Coherent Product Superposition for Downlink Multiuser MIMO

Yang Li, *Student Member, IEEE,* and Aria Nosratinia, *Fellow, IEEE*

Abstract

In a two-user broadcast channel where one user has full CSIR and the other has none, a recent result showed that TDMA is strictly suboptimal and a product superposition requiring non-coherent signaling achieves DoF gains under many antenna configurations. This work introduces product superposition in the domain of coherent signaling with pilots, demonstrates the advantages of product superposition in low-SNR as well as high-SNR, and establishes DoF gains in a wider set of receiver antenna configurations. Two classes of decoders, with and without interference cancellation, are studied. Achievable rates are established by analysis and illustrated by simulations.

Index Terms

CSIR, superposition, degrees of freedom, pilot, channel estimation

I. Introduction

Due to varying mobility and the effects of the propagation environment, wireless network nodes often have unequal capability to acquire CSIR (channel state information at receiver). Downlink (broadcast) transmission to nodes with unequal CSIR is therefore a subject of practical interest.

It has been known that if all downlink users have full CSIR, then orthogonal transmission (e.g. TDMA) achieves the optimal degrees of freedom (DoF) [1], [2], in the absence of CSIT under fast fading. A similar result is known to hold for certain antenna configurations in the absence of CSIR. Recently it was discovered [3] that a very different behavior emerges when one user has perfect CSIR and the other has none: in this case TDMA is highly suboptimal and a product superposition can achieve gains in the degrees of freedom (DoF). However, this result [3] required non-coherent Grassmannian signaling while most practical systems use pilots and employ coherent detection after channel estimation. In addition, the result [3] was limited to high-SNR and did not demonstrate optimality in all receiver antenna configurations.

In this paper we extend the product superposition to coherent signaling with pilots. This is motivated by several factors, among them the popularity and prevalence of coherent signaling in the practice of wireless communications, as well as the known results in the point-to-point channel [4] showing that pilot-based transmission can perform

The authors are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA, email: aria@utdallas.edu, yang@utdallas.edu

almost as well as Grassmannian signaling. We show that a similar result holds in the mixed-mobility broadcast channel. In the process, we demonstrate the DoF gains of product superposition for more antenna configurations than in [3], and in addition show that it has excellent performance in low-SNR as well as high-SNR.

A downlink scenario with two users is considered in this paper, where one user has a short coherence interval and is referred to as the *dynamic user*, and the other has a long coherence interval and is referred to as the *static user*. The main results of this paper are as follows.

- We propose a new signaling structure that is a product of two matrices representing the signals of the static and dynamic user, respectively, where the data for both users are transmitted using coherent signaling.
- We propose two decoding methods. The first method performs no interference cancellation at the receiver. We show that under this method, at both high SNR and low SNR, the dynamic user experiences almost no degradation due to the transmission of the static user. Therefore in the sense of the cost to the other user, the static user's rate is added to the system "for free." Avoiding interference cancellation gives this method the advantage of simplicity.
- The second method further improves the static user's rate by allowing it to decode and remove the dynamic user's signal. This increases the effective SNR for the static user and provides further rate gain.
- We show that the product superposition has DoF gains when the dynamic user has either more, less or equal number of antennas as the static user. Previously [3] the DoF gain was demonstrated only when the dynamic user had fewer or equal number of antennas compared with the static user.

The following notation is used throughout the paper: for a matrix $\mathbf{A}$, the transpose is denoted with $\mathbf{A}^t$, the conjugate transpose with $\mathbf{A}^H$, the pseudo inverse with $\mathbf{A}^\dagger$ and the element in row i and column j with $[\mathbf{A}]_{ij}$. The $k \times k$ identity matrix is denoted with $\mathbf{I}_k$. The set of $n \times m$ complex matrices is denoted with $\mathcal{C}^{n\times m}$. We denote $\mathcal{CN}(0,1)$ as the circularly symmetric complex Gaussian distribution with zero mean and unit variance. For all variables the subscripts "s" and "d" stand as mnemonics for "static" and "dynamic", respectively, and subscripts "τ" and "δ" stand for "training" and "data."

II. System Model and Preliminaries

We consider an M-antenna base-station transmitting to two users, where the dynamic user has N_d antennas and the static user has N_s antennas. The channel coefficient matrices of the two users are $\mathbf{H}_d \in \mathcal{C}^{N_d\times M}$ and $\mathbf{H}_s \in \mathcal{C}^{N_s\times M}$, respectively. In this paper we restrict our attention to $M = \max\{N_d, N_s\}$. The system operates under block-fading, where $\mathbf{H}_d$ and $\mathbf{H}_s$ remain constant for T_d and T_s symbols, respectively, and change independently across blocks. The coherence time T_d is small but T_s is large ($T_s \gg T_d$) due to different mobilities. The difference in coherence times means that the channel resources required by the static user to estimate its channel are negligible compared to the training requirements of the dynamic user. To reflect this in the model, it is assumed that $\mathbf{H}_s$ is known by the static user (but unknown by the dynamic user, naturally), while $\mathbf{H}_d$ is not known *a priori* by either user.

Over T_d time-slots (symbols) the base-station sends $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_M]^t$ across M antennas, where $\mathbf{x}_i \in \mathcal{C}^{T_d\times 1}$

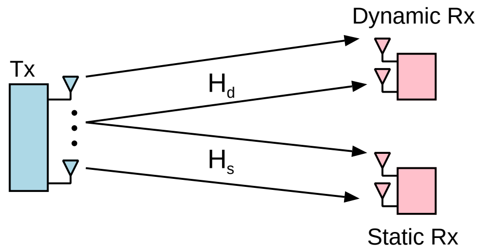

Fig. 1. Channel model.

is the signal vector sent by the antenna i. The signal at the dynamic and static users is respectively

$$\mathbf{Y}_d = \mathbf{H}_d\mathbf{X} + \mathbf{W}_d,$$
$$\mathbf{Y}_s = \mathbf{H}_s\mathbf{X} + \mathbf{W}_s, \tag{1}$$

where $\mathbf{W}_d \in \mathcal{C}^{N_d \times T_d}$ and $\mathbf{W}_s \in \mathcal{C}^{N_s \times T_d}$ are additive noise with i.i.d. entries $\mathcal{CN}(0,1)$. Each row of $\mathbf{Y}_d \in \mathcal{C}^{N_d \times T_d}$ (or $\mathbf{Y}_s \in \mathcal{C}^{N_s \times T_d}$) corresponds to the received signal at an antenna of the dynamic user (or the static user) over T_d time-slots. The base-station is assumed to have an average power constraint ρ

$$\mathbb{E}\big[\sum_{i=1}^{M} \mathrm{tr}(\mathbf{x}_i\mathbf{x}_i^H)\big] = \rho\, T_d. \tag{2}$$

The channels $\mathbf{H}_d$ and $\mathbf{H}_s$ have i.i.d. entries with the distribution $\mathcal{CN}(0,1)$. We assume $M = \max(N_d, N_s)$ and $T_d \geq 2N_d$ [4].

A. The Baseline Scheme

We start by establishing a baseline scheme and outlining its capacity for the purposes of comparison. In our system model, MIMO transmission schemes involving dirty paper coding, zero-forcing, or similar techniques [5]–[8] are not applicable since $\mathbf{H}_d$ varies too quickly for feedback to transmitter. Our baseline method uses orthogonal transmission, i.e., TDMA.

For the dynamic user, we consider the following near-optimal method. The base-station activates only N_d out of M antennas [4], sends an orthogonal pilot matrix $\mathbf{S}_\tau \in \mathcal{C}^{N_d \times N_d}$ during the first N_d time-slots, and then sends i.i.d. $\mathcal{CN}(0,1)$ data signal $\mathbf{S}_\delta \in \mathcal{C}^{N_d \times (T_d - N_d)}$ in the following $T_d - N_d$ time-slots [9], that is

$$\mathbf{X} = \left[\sqrt{\frac{\rho_\tau}{N_d}}\,\mathbf{S}_\tau \ \sqrt{\frac{\rho_\delta}{N_d}}\,\mathbf{S}_\delta\right] \tag{3}$$

where $\mathbf{S}_\tau\mathbf{S}_\tau^H = N_d\mathbf{I}$, and ρ_τ and ρ_δ are the average power used for training and data, respectively, and satisfy the power constraint in (2):

$$\rho_\tau N_d + \rho_\delta(T_d - N_d) \leq \rho T_d. \tag{4}$$

The dynamic user employs a linear minimum-mean-square-error (MMSE) estimation on the channel. The normalized channel estimate obtained in this orthogonal scheme is denoted $\overline{\mathbf{H}}_d \in \mathcal{C}^{N_d \times N_d}$. Under this condition, the rate attained by the dynamic user is [9]:

$$R_d \geq (1 - \frac{N_d}{T_d})\mathbb{E}\big[\log\det(\mathbf{I}_{N_d} + \frac{\rho_d}{N_d}\overline{\mathbf{H}}_d\overline{\mathbf{H}}_d^H)\big], \tag{5}$$

where ρ_d is the effective signal-to-noise ratio (SNR)

$$\rho_d = \frac{\rho_\delta\, \rho_\tau}{1 + \rho_\delta + \rho_\tau N_d}. \tag{6}$$

For the static user, the channel is known at the receiver, the base-station sends data directly using all M antennas. The rate achieved by the static user is [10]

$$R_s = \mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_s} + \frac{\rho}{N_s}\,\mathbf{H}_s\mathbf{H}_s^H\Bigg)\Bigg]. \tag{7}$$

Time-sharing ($0 \leq p \leq 1$) between R_d and R_s yields the rate region

$$\mathcal{R}_{OT} = \big(pR_d,\, (1-p)R_s\big). \tag{8}$$

B. Overview of Product Superposition [3]

In [3], a product superposition based on Grassmannian signaling was proposed and shown to achieve significant gain in DoF over orthogonal transmission. In the so-called *Grassmannian-Euclidean superposition* [3], the base-station transmits

$$\mathbf{X} = \mathbf{X}_s\mathbf{X}_d \in \mathcal{C}^{M\times T_d} \tag{9}$$

over T_d time-slots, where $\mathbf{X}_d \in \mathcal{C}^{N_d\times T_d}$ and $\mathbf{X}_s \in \mathcal{C}^{M\times N_d}$ are the signals for the dynamic and static user, respectively. For the dynamic user, a Grassmannian (unitary) signal is used to construct $\mathbf{X}_d$, so that information is carried only in the subspace spanned by the rows of $\mathbf{X}_d$. As long as $\mathbf{X}_s$ is full rank, its multiplication does not create interference for the dynamic user, since $\mathbf{X}_s\mathbf{X}_d$ and $\mathbf{X}_d$ span the same row-space.

The static user decodes and peels off $\mathbf{X}_d$ from the received signal, then decodes $\mathbf{X}_s$, which carries information in the usual manner of space-time codes.

In conventional point-to-point non-coherent methods [4], [11], power gain is obtained at low-SNR and yet no DoF gain is achieved. Compared with these method, the product superposition attains DoF gain by transmitting to two users.

III. Pilot-Based Product Superposition

We now develop a product superposition with coherent signaling for the two-user broadcast channel. We start with a simple method with single-user decoding (no interference cancellation).

A. Signaling Structure

Over T_d symbols (the coherence interval of the dynamic user) the base-station sends $\mathbf{X} \in \mathcal{C}^{M\times T_d}$ across N_s antennas:

$$\mathbf{X} = \mathbf{X}_s\mathbf{X}_d, \tag{10}$$

where $\mathbf{X}_s \in \mathcal{C}^{M \times N_d}$ is the data matrix for the static user and has i.i.d. $\mathcal{CN}(0,1)$ entries. The signal matrix $\mathbf{X}_d \in \mathcal{C}^{N_d \times T_d}$ is intended for the dynamic user and consists of the data matrix $\mathbf{X}_\delta \in \mathcal{C}^{N_d \times (T_d - N_s)}$ whose entries are i.i.d. $\mathcal{CN}(0,1)$ and the pilot matrix $\mathbf{X}_\tau \in \mathcal{C}^{N_d \times N_s}$ which is *unitary*, and is known to both static and dynamic users.

$$\mathbf{X}_d = \Big[\sqrt{c_\tau}\, \mathbf{X}_\tau \;\; \sqrt{c_\delta}\, \mathbf{X}_\delta \Big], \tag{11}$$

where the constant c_τ and c_δ satisfy the power constraint (2):

$$N_s N_d \big(c_\tau + (T_d - N_d) c_\delta \big) \leq \rho T_d. \tag{12}$$

Please make note of the normalization of pilot and data matrices in the product superposition: The pilot matrix is unitary, i.e., the entire pilot power is normalized, while the data matrix is normalized per time per antenna. This is only for convenience of mathematical expressions in the sequel; full generality is maintained via multiplicative constants c_δ and c_τ.

A sketch of the ideas involved in the decoding at the dynamic and static users is as follows. The signal received at the dynamic user is

$$\mathbf{Y}_d = \mathbf{H}_d \mathbf{X}_s \Big[\sqrt{c_\tau} \mathbf{X}_\tau \;\; \sqrt{c_\delta} \mathbf{X}_\delta \Big] + \mathbf{W}_d \tag{13}$$

where $\mathbf{W}_d$ is the additive noise. The dynamic user uses the pilot matrix to estimate the equivalent channel $\mathbf{H}_d \mathbf{X}_s$, and then decodes $\mathbf{X}_\delta$ based on the channel estimate.

For the static user, the signal received during the first N_d time-slots is

$$\mathbf{Y}_{s1} = \sqrt{c_\tau}\, \mathbf{H}_s \mathbf{X}_s \mathbf{X}_\tau + \mathbf{W}_{s1} \tag{14}$$

where $\mathbf{W}_{s1}$ is the additive noise at the static user during the first N_d samples. The static user multiplies its received signal by $\mathbf{X}_\tau^H$ from the right and then recovers [1] the signal $\mathbf{X}_s$.

Remark 1: Each of the dynamic user's codewords includes pilots because it needs frequent channel estimates. No pilots are included in the individual codewords of the static user because it only needs infrequent channel estimate updates. In practice static user's channel training occurs at much longer intervals outside the proposed signaling structure.

B. Main Result

Theorem 1: Consider an M-antenna base-station, a dynamic user with N_d-antennas and coherence time T_d, and a static user with N_s-antennas and coherence time $T_s \gg T_d$. Assuming the dynamic user does not know its channel $\mathbf{H}_d$ but the static user knows its channel $\mathbf{H}_s$, the pilot-based product superposition achieves the rates

$$R_d = (1 - \frac{N_d}{T_d}) \mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_d} + \frac{\rho_d}{N_d} \overline{\mathbf{H}}_d \overline{\mathbf{H}}_d^H \bigg) \bigg], \tag{15}$$

[1] The rate is assumed to be smaller than the channel capacity, so the codeword (multiple blocks of $\mathbf{X}_s$) can be always decoded as long as it is sufficient long.

$$R_s = \frac{N_d}{T_d}\,\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_s} + \frac{\rho_s}{N_s}\,\mathbf{H}_s\mathbf{H}_s^H\Bigg)\Bigg], \tag{16}$$

where $\overline{\mathbf{H}}_d$ is the *normalized* MMSE channel estimate of the equivalent dynamic channel $\mathbf{H}_d\mathbf{X}_s$, and ρ_d and ρ_s are the effective SNRs:

$$\rho_d = \frac{c_\tau c_\delta N_d N_s^2}{1 + c_\tau N_s + c_\delta N_d N_s}, \tag{17}$$

$$\rho_s = c_\tau N_s. \tag{18}$$

Proof: See Appendix I. ■

For the static user, the effective SNR ρ_s increases linearly with the power used in the training of the dynamic user. This is because the static user decodes based on the signal received during the training phase of the dynamic user.

For the dynamic user, the effective SNR ρ_d is unaffected by superimposing $\mathbf{X}_s$ on $\mathbf{X}_d$. To see this, compare (4) with (12) to arrive at $\rho_\tau = c_\tau N_s$ and $\rho_\delta = c_\delta N_d N_s$, therefore the two SNRs are equal to

$$\rho_d = \frac{c_\tau c_\delta N_d N_s^2}{1 + c_\tau N_s + c_\delta N_d N_s}. \tag{19}$$

Intuitively, the rate available to the dynamic user via orthogonal transmission (Eq. (5)) and via superposition (Eq. (15)) will be very similar: the normalized channel estimate $\overline{\mathbf{H}}_d$ in both cases has uncorrelated entries with zero mean and unit variance.[2] Thus the product superposition achieves the static user's rate "for free" in the sense that the rate for the dynamic user is approximately the same as in the single-user scenario. In the following, we discuss this phenomenon at low and high SNR.

1) Low-SNR Regime: We have $\rho_d, \rho_s \ll 1$. Let the eigenvalues of $\overline{\mathbf{H}}_d\overline{\mathbf{H}}_d^H$ be denoted $\bar{\lambda}_{di}^2$, $i = 1, \ldots, N_d$. Using (15) and a Taylor expansion of the log function at low SNR, the achievable rate for the dynamic user is approximately:

$$R_d \approx (1 - \frac{N_d}{T_d})\frac{\rho_d}{N_d}\,\mathbb{E}\big[\sum_{i=1}^{N_d}\bar{\lambda}_{di}^2\big] \tag{20}$$

$$= (1 - \frac{N_d}{T_d})\frac{\rho_d}{N_d}\,\mathrm{tr}\big(\mathbb{E}[\overline{\mathbf{H}}_d\overline{\mathbf{H}}_d^H]\big) \tag{21}$$

$$= (1 - \frac{N_d}{T_d})N_d\,\rho_d. \tag{22}$$

where higher-order Taylor terms have been ignored. Similarly, from (5), the baseline method achieves the rate

$$(1 - \frac{N_d}{T_d})N_d\,\rho_d. \tag{23}$$

Thus, the dynamic user attains the same rate as it would in the absence of the other user and its interference, i.e., a single-user rate. At low SNR, one cannot exceed this performance.

[2]The dynamic channel estimates in the orthogonal and superposition transmissions have the same mean and variance but are not identically distributed, because in the orthogonal case, $\overline{\mathbf{H}}_d$ is an estimate of $\mathbf{H}_d$, a Gaussian matrix, while in the superposition case it is an estimate of $\mathbf{H}_d\mathbf{X}_s$, the product of two Gaussian matrices. Therefore the expectations in Eq. (5) and (15) may produce slightly different results.

The rate available to the static user at low-SNR is obtained via (16), as follows:

$$R_s \approx \frac{\rho_s}{T_d}\,\mathrm{tr}\big(\mathbb{E}[\mathbf{H}_s\mathbf{H}_s^H]\big) \tag{24}$$

$$= \frac{N_s^2\,\rho_s}{T_d}. \tag{25}$$

2) High-SNR Regime: We have $\rho_d, \rho_s \gg 1$, therefore from (15) the achievable rate for the dynamic user is

$$R_d \approx (1-\frac{N_d}{T_d})\bigg(N_d \log\frac{\rho_d}{N_d} + \mathbb{E}\big[\sum_{i=1}^{N_d}\log\bar{\lambda}_{di}^2\big]\bigg). \tag{26}$$

where the approximation follows from the dominance of the channel gain term in the $\log$ det capacity formula. The dynamic user attains $N_d(1-N_d/T_d)$ degrees of freedom, which is the maximum DoF even in the absence of the static user [4]. Superimposing $\mathbf{X}_s$ only affects the distribution of eigenvalues $\bar{\lambda}_{di}^2$, whose impact is negligible at high-SNR.

For the static user, let the eigenvalues of $\mathbf{H}_s\mathbf{H}_s^H$ be denoted λ_{si}^2, $i=1,\ldots,N_s$. From (16), we have

$$R_s \approx \frac{N_d}{T_d}\bigg(N_s \log\frac{\rho_s}{N_s} + \mathbb{E}\big[\sum_{i=1}^{N_s}\log\lambda_{si}^2\big]\bigg), \tag{27}$$

which implies that the static user achieves N_dN_s/T_d degrees of freedom. Thus, the pilot-based product superposition achieves the DoF obtained in [3] for $N_d \le N_s$, and also for $N_d > N_s$.

C. Power Allocation

The effective SNRs of the dynamic and static users depend on c_τ and c_δ. We focus on c_τ and c_δ that maximize R_d (equivalently ρ_d) in a manner similar to [9]. From (62) and (69),

$$\rho_d = \frac{c_\tau c_\delta N_d N_s^2}{1 + c_\tau N_s + c_\delta N_d N_s}. \tag{28}$$

From (12), we have $c_\tau = \rho T_d/(N_dN_s) - c_\delta(T_d - N_d)$. Substitute c_τ into (28):

$$\rho_d = \frac{N_dN_s(T_d-N_d)}{T_d-2N_d}\cdot\frac{c_\delta(a-c_\delta)}{-c_\delta+b}, \tag{29}$$

where

$$a = \frac{\rho T_d}{N_dN_s(T_d-N_d)}, \tag{30}$$

$$b = \frac{N_d+\rho T_d}{N_dN_s(T_d-2N_d)}. \tag{31}$$

Noting that $0 \le c_\delta \le a$, we obtain the value of c_δ that maximizes R_d:

$$c_\delta^* = b - \sqrt{b^2-ab}, \tag{32}$$

which corresponds to

$$\rho_d^* = \frac{N_dN_s(T_d-N_d)}{T_d-2N_d}\big(2b-a-2\sqrt{b^2-ab}\big), \tag{33}$$

$$\rho_s^* = \frac{\rho T_d}{N_d} - N_s(T_d-N_d)(b-\sqrt{b^2-ab}). \tag{34}$$

In the low-SNR regime ($\rho \ll 1$), we have $a \ll b$, where $b \approx \frac{N_d}{N_d N_s (T_d - 2N_d)}$, and use Taylor expansion:

$$\sqrt{b^2 - ab} \approx b\big(1 - \frac{a}{2b} - \frac{a^2}{8b^2}\big).$$

We obtain

$$\rho_d^* \approx \frac{\rho^2 T_d^2}{4N_d(T_d - N_d)} \tag{35}$$

$$\rho_s^* \approx \frac{\rho T_d}{2N_d}. \tag{36}$$

This indicates that the static user has a much larger effective SNR, i.e., $\rho_d^* = o(\rho_s^*)$. In this case, from (22) and (25), the achievable rate is

$$R_d \geq \frac{T_d}{4}\rho^2, \tag{37}$$

$$R_s \approx \frac{N_s}{2}\rho. \tag{38}$$

In the high-SNR regime where $\rho \gg 1$ we have

$$\rho_d^* \approx \frac{\rho\, T_d}{(\sqrt{T_d - N_d} - \sqrt{N_d})^2}, \tag{39}$$

$$\rho_s^* \approx \frac{\rho T_d(\sqrt{T_d/N_d - 1} - 1)}{T_d - 2N_d}. \tag{40}$$

Both static and dynamic users attain SNR that increases linearly with ρ. When $T_d \gg N_d$, for the static user, $\rho_s^* \approx \rho\sqrt{T_d/N_d} \gg \rho_d^*$. For the dynamic user, we have $\rho_d^* \approx \rho$, which is the same SNR as if the dynamic user had perfect CSI; this is not surprising since the power used for training is negligible when the channel is very steady.

Remark 2: In the MIMO broadcast channel, conventional transmission schemes essentially divide the power between users. In the proposed product superposition the transmit power works for both users simultaneously instead of being divided between them. The training power used for the dynamic user also carries the static user's data. In this way, significant gains over TDMA is achieved, which is contrary to the conventional methods that at low-SNR produce little or no gain relative to TDMA.

Remark 3: In [3], the product superposition was shown to attain the following DoF region when $N_d \leq N_s$, i.e., achieving the coherent outer bound [2]:

$$\frac{d_d}{N_d} + \frac{d_s}{N_s} \leq 1, \quad d_d \leq N_d(1 - \frac{N_d}{T_d})$$

where d_d and d_s are the DoF of the dynamic and static user, respectively. Note that the developments in this section make no assumption about the relative number of antennas at the dynamic and static receivers. One can verify that Equations (15) and (16) meet the bounds shown above for both $N_d \leq N_s$ and $N_d > N_s$. Therefore, the achievable DoF of the product superposition is now established for all dynamic/static user antenna configurations.

IV. Improving Rates by Interference Cancellation

So far no interference cancellation was performed, therefore the users did not need to decode each other's signal. However, this had the effect that the static user utilizes only the portion of transmit power corresponding to the dynamic user's pilot, and not the portion corresponding to the dynamic user's data. In this section we explore the possibility of the static user decoding the signal of the dynamic user.[3] To facilitate this, we concentrate on the case $N_s \geq N_d$. The received signal at the static user is

$$\mathbf{Y}_s = \mathbf{H}_s \mathbf{X}_s [\sqrt{c_\tau}\, \mathbf{X}_\tau \ \sqrt{c_\delta}\, \mathbf{X}_\delta] + \mathbf{W}_s \tag{41}$$

where $\mathbf{Y}_s \in \mathcal{C}^{N_s \times T_d}$. The static user first estimates the product $\mathbf{H}_s \mathbf{X}_s \in \mathcal{C}^{N_s \times N_d}$ by using the pilot $\mathbf{X}_\tau$ sent during the first N_d time-slots, and then it decodes $\mathbf{X}_\delta$. Now $\mathbf{X}_d$ is known, therefore the entire observed signal at the static user can be used to decode its message. If $\mathbf{X}_\delta$ is decoded successfully, the static user can use the power used by the dynamic user data, in addition to the power used by the dynamic user pilot. Intuitively, harvesting additional power would improve the static user's rate relative to Section III.

Assuming the codeword used by the dynamic user is sufficiently long, so that the static user also experiences many channel realizations over the dynamic user codewords. The rate gain produced by the interference decoding is characterized by the following theorem.

Theorem 2: Assuming $N_s \geq N_d$ and sufficiently long codeword of the dynamic user, with interference decoding and cancellation, the pilot-based product superposition achieves the following rate for the static user

$$R_s = \frac{N_d}{T_d} \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_s} + \frac{\rho_s}{N_s} \mathbf{H}_s \mathbf{H}_s^H \right) \right], \tag{42}$$

where the effective SNR is

$$\rho_s = \frac{N_s}{\mathbb{E}[\lambda_i^{-2}]} \tag{43}$$

with λ_i^2 being any of the unordered eigenvalues of $\mathbf{X}_d \mathbf{X}_d^H$.

Proof: See Appendix II. ∎

Compared with Theorem 1, the SNR for the static user is improved by using the entire $\mathbf{X}_d$. To see this, we decompose $\mathbf{X}_\delta = \mathbf{U}_\delta \,\text{diag}(\gamma_1, \cdots, \gamma_{N_d})\, \mathbf{V}_\delta^H$, and obtain

$$\mathbf{X}_d \mathbf{X}_d^H = c_\tau \mathbf{I}_{N_d} + c_\delta\, \mathbf{U}_\delta \,\text{diag}(\gamma_1^2, \cdots, \gamma_{N_d}^2)\, \mathbf{U}_\delta^H \tag{44}$$

$$= \mathbf{U}_\delta \,\text{diag}(c_\tau + c_\delta \gamma_1^2, \cdots, c_\tau + c_\delta \gamma_{N_d}^2)\, \mathbf{U}_\delta^H. \tag{45}$$

Therefore, $\lambda_i^2 = c_\tau + c_\delta\, \gamma_i^2$, for $i = 1, \ldots, N_d$, and

$$\rho_s = \frac{N_s}{\mathbb{E}[(c_\tau + c_\delta\, \gamma_1^2)^{-1}]}. \tag{46}$$

which is greater than the effective power available to the previous scheme (compare with Eq. (18)). So knowing the dynamic user's data always produces a power gain.

[3]It is not necessary for the *dynamic* user to decode the other user's signal, even if it were possible, because we have shown the existence of static user does not significantly affect the capacity to the dynamic user.

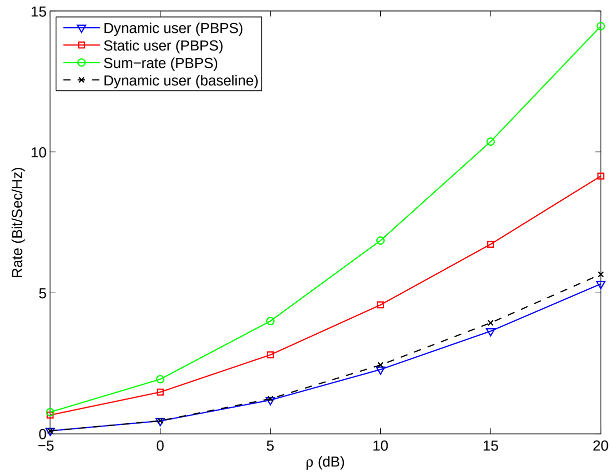

Fig. 2. Rate achieved by the pilot-based product superposition (PBPS): $N_d = 2$, $N_s = M = 4$ and $T_d = 5$.

V. Numerical Results

Unless specified otherwise, a power allocation is assumed (c_τ and c_δ) that maximizes the rate for the dynamic user.

Figure 2 illustrates the rate for dynamic and static users in the pilot-based product superposition, as shown in Theorem 1. We consider $N_d = 2$, $N_s = M = 4$ and $T_d = 5$. Numerical results correspond to the point on the rate region where the rate of the dynamic user is optimized. This is done to capture the corner point of the DoF region for the new scheme, and to highlight the most significant differences between the new scheme and the baseline scheme. At this operating point, in addition to near-optimal rate for the dynamic user, the proposed method provides significant rate for the static user. The degradation of the rate of the dynamic user, compared with the baseline scheme, is negligible in the low-SNR regime, and in the high-SNR regime the rate of the dynamic user has the optimal degrees of freedom (SNR slope). Thus the proposed method achieves the static user's rate almost "for free" in terms of the penalty to the dynamic user.

Figure 3 shows the impact of the available antenna of the static user. Here, $\rho = 10$ dB, $N_d = 2$, $M = N_s$ and $T_d = 5$. The static user's rate (thus the sum-rate) increases linearly with N_s, because the degrees of freedom is $N_d N_s / T_d$, as indicated by Theorem 1. The gap of the dynamic user's rate under the proposed method and the baseline method vanishes as N_s increases. Intuitively, the rate difference is because of the Jensen's loss: in the proposed method the equivalent channel is the product matrix $\mathbf{H}_d \mathbf{X}_s$ and is "more spread" than the channel in the baseline method. As N_s increases, by law of large numbers the columns of $\mathbf{X}_s$ will become orthonormal with probability one ($\mathbf{X}_s \mathbf{X}_s^H / N_s \to \mathbf{I}_{N_d}$) and thus will have a smaller impact on the distribution of $\mathbf{H}_d$.

Figure 4 demonstrates the impact of the coherence time of the dynamic user. Here, $\rho = 10$ dB, $N_d = 2$, and $N_s = M = 4$. As T_d increases, the rate for the dynamic user improves, since the portion of time-slots (overhead)

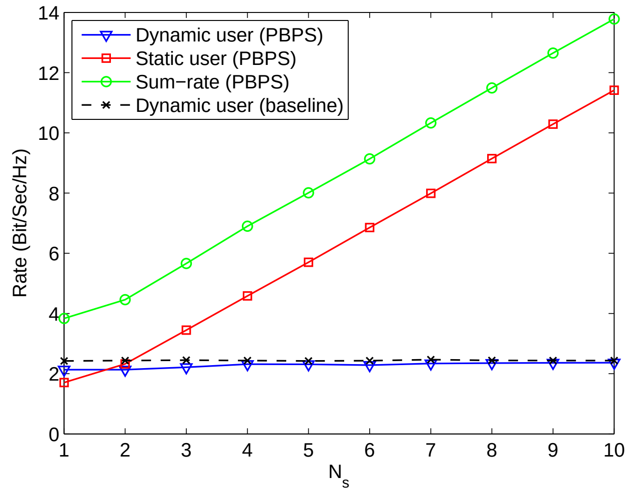

Fig. 3. Impact of the number of receive antennas of the static user: $\rho = 10$ dB, $N_d = 2$, $M = N_s$ and $T_d = 5$.

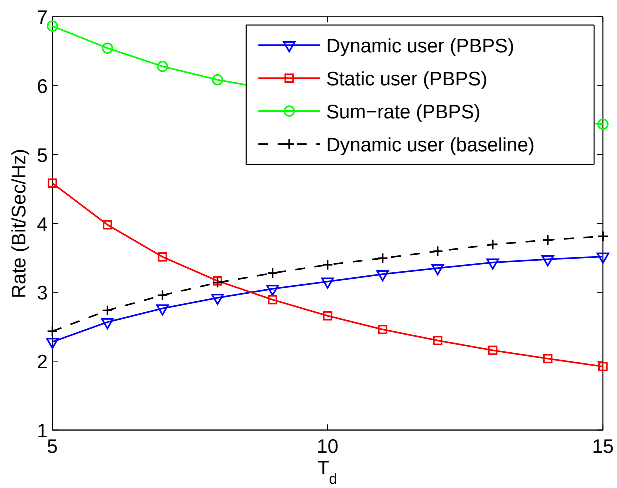

Fig. 4. Impact of channel coherence time: $\rho = 10$ dB, $N_d = 2$, and $N_s = M = 4$.

used for training is reduced. In contrast, the rate for the static user decreases with T_d, because the static user transmits new signal matrix over T_d period. Intuitively, as T_d increases, the dynamic user's channel becomes "more static", and therefore, the opportunity to explore its "insensitivity" to the channel is reduced.

Finally, in Figure 5, we show the gain of interference decoding in the pilot-based product superposition, where $N_d = 2$, $N_s = M$ and $T_d = 5$. By decoding the dynamic signal , the static rate is improved around 10%: the static user can now harvest the power carried not only by the dynamic user's pilot (the case without interference decoding) but also the dynamic user's data. This power gain does not increase the degrees of freedom of the static user, so the slope of the rate under two schemes are the same.

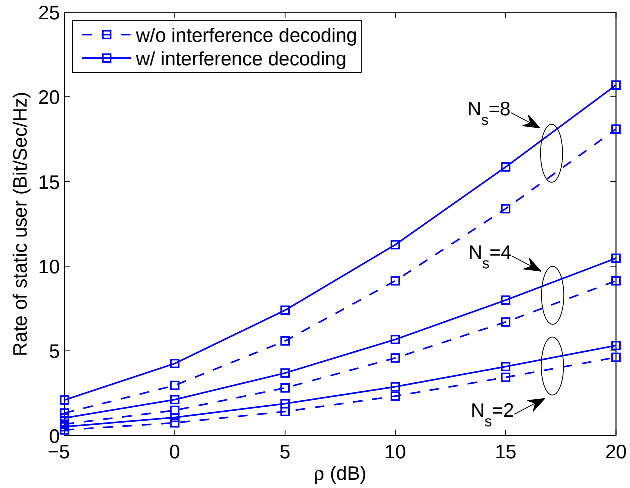

Fig. 5. Static user's rate with interference decoding: $N_d = 2$ and $T_d = 5$.

VI. Discussions, Extensions, and Conclusion

In this paper, we propose and analyze a pilot-based signaling that significantly improves the rate performance of the MIMO broadcast channel with varying CSIR. The proposed method sends a product of two signal matrices for the static and dynamic user, respectively, and each user decodes its own signal in a conventional manner. For the entire SNR range, the static user attains considerable rate almost without degrading the rate for the dynamic user. The static user's rate is further improved by allowing the static user to cancel the dynamic user's signal.

Remark 4: It is possible to extend the results of this paper to more than two receivers. The essence of the product superposition is to allow additional transmission for a static user when transmitting to a dynamic user. In case of more than two users, the static (dynamic) users can be grouped together. At each point in time, the transmitter uses product superposition to broadcast to one selected user from the static group and another user from the dynamic group.

Remark 5: Note that throughout this paper, both users are assumed to be in an ergodic mode of operation, i.e., the codewords are sufficiently long to allow coding arguments to apply. Simple extensions to this setup are easily obtained. For example, if the static user's coherence time is very long, one may adapt the transmission rate of the static user to its channel but allow the dynamic user to remain in an ergodic mode. Most expressions in this paper remain the same, except that for the rates and powers of the static user, expected values will be replaced with constant values.

Remark 6: As long as both users are in the ergodic mode, and the static user has more antennas than the dynamic user, it will be able to decode and cancel the interference caused by the dynamic user's signal. If we are in a mode where the static user's rate is adapted to the channel (as mentioned in Remark 5 above) and the dynamic user is in ergodic mode, then the static user may not always be able to decode the dynamic user's data because it cannot observe enough channel realizations to allow coding arguments to apply. In this case, sometimes the static user may experience an "outage" with respect to decoding the dynamic user's data. In this case, it can default to the oblivious method discussed in the early part of this paper and decode its own signal without peeling off the other user's data. The full exploration of such extensions is the subject of future research.

Remark 7: In each of the methods mentioned earlier in this paper, the static user operates under an equivalent single-user channel, by inverting either the pilot component or all components of the dynamic user's signal. Thus, any benefits available in single-user MIMO systems can also be available to the static user, including the benefits arising from CSIT. For example, water-filling can be applied to allocate power across multiple eigen-modes of the static user. However, this will change the effective channel seen by the dynamic user, thus complicating the analysis. The full analysis of this scenario is the subject of future research.

Acknowledgment

The authors gratefully acknowledge the valuable help of Mr. Mohamed Fadel.

Appendix I
Proof Of Theorem 1

A. Rate of the Static User

During the first N_d time-slots, the static user receives

$$\mathbf{Y}_{s1} = \sqrt{c_\tau}\,\mathbf{H}_s\mathbf{X}_s\mathbf{X}_\tau + \mathbf{W}_{s1}. \tag{47}$$

Because the static user knows $\mathbf{X}_\tau$, it removes the impact of $\mathbf{X}_\tau$ from $\mathbf{Y}_{2\tau}$:

$$\mathbf{Y}'_{s1} = \mathbf{Y}_{s1}\mathbf{X}_\tau^H \tag{48}$$

$$= \sqrt{c_\tau}\,\mathbf{H}_s\mathbf{X}_s + \mathbf{W}'_{s1} \tag{49}$$

where $\mathbf{Y}_{s1} \in \mathcal{C}^{N_s\times N_d}$ and $\mathbf{W}'_{s1}$ is the equivalent noise whose entries remain i.i.d. $\mathcal{CN}(0,1)$. Therefore, the channel seen by the static user becomes a point-to-point MIMO channel. Let $\mathbf{y}'_{si}$ and $\mathbf{x}_{si}$ be the column i of $\mathbf{Y}'_{s1}$ and $\mathbf{X}_s$, respectively. The mutual information

$$I(\mathbf{Y}_{s1};\mathbf{X}_s) = \sum_{i=1}^{N_d} I(\mathbf{y}'_{si};\mathbf{x}_{si}) \tag{50}$$

$$= N_d \log\det\Big(\mathbf{I}_{N_s} + c_\tau\,\mathbf{H}_s\mathbf{H}_s^H\Big), \tag{51}$$

which implies that the effective SNR for the static user is

$$\rho_s = c_\tau. \tag{52}$$

In the following $T_d - N_d$ time-slots, the static user disregards the received signal. The average rate achieved by the static user is

$$R_s = \frac{N_d}{T_d}\,\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_s} + \rho_s\,\mathbf{H}_s\mathbf{H}_s^H\Big)\Big], \tag{53}$$

where the expectation is over the channel realizations of $\mathbf{H}_s$.

B. Rate of the Dynamic User

The dynamic user first estimates the equivalent channel and then decodes its data. During the first N_d time-slots, the dynamic user receives the pilot signal

$$\mathbf{Y}_\tau = \sqrt{c_\tau}\,\mathbf{H}_d\mathbf{X}_s\mathbf{X}_\tau + \mathbf{W}_\tau \tag{54}$$

$$= \sqrt{c_\tau N_s}\,\widetilde{\mathbf{H}}_d\mathbf{X}_\tau + \mathbf{W}_\tau, \tag{55}$$

where $\widetilde{\mathbf{H}}_d \in \mathcal{C}^{N_d\times N_d}$ is the equivalent channel of the dynamic user

$$\widetilde{\mathbf{H}}_d \triangleq \frac{1}{\sqrt{N_s}}\,\mathbf{H}_d\mathbf{X}_s \tag{56}$$

Let $\tilde{h}_{ij} = [\widetilde{\mathbf{H}}_d]_{ij}$, then we have $\mathbb{E}[\tilde{h}_{ij}] = 0$ and

$$\mathbb{E}[\tilde{h}_{ij}\,\tilde{h}_{pq}^H] = \begin{cases} 1, & \text{if } (i,j) = (p,q) \\ 0, & \text{else} \end{cases}, \tag{57}$$

i.e., the entries of $\widetilde{\mathbf{H}}_d$ are uncorrelated and have zero-mean and unit variance.

The dynamic user estimates $\widetilde{\mathbf{H}}_d$ by the MMSE. Let

$$C_{YY} = (1 + c_\tau N_s)\mathbf{I}_{N_d}, \quad C_{YH} = \sqrt{c_\tau N_s}\,\mathbf{X}_\tau^H, \tag{58}$$

we have

$$\widehat{\mathbf{H}}_d = \mathbf{Y}_\tau C_{YY}^{-1} C_{YH} \tag{59}$$

$$= \frac{\sqrt{c_\tau N_s}}{1 + c_\tau N_s}\left(\sqrt{c_\tau N_s}\,\widetilde{\mathbf{H}}_d + \mathbf{W}_\tau\mathbf{X}_\tau^H\right) \tag{60}$$

Because $\mathbf{W}_\tau$ has i.i.d. $\mathcal{CN}(0,1)$ entries, the noise matrix $\mathbf{W}_\tau\mathbf{X}_\tau^H$ also has i.i.d. $\mathcal{CN}(0,1)$ entries. Define $\hat{h}_{1ij} = [\widehat{\mathbf{H}}_d]_{ij}$. Then, we have $\mathbb{E}[\hat{h}_{1ij}] = 0$ and

$$\mathbb{E}[\hat{h}_{ij}\hat{h}_{pq}^H] = \begin{cases} \alpha^2, & \text{if } (i,j) = (p,q) \\ 0, & \text{else} \end{cases}, \tag{61}$$

where

$$\alpha^2 \triangleq \frac{c_\tau N_s}{1 + c_\tau N_s}. \tag{62}$$

In other words, the estimate of the equivalent channel has uncorrelated elements with zero-mean and variance α^2.

During the remaining $T_d - N_d$ time-slots, the dynamic user regards the channel estimate $\widehat{\mathbf{H}}_d$ as the true channel and decodes the data signal. At the time-slot i, $N_d < i \le T_d$, the dynamic user receives

$$\mathbf{y}_{di} = \sqrt{c_\delta N_s}\,\widehat{\mathbf{H}}_d\mathbf{x}_{di} + \underbrace{\sqrt{c_\delta N_s}\,\widetilde{\mathbf{H}}_e\mathbf{x}_{di} + \mathbf{w}_{di}}_{\mathbf{w}'_{di}}, \tag{63}$$

where $\widetilde{\mathbf{H}}_e = \widetilde{\mathbf{H}}_d - \widehat{\mathbf{H}}_d$ is the estimation error for $\widetilde{\mathbf{H}}_d$, and $\mathbf{w}'_{di}$ is the equivalent noise that has zero mean and autocorrelation

$$\mathbf{R}_{w'_d} = c_\delta N_s\,\mathbb{E}\big[\widetilde{\mathbf{H}}_e\widetilde{\mathbf{H}}_e^H\big] + \mathbf{I}_{N_d} \tag{64}$$

$$= \big(1 + \frac{c_\delta N_d N_s}{1 + c_\tau N_s}\big)\mathbf{I}_{N_d}. \tag{65}$$

The equivalent noise $\mathbf{w}'_{di}$ is uncorrelated with the signal $\mathbf{x}_{di}$, because $\mathbb{E}[\tilde{\mathbf{H}}_e \mathbf{x}_{di}\mathbf{x}_{di}^H] = \mathbb{E}[\tilde{\mathbf{H}}_e]\mathbb{E}[\mathbf{x}_{di}\mathbf{x}_{di}^H] = 0$. Therefore, from [9, Thm.1], the mutual information is lower bounded by:

$$I(\mathbf{y}_{di}; \mathbf{x}_{di}|\widehat{\mathbf{H}}_d) \geq \log\det\Bigg(\mathbf{I}_{N_d} + \frac{c_\delta N_s\, \widehat{\mathbf{H}}_d \widehat{\mathbf{H}}_d^H}{1 + c_\delta N_d N_s/(1 + c_\tau N_s)}\Bigg) \tag{66}$$

$$= \log\det\Bigg(\mathbf{I}_{N_d} + \frac{c_\delta \alpha^2 N_s\, \overline{\mathbf{H}}_d \overline{\mathbf{H}}_d^H}{1 + c_\delta N_d N_s/(1 + c_\tau N_s)}\Bigg), \tag{67}$$

where $\overline{\mathbf{H}}_d$ is the normalized channel whose elements have unit variance

$$\overline{\mathbf{H}}_d = \frac{1}{\alpha}\widehat{\mathbf{H}}_d. \tag{68}$$

From (67), the effective SNR for the dynamic user can be defined as

$$\rho_d = \frac{c_\delta \alpha^2 N_d N_s}{1 + c_\delta N_d N_s/(1 + c_\tau N_s)}. \tag{69}$$

The average rate that the dynamic user achieves is

$$R_d \geq (1 - \frac{N_d}{T_d})\mathbb{E}\big[\log\det(\mathbf{I}_{N_d} + \frac{\rho_d}{N_d}\overline{\mathbf{H}}_d \overline{\mathbf{H}}_d^H)\big], \tag{70}$$

where the expectation is over the dynamic user's channel realizations.

APPENDIX II
PROOF OF THEOREM 2

We first show that if the codeword used by the dynamic user is sufficiently long, the static user always decodes the dynamic user's signal.

Similar to the dynamic user, the equivalent channel of the static user $\widetilde{\mathbf{H}}_s \triangleq \mathbf{H}_s \mathbf{X}_s/\sqrt{N_s}$ can be estimated as $\widehat{\mathbf{H}}_s \in \mathcal{C}^{N_s \times N_d}$ by using the pilot $\mathbf{X}_\tau$. During time-slots $i = N_d + 1, \ldots, T_d$, the static user receives:

$$\mathbf{y}_{si} = \sqrt{c_\delta N_s}\, \widehat{\mathbf{H}}_s \mathbf{x}_{di} + \underbrace{\sqrt{c_\delta N_s}\, \widetilde{\mathbf{H}}_e \mathbf{x}_{di} + \mathbf{w}_{si}}_{\mathbf{w}'_{si}}, \tag{71}$$

where $\mathbf{x}_{di} \in \mathcal{C}^{N_d \times 1}$ is the i-th column of X_d. The mutual information

$$I(\mathbf{y}_{si}; \mathbf{x}_{di}|\widehat{\mathbf{H}}_s) \geq \log\det\Bigg(\mathbf{I}_{N_s} + \frac{c_\delta N_s\, \widehat{\mathbf{H}}_s \widehat{\mathbf{H}}_s^H}{1 + c_\delta N_d N_s/(1 + c_\tau N_s)}\Bigg) \tag{72}$$

$$= \log\det\big(\mathbf{I}_{N_d} + \frac{\rho_d}{N_d}\overline{\mathbf{H}}_s \overline{\mathbf{H}}_s^H\big), \tag{73}$$

where $\overline{\mathbf{H}}_s = \frac{1}{\alpha}\widehat{\mathbf{H}}_s$ is the normalized channel estimate and ρ_d was given in (19). For the static user, the effective SNR for decoding the dynamic signal is identical to that of the dynamic user.

The static user also experiences many channel realizations over the dynamic user codewords. Write $\overline{\mathbf{H}}_s = [\overline{\mathbf{H}}_{s1}; \overline{\mathbf{H}}_{s2}]$, where $\overline{\mathbf{H}}_{s1} \in \mathcal{C}^{N_d \times N_d}$ and $\overline{\mathbf{H}}_{s2} \in \mathcal{C}^{(N_s - N_d) \times N_d}$. Then,

$$\mathbb{E}\big[I(\mathbf{y}_{si}; \mathbf{x}_{di}|\widehat{\mathbf{H}}_s)\big]$$

$$\geq \mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_d} + \rho_d\big(\overline{\mathbf{H}}_{s1}\overline{\mathbf{H}}_{s1}^H + \overline{\mathbf{H}}_{s2}\overline{\mathbf{H}}_{s2}^H\big)\bigg)\bigg] \tag{74}$$

$$\geq \mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_d} + \rho_d\overline{\mathbf{H}}_{s1}\overline{\mathbf{H}}_{s1}^H\bigg)\bigg], \tag{75}$$

$$= \mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_d} + \rho_d\mathbf{H}_d\mathbf{H}_d^H\bigg)\bigg], \tag{76}$$

$$= R_d \tag{77}$$

where (75) uses $\log\det(\mathbf{A}+\mathbf{B}) \geq \log\det\mathbf{A}$ for positive definite matrices $\mathbf{A}, \mathbf{B}$, and (76) uses the fact that $\overline{\mathbf{H}}_{s1}$ has the same distribution as $\overline{\mathbf{H}}_d$. Therefore the static user can decode the dynamic user's signal, and from here on we assume the static user has access to the dynamic user signal.

We now use the singular value decomposition of the dynamic signal $\mathbf{X}_d = \mathbf{U}_d\mathbf{\Sigma}_d\mathbf{V}_d^H$, where $\mathbf{U}_d \in \mathcal{C}^{N_d\times N_d}$, $\mathbf{V}_d \in \mathcal{C}^{T_d\times N_d}$ are unitary matrices, and $\mathbf{\Sigma}_d = \text{diag}(\lambda_1, \cdots, \lambda_{N_d})$. Then, we have

$$\mathbf{Y}_s' = \mathbf{Y}_s\mathbf{V}_d\mathbf{\Sigma}_d^{-1} \tag{78}$$

$$= \mathbf{H}_s\mathbf{X}_s\mathbf{U}_d + \mathbf{W}_s\mathbf{V}_d\mathbf{\Sigma}_d^{-1} \tag{79}$$

$$\triangleq \mathbf{H}_s\mathbf{X}_s' + \mathbf{W}_s'\mathbf{\Sigma}_d^{-1}, \tag{80}$$

where $\mathbf{X}_s' = \mathbf{X}_s\mathbf{U}_d, \mathbf{W}_s' = \mathbf{W}_s\mathbf{V}_d$. Because $\mathbf{U}_d$, $\mathbf{V}_d$ are unitary, the entries of $\mathbf{X}_s', \mathbf{W}_s' \in \mathcal{C}^{N_s\times N_d}$ remain i.i.d. $\mathcal{CN}(0,1)$. Define $\mathbf{y}_s' = \mathbf{vec}(\mathbf{Y}_s')$, $\mathbf{x}_s' = \mathbf{vec}(\mathbf{X}_s')$, $\mathbf{H}_s' = \mathbf{I}_{N_d} \otimes \mathbf{H}_s$ and

$$\mathbf{w}_s' = \mathbf{vec}(\mathbf{W}_s'\mathbf{\Sigma}_d^{-1}) = \begin{bmatrix} \frac{1}{\lambda_1}\mathbf{w}_{s1}' \\ \vdots \\ \frac{1}{\lambda_{N_d}}\mathbf{w}_{sN_d}' \end{bmatrix}. \tag{81}$$

Then, from (80), we write $\mathbf{y}_s' \in \mathcal{C}^{N_dN_s\times 1}$ as

$$\mathbf{y}_s' = \mathbf{H}_s'\mathbf{x}_s' + \mathbf{w}_s'. \tag{82}$$

The mutual information

$$I(\mathbf{Y}_s;\mathbf{X}_s|\mathbf{H}_s,\mathbf{X}_d) = I(\mathbf{y}_s';\mathbf{x}_s'|\mathbf{H}_s,\mathbf{X}_d) \tag{83}$$

$$= \log\det\bigg(\mathbf{I}_{N_dN_s} + \mathbf{R}_{w_s'}^{-1}\ \mathbf{H}_s'\mathbf{H}_s'^H\bigg), \tag{84}$$

where $\mathbf{R}_{w_s'} = \mathbb{E}[\mathbf{w}_s'\mathbf{w}_s'^H]$ is the noise autocorrelation matrix that is given by

$$\mathbf{R}_{w_s'} = \begin{bmatrix} \mathbb{E}[\lambda_1^{-2}]\mathbf{I}_{N_s} & & \\ & \ddots & \\ & & \mathbb{E}[\lambda_{N_d}^{-2}]\mathbf{I}_{N_s} \end{bmatrix}. \tag{85}$$

Therefore, the average rate attained by the static user is

$$R_s = \frac{1}{T_d}\mathbb{E}[I(\mathbf{Y}_s;\mathbf{X}_s|\mathbf{H}_s,\mathbf{X}_d)] \tag{86}$$

$$= \frac{1}{T_d}\mathbb{E}\left[\sum_{i=1}^{N_d}\log\det\left(\mathbf{I}_{N_s} + \frac{1}{\mathbb{E}[\lambda_i^{-2}]}\mathbf{H}_s\mathbf{H}_s^H\right)\right] \tag{87}$$

$$= \frac{N_d}{T_d}\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_s} + \frac{1}{\mathbb{E}[\lambda_1^{-2}]}\mathbf{H}_s\mathbf{H}_s^H\right)\right], \tag{88}$$

where the last equality holds because the marginal distributions of $\{\lambda_i\}$ are identical.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691 – 1706, July 2003.

[2] C. Huang, S. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without channel state information at transmitters," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 849 –857, Feb. 2012.

[3] Y. Li and A. Nosratinia, "Product superposition for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 58, no. 11, pp. 6839–6852, Nov. 2012.

[4] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[5] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936 –3964, Sept. 2006.

[6] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045 –5060, Nov. 2006.

[7] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528 – 541, Mar. 2006.

[8] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[10] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. on Telecomm.*, vol. 10, no. 6, pp. 585–595, 1999.

[11] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2383–2399, Sept. 2001.