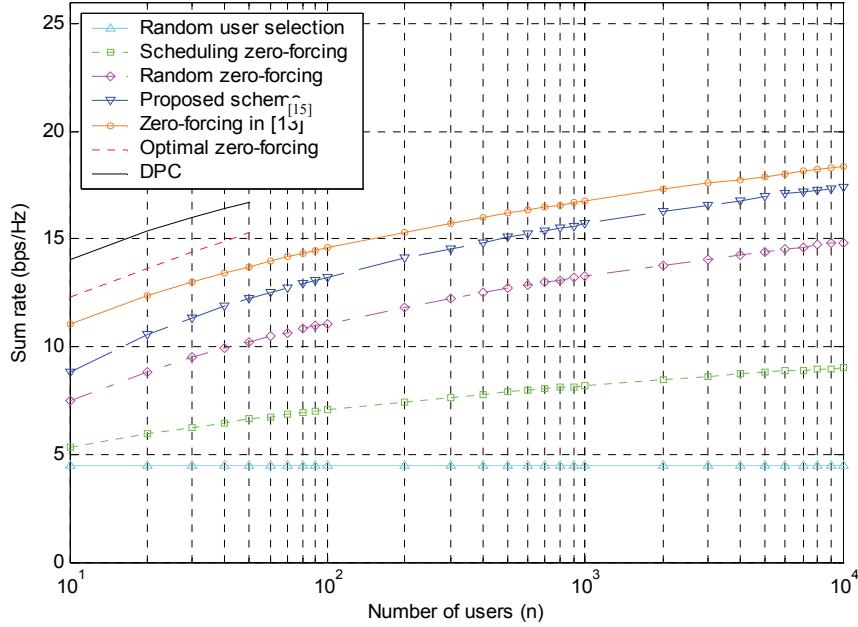


Figure 5. Sum rates for different schemes with $M = 4$ and $P = 10$ dB.

scheme in [15], optimal zero forcing described in Section 3 and dirty paper coding. Due to high level of complexity, the sum rate of DPC and optimal zero-forcing has been just depicted for small number of users. As shown, NUS-CAP outperforms the schemes based on partial channel knowledge. Furthermore, its achievable sum rate is close to that of full channel knowledge based schemes and the rate of growth is the same as that of optimal schemes at the limit of large number of users. Note that, NUS-CAP works better in high number of users; however, by increasing the number of basis sets it is possible to compensate this drawback in low number of users. As mentioned we have just considered one basis vector set for the proposed schemes. By increasing the number of basis sets, the proposed schemes perform better even when the number of users is small.

8. Concluding Remarks

In this paper, we investigated the design of a partial knowledge based precoding scheme for broadcast MIMO systems. By partial knowledge we mean that the transmitter does not have the instantaneous channel realizations of all active users in the network. This assumption arises because the fast changing nature of wireless channels makes full channel knowledge difficult to obtain at the transmitter.

We proposed various combined user selection and pre-coding schemes with no or small amount of channel feedback and low complexity, which can achieve the performance of full knowledge schemes in the limit of large number of users. More precisely, it was shown that the ergodic capacity offered by the proposed schemes can

achieve the same growing rate with the number of users ($\log \log n$) as that of DPC and optimal zero-forcing pre-coding.

The ability of the proposed schemes in achieving optimality in term of ergodic capacity growth rate was demonstrated by analysis and numerical results. It was shown that as the number of users increases, the proposed schemes need lower amount of feedback and processing load than other schemes. Moreover, the proposed schemes are superior to each other in different propagation scenarios and SNR regimes. Nevertheless, as their structures are very similar, it is possible for the base station to switch between them in order to adapt to the propagation environment. This feature gives flexibility in system configuration as well as more robustness in compensating for channel changes.

9. References

- [1] E. Telatar, "Capacity of Multi-Antenna Gaussian Channel," *European Transactions on Telecommunications*, Vol. 10, November 1999, pp. 585-595.
- [2] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas," *Wireless Personal Communication*, Vol. 6, March 1998, pp. 311-335.
- [3] Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis and H. Sampath, "Optimal Design for Space-Time Linear Precoders and Decoders," *IEEE Transactions on Signal Processing*, Vol. 50, No. 5, May 2002, pp. 1051-1063.
- [4] J. Akhtar and D. Gesbert, "Spatial Multiplexing over

- Correlated MIMO Channels with a Closed-Form Precoder," *IEEE Transactions on Wireless Communications*, No. 4, Vol. 5, September 2005, pp. 2400-2409.
- [5] H. R. Bahrami and T. Le-Ngoc, "Precoder Design Based on Correlation Matrices for MIMO Systems," *IEEE Transactions on Wireless Communications*, Vol. 5, No. 12, December 2006, pp. 3579-3587.
- [6] G. Caire and S. Shamai (Shitz), "On the Achievable throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Transactions on Information Theory*, Vol. 49, No. 7, July 2003, pp. 1691-1706.
- [7] M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory*, Vol. 29, No. 3, May 1983, pp. 439-441.
- [8] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, Achievable Rates and Sum Rate Capacity of Gaussian MIMO Broadcast Channel," *IEEE Transactions on Information Theory*, Vol. 49, No. 10, October 2003, pp. 2658-2668.
- [9] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Downlink-Uplink Duality," *IEEE Transactions on Information Theory*, Vol. 49, No. 8, August 2003, pp. 1912-1921.
- [10] W. Yu and J. M. Cioffi, "Sum Capacity of Vector Gaussian Broadcast Channels, *IEEE Transactions on Information Theory*, Vol. 50, No. 9, September 2004, pp. 1875-1892.
- [11] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 5, June 2003, pp. 684-702.
- [12] G. Caire, N. Jindal, M. Kobayashi and N. Ravindran, "Achievable Rates with Simple Channel State Estimation and Feedback Schemes," Submitted to *IEEE Transactions on Information Theory*, November 2007.
- [13] M. Sharif and B. Hassibi, "On the Capacity of MIMO Broadcast Channels with Partial Side Information," *IEEE Transactions on Information Theory*, Vol. 51, No. 2, February 2005, pp. 506-522.
- [14] M. Sharif and B. Hassibi, "On the Capacity of MIMO Broadcast Channel with Partial Side Information," *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Vol. 1, November 2003, pp. 958-962.
- [15] T. Yoo and A. Goldsmith, "On the Optimality of Multi-antenna Broadcast Scheduling Using Zero-Forcing Beam-Forming," *IEEE Journal on Selected Areas in Communication*, Vol. 24, No. 3, March 2006, pp. 528-541.
- [16] T. Yoo, N. Jindal and A. Goldsmith, "Multi-Antenna Broadcast Channels with Limited Feedback and User Selection," *IEEE Journal on Selected Areas in Communication*, Vol. 25, No. 6, August 2007, pp. 1069-1078.
- [17] K. Huang, J. G. Andrews and R. W. Heath, "Performance of Orthogonal Beamforming for SDMA with Limited Feedback," *IEEE Transactions on Vehicular Technology*, January 2009, pp. 152-164.