Statistically Adaptive Spatial Multiplexing for Time-varying Correlated MIMO Channels*

Yan Zhou and Akbar M. Sayeed

Department of Electrical and Computer Engineering
University of Wisconsin–Madison
1415 Engineering Drive
Madison,WI 53706

Corresponding author:

Akbar M. Sayeed

Department of Electrical and Computer Engineering
University of Wisconsin–Madison
1415 Engineering Drive
Madison,WI 53706
Email: akbar@engr.wisc.edu
Phone: (608)265-4731, Fax: (608)265-4623

*This work is supported by the NSF under grant #CCF-0431088.

Abstract

This work presents a statistically adaptive spatial multiplexing scheme for time-varying correlated MIMO channels. Unlike the conventional spatially uncorrelated block fading channel model, both channel variation in each data block and spatial correlation are accommodated in the design. The proposed system continuously transmits packets with identical structure. Each packet consists of a training phase followed by a data transmission phase. The training signal is used to predict instantaneous channel state at the receiver. In each symbol period of data transmission phase, the transmitter simultaneously transmits multiple data streams, which are decoded at the receiver based on the predicted channel state. The transmitter is assumed to know only the channel statistics, with which the power and rate for each data stream as well as the number of data streams are adjusted in each symbol period to achieve a target bit error rate. The packet length is further optimized to maximize the spectral efficiency. The results reveal that the rate-maximizing transmission strategy makes a judicious diversity-multiplexing tradeoff by allocating power to an optimum number of data streams. Furthermore, the optimum packet length decreases according to the order MIMO $>\{$MISO, SIMO$\}>$ SISO and decreases for fewer antennas in each case.

Index Terms – Multi-antenna systems, time-varying channels, adaptive modulation

I. Introduction

Multiple antennas at the transmitter and receiver have been shown to dramatically improve the capacity and reliability of wireless communication links and are often referred to as multiple-input multiple-output (MIMO) systems [1, 2]. However, design of communication strategies for time-varying channels is challenging due to temporal channel variations that lead to outdated channel state information at both ends. The impact of feedback delay on spatial multiplexing in time-varying MIMO channels has been studied in [3, 4], which assume that error-free but delayed instantaneous channel information is available at the transmitter. The use of correction matrices at the receiver is suggested by [3] to partially compensate the effect of delay, while [4] incorporates the channel uncertainty due to time variations in

the precoder design. In the space-time coded adaptive modulation scheme [5], the transmitter optimizes the symbol modulation levels in orthogonal space-time codes based on the feedback of instantaneous received SNR, and the effect of feedback delay has been investigated. In [6], Alamouti coded data are transmitted along two orthogonal basis beams with optimum power allocation and modulation levels. In contrast to [3]-[5], the transmitter is assumed to only know the instantaneous estimated channel information with feedback delay, which better matches the realistic situation. In [7], the proposed adaptive transmit beamformer exploits the predicted channel information to mitigate the effect of feedback delay, and the results show that a critical value of normalized prediction error is a good indicator of performance degradation. Compared with the time-varying MIMO channels, the time-varying SISO channels have been studied more extensively. Some representative work is described here. The pilot-based estimation of time-varying SISO channels and its impact on receiver performance have been studied in [8, 9]. A closed-form lower bound of average channel capacity is derived in [10] in terms of channel estimation error variance, and the optimal pilot symbol spacing and power allocation are determined by maximizing this lower bound. The effects of feedback delay and error on the performance of adaptive systems with instantaneous channel feedback have been investigated in [11, 12].

This work is motivated by the following two reasons. First, most existing research on time-varying MIMO systems assumes a spatially uncorrelated block fading channel model, which ignores temporal channel variation in each data block as well as the spatial correlation. This will in general lead to inaccurate performance predictions. Secondly, most existing work assumes that the transmitter knows the instantaneous channel information, which can be either fed back from the receiver in FDD systems [7] or estimated at the transmitter in TDD systems [3]. However, both methods will induce delay and error in the channel information, which will inevitably degrade performance. In this work, a statistically adaptive spatial multiplexing scheme is proposed for narrowband time-varying correlated MIMO channels and has the following features: (1) The design is based on the virtual channel representation [15, 16], which jointly incorporates temporal channel variations and spatial correlation; (2) Instead of instantaneous channel information, the transmitter is assumed to only know the channel statistics, which

usually change at much slower time scales and can thus be tracked more easily and reliably; (3) Unlike conventional adaptive modulation schemes whose power and rate for each data stream are invariant in each data block, the proposed adaptive spatial multiplexing scheme adjusts the power and rate for each data stream as well as the number of data streams in every symbol period based on the channel statistics, and the data block length is further optimized to maximize the spectral efficiency. The results reveal that the rate-maximizing transmission strategy makes a judicious diversity-multiplexing tradeoff by allocating power to an optimum number of data streams. Furthermore, the optimum data block length decreases according to the order MIMO $>$ $\{$MISO, SIMO$\}$ $>$ SISO and decreases for fewer antennas in each case, since each virtual channel coefficient is contributed by more paths with larger Doppler spread due to the coarser spatial resolution, and the larger spread reduces the channel coherence time in virtual domain and hence the optimum data block length. The proposed statistics-based design is also compared to that with perfect instantaneous Channel State Information at Transmitter (CSIT), which on average improves the rate by 24%. Therefore, the former might be a good complexity-performance tradeoff by avoiding the instantaneous feedback of CSIT, which will both reduce the resource and suffer from the imperfections in practice.

The paper is organized as follows. Section II describes the virtual channel representation for time-varying correlated MIMO channels. Section III provides an overview of the proposed scheme as well as the design objective. The optimization of system components is elaborated in Section IV. For comparison, the scheme with perfect instantaneous channel information at transmitter is described in Section V. Numerical results including the performance comparison of both schemes are presented in Section VI, and Section VII contains concluding remarks.

II. Signaling in Virtual Angle Domain

Consider a MIMO system with Uniform Linear Arrays (ULA) at both ends with N_T transmit antennas and N_R receive antennas. For a narrowband time-varying MIMO channel, the $N_R \times 1$ received

signal vector at time t can be written as

$$\boldsymbol{y}_c(t) = \mathbf{H}_c(t)\boldsymbol{x}_c(t) + \boldsymbol{n}_c(t), \qquad 0 \le t \le T \tag{1}$$

where $\boldsymbol{x}_c(t)$ is the $N_T \times 1$ transmitted signal vector, $\boldsymbol{n}_c(t)$ is the i.i.d. Gaussian noise vector, and T presents the overall signaling duration. The $N_R \times N_T$ channel matrix at time t is modeled as

$$\mathbf{H}_c(t) = \sum_{l=1}^{L} \beta_l \boldsymbol{a}_R(\theta_{R,l})\boldsymbol{a}_T^H(\theta_{T,l})e^{j2\pi\nu_l t} \tag{2}$$

where L is the number of paths. For the l-th path, the array steering and response vectors are given by

$$\boldsymbol{a}_T(\theta_{T,l}) = \left[1, e^{-j2\pi\theta_{T,l}}, \ldots, e^{-j2\pi(N_T-1)\theta_{T,l}}\right]^T, \quad \boldsymbol{a}_R(\theta_{R,l}) = \left[1, e^{-j2\pi\theta_{R,l}}, \ldots, e^{-j2\pi(N_R-1)\theta_{R,l}}\right]^T \tag{3}$$

where the parameter θ is related to the physical path angle[1] ϕ as $\theta = d\sin\phi/\lambda$ with λ and d denoting the wavelength and the antenna spacing, respectively. Without loss of generality (WLOG), we focus on the critical spacing $d = \lambda/2$ in this paper and, hence, $\theta = \sin\phi/2 \in [-0.5, 0.5]$. The impact of antenna spacing on capacity and diversity performance has been investigated in [15]. The complex path amplitude $\beta_l = \alpha_l e^{j\varphi_l}$ has envelope $\alpha_l > 0$ and phase φ_l uniformly distributed in $[0, 2\pi]$. WLOG, the transmitter and receiver are assumed to move in the array-broadside directions at speeds of v_T and v_R. The resultant Doppler frequency shift of the l-th path is given by $\nu_l = (v_R\cos\phi_{R,l} + v_T\cos\phi_{T,l})/\lambda$ with the maximum shift as $f_{\max} = (v_R + v_T)/\lambda$. Moreover, $\{\alpha_l\}$, $\{\theta_{R,l}\}$, $\{\theta_{T,l}\}$, $\{\nu_l\}$ are fixed for a given scattering environment, while $\{\varphi_l\}$ randomly change for different channel realizations [13, 14].

In contrast to (1), signaling can be realized in virtual angle domain (beamspace) instead of in spatial domain. The signal relation in virtual angle domain can be written as

$$\boldsymbol{y}(t) = \mathbf{H}(t)\boldsymbol{x}(t) + \boldsymbol{n}(t) \tag{4}$$

where the virtual channel matrix $\mathbf{H}(t) = \mathbf{A}_R^H\mathbf{H}_c(t)\mathbf{A}_T$ is the 2D Discrete Fourier Transform (DFT) of $\mathbf{H}_c(t)$, and the transformed vectors $\boldsymbol{y}(t) = \mathbf{A}_R^H\boldsymbol{y}_c(t)$, $\boldsymbol{x}(t) = \mathbf{A}_T^H\boldsymbol{x}_c(t)$, $\boldsymbol{n}(t) = \mathbf{A}_R^H\boldsymbol{n}_c(t)$ represent the received, transmitted, and noise vectors in beamspace. The $N_R \times N_R$ and $N_T \times N_T$ unitary DFT matrices

[1] measured w.r.t. the array broadside.

are given by $\mathbf{A}_R = [\boldsymbol{a}_R(\widetilde{\theta}_{R,1}), \ldots, \boldsymbol{a}_R(\widetilde{\theta}_{R,N_R})]/\sqrt{N_R}$ and $\mathbf{A}_T = [\boldsymbol{a}_T(\widetilde{\theta}_{T,1}), \ldots, \boldsymbol{a}_T(\widetilde{\theta}_{T,N_T})]/\sqrt{N_T}$ with the fixed receive and transmit virtual angles defined as

$$\widetilde{\theta}_{R,q} = \frac{q - \widetilde{N}_R}{N_R},\ q = 1, \ldots, N_R, \quad \widetilde{\theta}_{T,p} = \frac{p - \widetilde{N}_T}{N_T},\ p = 1, \ldots, N_T \tag{5}$$

where, WLOG, we assume the number of antennas is odd and define $\widetilde{N}_R = (N_R - 1)/2 + 1,\ \widetilde{N}_T = (N_T - 1)/2 + 1$. The signaling in virtual angle domain is illustrated in Fig.1. Through DFT, the p-th element of $\boldsymbol{x}(t)$ is transmitted from the p-th transmit beam directed at $\widetilde{\theta}_{T,p}$, while the q-th element of $\boldsymbol{y}(t)$ denotes the signal captured by the q-th receive beam at $\widetilde{\theta}_{R,q}$. As implied by (4), the virtual channel coefficient $H(q, p, t)$ in $\mathbf{H}(t)$ represents the channel coupling between the p-th transmitted element $x_p(t)$ and the q-th received element $y_q(t)$ in beamspace.

One important property of virtual channel coefficients is their approximately uncorrelated nature. This can be interpreted via virtual path partitioning, which introduces the following subsets of paths

$$\begin{aligned} S_{R,q} &= \{l : -1/(2N_R) \le (\theta_{R,l} - \widetilde{\theta}_{R,q}) < 1/(2N_R)\}, \quad q = 1, \ldots, N_R \\ S_{T,p} &= \{l : -1/(2N_T) \le (\theta_{T,l} - \widetilde{\theta}_{T,p}) < 1/(2N_T)\}, \quad p = 1, \ldots, N_T \end{aligned} \tag{6}$$

corresponding to the spatial resolutions: $\Delta\theta_R = 1/N_R$ and $\Delta\theta_T = 1/N_T$. The partitioning in case of $N_R = N_T = 3$ is illustrated in Fig.2, where each dot represents a path angular position $(\theta_{R,l}, \theta_{T,l})$ in the 2D domain $[-0.5, 0.5] \times [-0.5 \times 0.5]$. The 3 receive beams partition the paths into 3 rows $\{S_{R,q}\}_{q=1}^3$ with height $1/3$, and the 3 transmit beams partition the paths into 3 columns $\{S_{T,p}\}_{p=1}^3$ with width $1/3$. Based on the above partitioning, each virtual channel coefficient in (4) can be approximated as

$$H(q, p, t) = \frac{1}{\sqrt{N_R N_T}} \sum_{l=1}^{L} \beta_l \boldsymbol{a}_R^H(\widetilde{\theta}_{R,q}) \boldsymbol{a}_R(\theta_{R,l}) \boldsymbol{a}_T^H(\theta_{T,l}) \boldsymbol{a}_T(\widetilde{\theta}_{T,p}) e^{j2\pi\nu_l t} \approx \sqrt{N_R N_T} \sum_{l \in S_{q,p}} \beta_l e^{j2\pi\nu_l t} \tag{7}$$

where $S_{q,p} = S_{R,q} \cap S_{T,p}$ represents the paths jointly captured by the p-th transmit beam and the q-th receive beam, as illustrated in Fig.2. The mathematical reasoning of the approximation can be found in [15]. It indicates that $\{H(q, p, t)\}$ for different (q, p)'s are contributed by disjoint subsets of paths and, hence, they are uncorrelated due to the independent path amplitudes. Based on the uncorrelated nature, it has been shown that the capacity-achieving input covariance matrix in virtual domain $E\left[\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right]$

has a diagonal structure [20, 21]. This motivates the consideration of transmitting independent data streams via different virtual angles in the proposed scheme.

Furthermore, the temporal correlation of $H(q,p,t)$ can be derived from (7) as

$$r_{q,p}(t-t') \triangleq E\left[H(q,p,t)H^*(q,p,t')\right] = \int G_{q,p}(\nu)e^{j2\pi\nu(t-t')}d\nu \tag{8}$$

where the path Doppler power spectrum conditioned on (q,p) can be approximated as

$$\begin{aligned} G_{q,p}(\nu) &= \frac{1}{N_R N_T}\sum_{l=1}^{L}\alpha_l^2\left|\boldsymbol{a}_R^H(\widetilde{\theta}_{R,q})\boldsymbol{a}_R(\theta_{R,l})\right|^2\left|\boldsymbol{a}_T^H(\theta_{T,l})\boldsymbol{a}_T(\widetilde{\theta}_{T,p})\right|^2\delta(\nu-\nu_l) \\ &\approx N_R N_T\sum_{l\in S_{q,p}}\alpha_l^2\delta(\nu-\nu_l). \end{aligned} \tag{9}$$

The above approximation implies that the conditional Doppler power spectrum is contributed by the paths in $S_{q,p}$ and hence has a smaller spread than that in SISO channel, which is contributed by all paths. The smaller spread yields a slower decay of the corresponding temporal correlation and hence slows down the effective channel variation. This explains why MIMO systems can reduce the training update frequency, as reflected by the numerical results.

III. System Design Overview

The system continuously transmits and receives space-time packets with identical structure. WLOG, the structure of the first packet is shown in Fig. 3. It has a training phase with N_{tr} symbol periods followed by a data transmission phase with N_D symbol periods. The fixed symbol period is denoted as T_s, and the time 0 represents the end of training phase. The packet is transmitted and received in virtual angle domain. WLOG, the N_T transmit virtual angles are sorted in descending order according to their statistical strengths[2]. In the p-th training symbol period, a training symbol is transmitted from the beam at the p-th transmit virtual angle with full power ρ. This implies $N_{tr} \leq N_T$. Accordingly, the transmitted training signal matrix is $\mathbf{D} = \sqrt{\rho}\mathbf{I}_{N_{tr}}$ with $\mathbf{I}_{N_{tr}}$ as the identity matrix. In the n-th data symbol period, a $k(n)\times 1$ signal vector $\boldsymbol{x}(n)$ is launched from the first $k(n)$ transmit virtual beams.

[2]which means $E\|\mathbf{H}(:,1,t)\|_F^2 \geq ,\ldots, \geq E\|\mathbf{H}(:,N_T,t)\|_F^2$, where $\mathbf{H}(:,p,t)$ represents the p-th column of $\mathbf{H}(t)$ and $\|\cdot\|_F^2$ denotes the Frobenius norm. It can be shown that the order is independent of t for a given scattering environment.

The system block diagram in the n-th data symbol period is illustrated in Fig.4. At the transmitter side, the transmitted signal vector in virtual angle domain is formed as $\boldsymbol{x}(n) = \boldsymbol{\Lambda}^{1/2}(n)\boldsymbol{s}(n)$, where $\boldsymbol{\Lambda}(n)$ represents a diagonal power-shaping matrix. The $k(n)\times 1$ data vector $\boldsymbol{s}(n)$ consists of independent QAM symbols with $E[\boldsymbol{s}(n)\boldsymbol{s}^H(n)] = \mathbf{I}_{k(n)}$, and $R_p(n)$ represents the number of bits in the p-th symbol. Accordingly, the covariance matrix of $\boldsymbol{x}(n)$ has a diagonal structure, which is capacity-achieving [20, 21]. The vector $\boldsymbol{x}(n)$ is further zero-padded and DFT transformed to the spatial domain for antenna transmission: $\boldsymbol{x}_c(n) = \mathbf{A}_T[\boldsymbol{x}^T(n) \vdots \mathbf{0}^T]^T$, where $\mathbf{0}$ is a $(N_T - k(n))\times 1$ all-zero vector. Due to the DFT, the $k(n)$ data symbols in $\boldsymbol{x}(n)$ are respectively launched from the first $k(n)$ transmit virtual beams. At the receiver side, the signal vector on the N_R receive antennas is transformed to the beamspace through $\boldsymbol{y}(n) = \mathbf{A}_R^H \boldsymbol{y}_c(n)$, where $\boldsymbol{y}(n)$ consists of signals captured by the N_R receive virtual beams. The vector $\boldsymbol{y}(n)$ is then fed to a linear decoder to estimate the data vector: $\widehat{\boldsymbol{s}}(n) = \mathbf{G}(n)\boldsymbol{y}(n)$. The decoder is formed based on the predicted virtual channel matrix $\widehat{\mathbf{H}}(n)$, which is generated by feeding the $N_R \times N_{tr}$ signal matrix $\mathbf{Z}$ received in training phase to a linear channel predictor $\mathbf{L}(n)$.

The general design objective is to maximize the average rate per packet $\sum_{n=1}^{N_D}\sum_{p=1}^{k(n)} R_p(n)/(N_{tr} + N_D)$ subject to the transmit power constraint and the BER requirement for each data stream. The optimization is over $\mathbf{L}(n)$, $\mathbf{G}(n)$, $\boldsymbol{\Lambda}(n)$, $\{R_p(n)\}$, $k(n)$, and N_D with channel statistics known at both sides. The procedure is briefly described below.

At the receiver side:

Step 1 In each data symbol period, $\mathbf{L}(n)$ is formed by minimizing the Mean Square Error (MSE) between the true channel state and its prediction. The MSE is averaged over channel statistics. Therefore, $\mathbf{L}(n)$ is determined by channel statistics and hence is the same for different packets;

Step 2 The decoder $\mathbf{G}(n)$ is formed by minimizing the MSE between $\boldsymbol{s}(n)$ and $\widehat{\boldsymbol{s}}(n)$, which is averaged over $\boldsymbol{s}(n)$ and the noise. As shown later, $\mathbf{G}(n)$ is a function of $\widehat{\mathbf{H}}(n)$ and $\boldsymbol{\Lambda}(n)$. Therefore, it varies with different realizations of $\widehat{\mathbf{H}}(n)$ in different packets;

At the transmitter side:

Step 1 For a given number of data streams $k(n)$, $\boldsymbol{\Lambda}(n)$ is optimized by minimizing the MSE between

$\boldsymbol{s}(n)$ and $\widehat{\boldsymbol{s}}(n)$, which is averaged over channel statistics, since the transmitter only knows channel statistics. Next, the rate of the p-th stream $R_p(n)$ is chosen as the maximum number of bits keeping the corresponding average BER under a target BER. Both $\mathbf{\Lambda}(n)$ and $\{R_p(n)\}$ are optimized over channel statistics and hence do not change across the packets;

Step 2 The $k(n)$ is further optimized by maximizing the total rate in the n-th data symbol period $\sum_{p=1}^{k(n)} R_p(n)$, and the corresponding $\mathbf{\Lambda}(n)$ and $\{R_p(n)\}_{p=1}^{k(n)}$ are selected as the final designs;

Step 3 The data block length N_D is optimized by maximizing the average rate per packet defined earlier with the optimum $k(n)$ and the corresponding $\{R_p(n)\}_{p=1}^{k(n)}$ in the expression. For simplicity, the length of training phase N_{tr} can be first fixed as N_T, though the optimum value can be smaller as described later;

The proposed statistics-based design has the following two features. (1) It has low complexity compared with the CSIT-based design [3]-[7]: instantaneous channel feedback is not required, and the same optimized packet structure is applied over the period when the channel statistics are static; (2) The temporal and spatial channel correlations are exploited by optimizing the packet length and the number of data streams, which significantly improves the average rate, as shown in the results;

IV. Optimization of System Components

WLOG, we focus on the first packet and list major design assumptions as follows.

- $\mathbf{A1}$) For simplicity, $\mathbf{H}(t)$ in the training phase is assumed to be the same as that at time 0. However, extension to the case considering channel variation in the training phase is feasible;
- $\mathbf{A2}$) Virtual channel matrix $\mathbf{H}(t)$ in the n-th data symbol period in Fig. 3 is assumed to be the same as $\mathbf{H}(nT_s)$, which is denoted as $\mathbf{H}(n)$ hereafter;
- $\mathbf{A3}$) $H(q,p,t)$ has zero-mean Gaussian distribution. It can be seen from (7) that $H(q,p,t)$ is a weighted sum of independent path amplitudes. According to the central limit theorem, the distribution would be approximately Gaussian if the number of paths is large;
- $\mathbf{A4}$) $\{H(q,p,t)\}$ are independent for different (q,p)'s: $E[H(q,p,t)H^*(q',p',t')] = r_{q,p}(t-t')\,\delta_{q-q'}\delta_{p-p'}$ with $r_{q,p}(t-t')$ specified in (8). This assumption stems from (7);

- $\mathbf{A5}$) Both receiver and transmitter know the temporal channel correlations $\{r_{q,p}(nT_s), n = 1, \ldots, N_{max}\}$ for all (q, p)'s, where N_{max} denotes an upper bound of the data block length N_D. Those temporal correlations represent the channel statistics needed in the design;
- $\mathbf{A6}$) To get a tractable BER expression, the interference plus noise seen by each data stream at the linear decoder output is assumed to be Gaussian. This assumption is justified in [23] and has been widely applied in linear transceiver designs [26];
- $\mathbf{A7}$) The number of data streams $k(n)$ has an upper bound of $\min\{N_R, N_T\}$, which is the maximum rank of $\mathbf{H}(n)$;

A. Optimization of Channel Predictor

For simplicity, the length of training phase is first assumed to be $N_{tr} = N_T$. With $\mathbf{A1}$, the $N_R \times N_T$ received training matrix in virtual angle domain can be written as $\mathbf{Z} = \mathbf{H}(0)\mathbf{D} + \mathbf{W}$, where $\mathbf{D} = \sqrt{\rho}\mathbf{I}_{N_T}$ and the $N_R \times N_T$ noise matrix $\mathbf{W}$ has i.i.d. complex Gaussian entries with zero mean and variance σ_n^2. The predicted virtual channel matrix in the n-th data symbol period is given by

$$\widehat{\mathbf{H}}(n) = \overline{\text{vec}}(\widehat{\boldsymbol{h}}(n)),\ \widehat{\boldsymbol{h}}(n) = \mathbf{L}(n)\boldsymbol{z},\ \boldsymbol{z} = \text{vec}(\mathbf{Z}) = \left(\mathbf{D}^T \otimes \mathbf{I}_{N_R}\right)\boldsymbol{h}(0) + \boldsymbol{w} = \sqrt{\rho}\boldsymbol{h}(0) + \boldsymbol{w} \tag{10}$$

where $\mathbf{L}(n)$ is the $N_R N_T \times N_R N_T$ linear channel predictor, $\boldsymbol{z} = \text{vec}(\mathbf{Z})$, $\boldsymbol{h}(0) = \text{vec}(\mathbf{H}(0))$, and $\boldsymbol{w} = \text{vec}(\mathbf{W})$ are obtained by stacking[3] the columns of $\mathbf{Z}$, $\mathbf{H}(0)$, and $\mathbf{W}$, respectively, $\overline{\text{vec}}(\cdot)$ represents the inverse operation of $\text{vec}(\cdot)$, and $\boldsymbol{z}$ is specified via the identity $\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\text{vec}(\mathbf{B})$ with $\otimes$ denoting the Kronecker product. The MMSE channel predictor can be derived from the orthogonality principle [18] as

$$\begin{aligned} \mathbf{L}_o(n) &= \arg\min_{\mathbf{L}(n)} E\,\|\,\mathbf{L}(n)\boldsymbol{z} - \boldsymbol{h}(n)\,\|_F^2 \\ &= E\left[\boldsymbol{h}(n)\boldsymbol{z}^H\right]\left(E\left[\boldsymbol{z}\boldsymbol{z}^H\right]\right)^{-1} = \sqrt{\rho}\boldsymbol{\Omega}(n)\left(\rho\boldsymbol{\Omega}(0) + \sigma_n^2\mathbf{I}_{N_R N_T}\right)^{-1} \end{aligned} \tag{11}$$

where $\boldsymbol{h}(n) = \text{vec}(\mathbf{H}(n))$ with $\mathbf{H}(n)$ given in $\mathbf{A2}$, and the expectation is over both $\boldsymbol{h}$ and $\boldsymbol{w}$. According to $\mathbf{A4}$, $\boldsymbol{\Omega}(n) = E\left[\boldsymbol{h}(n)\boldsymbol{h}^H(0)\right]$ has a diagonal structure and, hence, $\mathbf{L}_o(n)$ is diagonal as

[3]Mathematically, $\boldsymbol{z} = \text{vec}(\mathbf{Z}) = [\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_{N_T}^T]^T$ with $\boldsymbol{z}_p$ as the p-th column of $\mathbf{Z}$, and $\overline{\text{vec}}(\boldsymbol{z}) = \mathbf{Z}$.

well. It can be shown that the $((p-1)N_R+q)$-th elements of $\widehat{\boldsymbol{h}}(n)$ and $\boldsymbol{z}$ correspond to $\widehat{H}(q,p,n)$ and $Z(q,p)$, respectively, and the $((p-1)N_R+q)$-th diagonal element of $\mathbf{L}_o(n)$ can be specified as $r_{q,p}(nT_s)\sqrt{\rho}/\left(r_{q,p}(0)\rho+\sigma_n^2\right)$. The above relations together with $\widehat{\boldsymbol{h}}(n)=\mathbf{L}_o(n)\boldsymbol{z}$ give the expressions of $\widehat{H}(q,p,n)$ and the associated prediction error as

$$\begin{aligned}\widehat{H}(q,p,n)&=\frac{r_{q,p}(nT_s)\sqrt{\rho}}{r_{q,p}(0)\rho+\sigma_n^2}Z(q,p)=\frac{r_{q,p}(nT_s)\sqrt{\rho}}{r_{q,p}(0)\rho+\sigma_n^2}\left(H(q,p,0)\sqrt{\rho}+W(q,p)\right)\\ E(q,p,n)&=H(q,p,n)-\widehat{H}(q,p,n)\end{aligned}\tag{12}$$

and, according to $\mathbf{A3}$, both have zero-mean complex Gaussian distribution

$$\begin{aligned}&\widehat{H}(q,p,n)\sim\mathbb{CN}\left(0,\widehat{\sigma}^2(q,p,n)\right),\quad \widehat{\sigma}^2(q,p,n)=|r_{q,p}(nT_s)|^2\rho/\left(r_{q,p}(0)\rho+\sigma_n^2\right)\\ &E(q,p,n)\sim\mathbb{CN}\left(0,\epsilon^2(q,p,n)\right),\quad \epsilon^2(q,p,n)=r_{q,p}(0)-\widehat{\sigma}^2(q,p,n).\end{aligned}\tag{13}$$

The matrix version of (12) is given by $\mathbf{H}(n)=\widehat{\mathbf{H}}(n)+\mathbf{E}(n)$ with $\widehat{\mathbf{H}}(n)$ known at the receiver. Based on $\mathbf{A4}$ and (12), we have the following properties

$$\begin{aligned}&E\left[\widehat{H}(q,p,n)\widehat{H}^*(q',p',n)\right]=\widehat{\sigma}^2(q,p,n)\delta_{qq'}\delta_{pp'}\\ &E\left[E(q,p,n)E^*(q',p',n)\right]=\epsilon^2(q,p,n)\delta_{qq'}\delta_{pp'},\quad E\left[E(q,p,n)\widehat{H}^*(q',p',n)\right]=0\end{aligned}\tag{14}$$

which indicate that $\widehat{\mathbf{H}}(n)$ and $\mathbf{E}(n)$ have independent entries, and they are also mutually independent. The channel estimation for correlated MIMO channels has been studied in [19], which shows that the optimum training signal corresponds to transmitting beams in successive symbol intervals along different transmit angles as assumed in Fig.3. The proposed scheme also considers the channel variation in each data block, and the channel in the n-th symbol period is predicted by the predictor in (11).

B. Optimization of Decoder and Power-Shaping Matrix

As described in Section III, the transmitted signal vector in virtual angle domain $\boldsymbol{x}(n)=\mathbf{\Lambda}^{1/2}(n)\boldsymbol{s}(n)$ has dimension $k(n)$, which is simplified as k in the following. The corresponding $N_R\times 1$ received signal vector in virtual domain can be written as

$$\boldsymbol{y}(n)=\mathbf{H}_k(n)\boldsymbol{x}(n)+\boldsymbol{n}(n)=\widehat{\mathbf{H}}_k(n)\boldsymbol{x}(n)+\boldsymbol{v}(n)\tag{15}$$

where the matrix with subscript k contains the first k columns of the original matrix, $\boldsymbol{n}(n)$ has i.i.d. complex Gaussian noise entries with zero mean and variance σ_n^2, the virtual channel matrix is decomposed as $\mathbf{H}_k(n) = \widehat{\mathbf{H}}_k(n) + \mathbf{E}_k(n)$, and $\boldsymbol{v}(n) = \mathbf{E}_k(n)\boldsymbol{x}(n) + \boldsymbol{n}(n)$ represents the effective noise vector. The MMSE decoder can be derived as

$$\begin{aligned}\mathbf{G}_o(n) &= \arg\min_{\mathbf{G}(n)} E \| \mathbf{G}(n)\boldsymbol{y}(n) - \boldsymbol{s}(n) \|_F^2 = E\left[\boldsymbol{s}(n)\boldsymbol{y}^H(n)\right]\left(E\left[\boldsymbol{y}(n)\boldsymbol{y}^H(n)\right]\right)^{-1} \\ &= \boldsymbol{\Lambda}^{1/2}(n)\widehat{\mathbf{H}}_k^H(n)\left(\widehat{\mathbf{H}}_k(n)\boldsymbol{\Lambda}(n)\widehat{\mathbf{H}}_k^H(n) + \boldsymbol{\Delta}(n)\right)^{-1}\end{aligned} \tag{16}$$

where the expectation is over $\mathbf{E}_k(n)$, $\boldsymbol{s}(n)$, $\boldsymbol{n}(n)$, and the effective noise covariance matrix is given by

$$\begin{aligned}\boldsymbol{\Delta}(n) &= E\left[\boldsymbol{v}(n)\boldsymbol{v}^H(n)\right] = E\left[\mathbf{E}_k(n)\boldsymbol{\Lambda}(n)\mathbf{E}_k^H(n)\right] + \sigma_n^2\mathbf{I}_{N_R} \\ &= \text{diag}\left(\text{tr}\left(\boldsymbol{\Sigma}_1(n)\boldsymbol{\Lambda}(n)\right), \ldots, \text{tr}\left(\boldsymbol{\Sigma}_{N_R}(n)\boldsymbol{\Lambda}(n)\right)\right) + \sigma_n^2\mathbf{I}_{N_R}\end{aligned} \tag{17}$$

where $\boldsymbol{\Sigma}_q(n) = E\left[\mathbf{E}_k^H(q,:,n)\mathbf{E}_k(q,:,n)\right] = \text{diag}\left(\epsilon^2(q,1,n), \ldots, \epsilon^2(q,k,n)\right)$ with $\mathbf{E}_k(q,:,n)$ as the q-th row of $\mathbf{E}_k(n)$ and $\epsilon^2(q,p,n)$ given in (13). The MSE for the optimization of $\boldsymbol{\Lambda}(n)$ is defined as

$$\begin{aligned}MSE(n) &= E\,\text{tr}\left[\left(\mathbf{G}_o(n)\boldsymbol{y}(n) - \boldsymbol{s}(n)\right)\left(\mathbf{G}_o(n)\boldsymbol{y}(n) - \boldsymbol{s}(n)\right)^H\right] \\ &= E\,\text{tr}\left(\mathbf{I}_k + \boldsymbol{\Lambda}^{1/2}(n)\widehat{\mathbf{H}}_k^H(n)\boldsymbol{\Delta}^{-1}(n)\widehat{\mathbf{H}}_k(n)\boldsymbol{\Lambda}^{1/2}(n)\right)^{-1} \\ &\le E\,\text{tr}\left(\mathbf{I}_k + \boldsymbol{\Lambda}^{1/2}(n)\widehat{\mathbf{H}}_k^H(n)\bar{\boldsymbol{\Delta}}^{-1}(n)\widehat{\mathbf{H}}_k(n)\boldsymbol{\Lambda}^{1/2}(n)\right)^{-1} \triangleq \overline{MSE}(n) \\ \bar{\boldsymbol{\Delta}}(n) &= \text{tr}\left(\boldsymbol{\Lambda}(n)\right)\text{diag}\left(\epsilon^2(1,n), \ldots, \epsilon^2(N_R,n)\right) + \sigma_n^2\mathbf{I}_{N_R}, \quad \epsilon^2(q,n) = \max_{p=1,\ldots,k}\epsilon^2(q,p,n)\end{aligned} \tag{18}$$

where the readers can refer to [25, 26] for the 2nd step, the inequality stems from $\bar{\boldsymbol{\Delta}}(n) \ge \boldsymbol{\Delta}(n)$ as well as the properties of matrix ordering [22], and the expectation is over $\boldsymbol{s}(n)$, $\boldsymbol{n}(n)$, $\mathbf{E}_k(n)$, and $\widehat{\mathbf{H}}_k(n)$, since the transmitter only knows channel statistics. The optimum $\boldsymbol{\Lambda}(n)$ is aimed at minimizing $MSE(n)$. However, the optimization may not be convex due to the term $\text{tr}\left(\boldsymbol{\Sigma}_q(n)\boldsymbol{\Lambda}(n)\right)$ in (17) and the inverse on $\boldsymbol{\Delta}(n)$ in (18). Therefore, $\boldsymbol{\Lambda}(n)$ is optimized by minimizing the upper bound of $MSE(n)$

$$\boldsymbol{\Lambda}_o(n,k) = \arg\min_{\boldsymbol{\Lambda}(n)} \overline{MSE}(n), \quad \text{s.t. } \text{tr}(\boldsymbol{\Lambda}(n)) = \rho \tag{19}$$

where the variable k emphasizes that the solution is conditioned on k. It can be shown with the techniques in [20] that (19) is convex programming and, therefore, the global minimizer can be reliably obtained via optimization routines.

C. Optimization of Transmission Rates

With the power-shaping matrix, we next optimize the transmission rates of the k streams. The SINR corresponding to the p-th received stream of $\widehat{\boldsymbol{s}}(n)=\mathbf{G}_o(n)\boldsymbol{y}(n)$ can be straightforwardly shown as

$$\gamma_p(n) = \lambda_p(n)\widehat{\mathbf{H}}_k^H(:,p,n)\left(\sum_{j\neq p}\lambda_j(n)\widehat{\mathbf{H}}_k(:,j,n)\widehat{\mathbf{H}}_k^H(:,j,n)+\boldsymbol{\Delta}(n)\right)^{-1}\widehat{\mathbf{H}}_k(:,p,n) \tag{20}$$

where $\widehat{\mathbf{H}}_k(:,p,n)$ represents the p-th column of $\widehat{\mathbf{H}}_k(n)$, and $\lambda_p(n)$ denotes the p-th diagonal element of $\boldsymbol{\Lambda}_o(n,k)$. With $\mathbf{A6}$, a tight BER approximation for QAM constellation with 2^i points is given by [27]: $BER(i,\gamma_p(n)) = 0.2\exp\left(-1.5\gamma_p(n)/(2^i-1)\right)$. Since the transmitter only knows channel statistics, the rate of the p-th stream in the n-th data symbol period is defined as the maximum number of bits keeping the average BER under the target BER

$$R_p(n,k) = \max_{i\in\{0,1,2,\dots\}} i, \quad \text{s.t. } \overline{BER}_p(n) = E\left[BER(i,\gamma_p(n))\right] \leq BER_{tar} \tag{21}$$

where the expectation is over $\widehat{\mathbf{H}}_k(n)$, and the variable k emphasizes that the solution is conditioned on k. Deriving a closed-form expression of $R_p(n,k)$ is difficult, since the distribution of $\gamma_p(n)$ for arbitrary channel statistics is unknown. The asymptotic distribution for i.i.d. $\widehat{\mathbf{H}}_k(n)$ with large number of antennas can be found in [28], and the distribution for i.i.d. $\widehat{\mathbf{H}}_k(n)$ with arbitrary number of antennas is given in [29] but is complicated for numerical evaluation. Therefore, $R_p(n,k)$ is numerically searched for general MIMO channels. However, closed-form solutions can be obtained for SISO, MISO, and SIMO channels with arbitrary correlations. They are ignored due to the space limitation.

D. Optimization of Data Stream Number and Data Block Length

The number of data streams k can be optimized to maximize the total rate in the n-th data symbol period

$$k_o(n) = \arg\max_{k\in S_1}\sum_{p=1}^{k} R_p(n,k), \quad S_1 = \{1,\dots,\min\{N_T,N_R\}\} \tag{22}$$

where $\min\{N_T,N_R\}$ represents the upper bound of k, as described in $\mathbf{A7}$. The optimized power shaping matrix and transmission rates associated with $k_o(n)$ are selected as the final designs

$$\boldsymbol{\Lambda}_o(n) = \boldsymbol{\Lambda}_o(n,k_o(n)), \quad \boldsymbol{R}_o(n) = \left[R_1(n,k_o(n)),\dots,R_{k_o(n)}(n,k_o(n))\right]^T. \tag{23}$$

To assess the performance, we define the average rate per packet as

$$AR(N_D) = \frac{\sum_{n=1}^{N_D} R_o(n)}{N_D + N_{tr}}, \quad R_o(n) = \sum_{p=1}^{k_o(n)} R_p(n, k_o(n)), \quad N_{tr} = \max_{n \in \{1,\ldots,N_D\}} k_o(n) \tag{24}$$

where $R_o(n)$ denotes the instantaneous total rate in the n-th data symbol period, and the number of training symbol periods N_{tr} is equal to the number of transmit angles involved in the data transmission. Note that N_{tr} could be less than N_T initially assumed in Section IV-A, since the data transmission may not use all N_T angles. The optimum data block length is the one maximizing the average rate

$$N_o = \arg \max_{N_D \in S_2} AR(N_D), \qquad S_2 = \{1, \ldots, N_{max}\} \tag{25}$$

where N_{max} represents the maximum length for searching N_o and can be predetermined by simulating typical channels in practice.

V. Adaptation with Instantaneous Feedback

The statistics-based design in Section III only requires channel statistics at the transmitter. However, most existing work [3]-[7] assumes the availability of Channel State Information at Transmitter (CSIT). It would be instructive to compare the statistics-based design to the design with CSIT. In the latter case, the transmitter is assumed to know the received training signal matrix $\mathbf{Z}$, which is perfectly fed back from the receiver without delay and error. Therefore, the transmitter can know the predicted virtual channel matrix $\widehat{\mathbf{H}}(n)$ via (10).

The transceiver with CSIT has the same structure as Fig.4 except that the power-shaping matrix and the zero-padding block are replaced by a linear precoder. The optimization is also the same as the statistics-based design except that the precoder and rates are optimized for each realization of $\widehat{\mathbf{H}}(n)$. For a fixed number of data streams k, the $N_R \times 1$ received signal vector can be written as

$$\boldsymbol{y}(n) = \mathbf{H}(n)\mathbf{F}(n)\boldsymbol{s}(n) + \boldsymbol{n}(n) = \widehat{\mathbf{H}}(n)\mathbf{F}(n)\boldsymbol{s}(n) + \boldsymbol{v}(n) \tag{26}$$

where $\mathbf{F}(n)$ is the $N_T \times k$ precoder, and $\boldsymbol{v}(n) = \mathbf{E}(n)\mathbf{F}(n)\boldsymbol{s}(n) + \boldsymbol{n}(n)$. Eqn (26) will reduce to (15) if

$\mathbf{F}(n)=[\mathbf{\Lambda}^{1/2}(n)\vdots\mathbf{0}_{k\times(N_T-k)}]^T$. The precoder is aimed at minimizing the following MSE

$$\begin{aligned} MSE\left(\widehat{\mathbf{H}}(n)\right) &= \text{tr}\left(\mathbf{I}_k+\mathbf{F}^H(n)\widehat{\mathbf{H}}^H(n)\mathbf{\Delta}^{-1}(n)\widehat{\mathbf{H}}(n)\mathbf{F}(n)\right)^{-1} \\ &\le \text{tr}\left(\mathbf{I}_k+\mathbf{F}^H(n)\widehat{\mathbf{H}}^H(n)\bar{\mathbf{\Delta}}^{-1}(n)\widehat{\mathbf{H}}(n)\mathbf{F}(n)\right)^{-1} \end{aligned} \tag{27}$$

which is derived in the same way as (18) except that the expectation is not on $\widehat{\mathbf{H}}(n)$, since the transmitter now knows $\widehat{\mathbf{H}}(n)$. Let $\mathbf{V}\mathbf{\Psi}\mathbf{V}^H$ be the eigen-decomposition of $\widehat{\mathbf{H}}^H(n)\bar{\mathbf{\Delta}}^{-1}(n)\widehat{\mathbf{H}}(n)$. The upper bound of $MSE\left(\widehat{\mathbf{H}}(n)\right)$ can be minimized by choosing $\mathbf{F}(n)=\mathbf{V}[\mathbf{\Phi}^{1/2}\vdots\mathbf{0}_{k\times(N_T-k)}]^T$ with the optimum power-shaping matrix $\mathbf{\Phi}=\text{diag}\,(\phi_1,\ldots,\phi_k)$ specified in [25]. Furthermore, the SINR of the p-th stream at the MMSE decoder output can be straightforwardly shown as

$$\gamma_p(n)=\mathbf{F}^H(:,p,n)\widehat{\mathbf{H}}^H(n)\left(\sum_{j\neq p}\widehat{\mathbf{H}}(n)\mathbf{F}(:,j,n)\mathbf{F}^H(:,j,n)\widehat{\mathbf{H}}^H(n)+\mathbf{\Delta}(n)\right)^{-1}\widehat{\mathbf{H}}(n)\mathbf{F}(:,p,n) \tag{28}$$

where $\mathbf{F}(:,p,n)$ denotes the p-th column of $\mathbf{F}(n)$. The maximum rate of the p-th stream under the BER requirement[4] is given by $R_p(n,k)=\lfloor\log_2\left(1-1.5\gamma_p(n)/\ln(BER_{tar}/0.2)\right)\rfloor$, and the optimum number of data streams is searched according to $k_o(n)=\arg\max_{k\in S_1}\sum_{p=1}^{k}R_p(n,k)$. In contrast to the statistics-based design, $\mathbf{F}(n)$, $\{R_p(n,k)\}$, and $k_o(n)$ are optimized for a given $\widehat{\mathbf{H}}(n)$ instead of its statistics. The corresponding average rate can be expressed as

$$AR(N_D)=\frac{\sum_{n=1}^{N_D}R_o(n)}{N_D+N_{tr}},\qquad R_o(n)=E\left(\sum_{p=1}^{k_o(n)}R_p(n,k_o(n))\right) \tag{29}$$

where the expectation is on $\widehat{\mathbf{H}}(n)$, and the training block length N_{tr} is equal to the number of non-vanishing transmit virtual angles[5]. This is because unlike the statistics-based design, the precoder and rates are optimized for each estimated $\mathbf{H}(n)$. To estimate $\mathbf{H}(n)$, the pilot symbols have to be sent from all non-vanishing transmit virtual angles.

[4]Mathematically, $R_p(n,k)=\max_{i\in\mathbb{Z}} i,$ s.t. $BER(i,\gamma_p(n))\le BER_{tar}$ with $BER(i,\gamma_p(n))$ defined above (21).

[5]which equals the number of columns of $\mathbf{H}(n)$ whose Frobenious norms averaged over channel statistics are not zero.

VI. Results and Discussions

A. Simulation Parameters and Procedure

We consider $L = 100$ paths whose arrival and departure angles $\{\theta_{R,l}, \theta_{T,l}\}$ are randomly uniformly distributed within a 2D angular region in the virtual domain: $[-\theta_{\max}, \theta_{\max}] \times [-\theta_{\max}, \theta_{\max}]$, where $\theta_{\max}$ determines the angular spread. Each path has the same strength $\alpha_l^2 = 1/L$, which implies the channel normalization $E \| \mathbf{H}_c(t) \|_F^2 = N_R N_T$. Both the transmitter and receiver have the same speed $v_T = v_R = 10$km/h. The carrier frequency is 1.8GHz, and the resultant maximum Doppler shift is $f_{\max} = 33.3$Hz. The symbol period T_s is specified via the product $f_{\max} T_s$, which determines the fading rate. The BER target is 10^{-3}, the noise power per receive antenna σ_n^2 is 1, and the other parameters N_R, N_T, $\theta_{\max}$, $f_{\max} T_s$, ρ are specified in the results.

Based on the above physical parameters, $r_{q,p}(t - t')$ is calculated via the exact expression in (8) for all (q, p)'s, and both $\widehat{\sigma}^2(q, p, n)$ and $\epsilon^2(q, p, n)$ are computed by (13), which determines the statistics of $\widehat{\mathbf{H}}(n)$ and $\mathbf{E}(n)$. The statistics-based design follows the procedure in Section IV, where the power-shaping matrix and transmission rates are optimized with 10^3 realizations of $\widehat{\mathbf{H}}(n)$. For the design with CSIT, the precoder and rates are optimized for each realization of $\widehat{\mathbf{H}}(n)$.

B. Instantaneous and Average Rates

The results of instantaneous rate $R_o(n)$ for the statistics-based design are shown in Fig.5, where $\rho = 30$dB, $f_{\max} T_s = 10^{-2}$, and $\theta_{\max} = 0.5$. The "MIMO, $\mathbf{\Lambda}_o(n)$" and "MIMO, $\mathbf{\Lambda}_o(n, N_T)$" represent the performance with $\mathbf{\Lambda}_o(n)$ in (23) and $\mathbf{\Lambda}_o(n, N_T)$ in (19), respectively. It is obvious that $R_o(n)$ decays as the time index n increases in each case. This can be understood from (13), which indicates that larger n impairs the power[6] of estimated channel coefficient $\widehat{\sigma}^2(q, p, n)$ while boosts that of estimation error $\epsilon^2(q, p, n)$ due to the vanishing temporal correlation $r_{q,p}(nT_s)$. Accordingly, the SINR of each data stream in (20) and hence $R_o(n)$ diminish as time progresses. The corresponding average rate $AR(N_D)$ is plotted in Fig.6. As N_D increases, $AR(N_D)$ in each case goes up before reaching the peak and then

[6] which is the variance of the estimable part and represents the estimation quality.

gradually decreases. This is because smaller N_D reduces the portion of data transmission time in each packet and hence degrades transmission efficiency, while larger N_D extends data transmission time at the cost of lower $R_o(n)$ in the extended period, which eventually brings down $AR(N_D)$.

Next, we study the impact of number of data streams on the MIMO performance, which is best revealed by the comparison of $\mathbf{\Lambda}_o(n, N_T)$ and $\mathbf{\Lambda}_o(n)$. The former distributes power over all N_T transmit angles, while the latter allocates power only to the $k_o(n)$ strongest angles. In each case, the number of excited data streams is equal to the rank of the power-shaping matrix. Compared with $\mathbf{\Lambda}_o(n)$, $\mathbf{\Lambda}_o(n, N_T)$ induces a 55% to 100% reduction of $R_o(n)$ in Fig.5 and a 64% to 71% reduction of $AR(N_D)$ in Fig.6. The ranks of both power-shaping matrices are plotted in Fig.7, which shows that $\mathbf{\Lambda}_o(n, N_T)$ has a fixed rank of 11, while the rank of $\mathbf{\Lambda}_o(n)$ ranges between 4 and 8. Therefore, activating all 11 data streams with less rate per stream is inferior to concentrating power on fewer streams with more power and rate in each.

The performance of non-MIMO configurations is investigated next. The instantaneous rate $R_o(n)$ for MISO and SISO is plotted in Fig.5, where MISO achieves improvement over SISO due to the array gain. However, the improvement is not significant, since multiplexing doesn't exploit transmit diversity, which can be achieved via space-time codes [30, 31]. Compared with MISO and SISO, SIMO achieves significant improvement in both $R_o(n)$ and $AR(N_D)$. This is because SIMO captures more channel power by the N_R receive antennas, and the received SNR is stabilized by the antenna diversity. In addition, SIMO has a lower training cost than MIMO due to the use of single transmit antenna. Both antenna diversity and low training cost make the maximum $AR(N_D)$ of SIMO roughly match that of MIMO with $\mathbf{\Lambda}_o(n, N_T)$ in Fig.6. In fact, MIMO with $\mathbf{\Lambda}_o(n, N_T)$ and SIMO correspond to the full-multiplexing and full-diversity schemes, respectively, while MIMO with $\mathbf{\Lambda}_o(n)$ makes a judicious diversity-multiplexing tradeoff by choosing the optimum number of data streams. As demonstrated in Fig.5 and 6, the tradeoff brings significant improvement over SIMO and MIMO with $\mathbf{\Lambda}_o(n, N_T)$.

Finally, it would be instructive to compare the statistics-based design to that with CSIT, whose performance is represented by "MIMO, CSIT". It can be observed in Fig.5 that MIMO with CSIT

on average achieves a 35% improvement in $R_o(n)$ over MIMO with $\mathbf{\Lambda}_o(n)$, and the improvement in $AR(N_D)$ is on average 24% as shown in Fig.6. These results indicate that the loss may be acceptable if there is no instantaneous channel state feedback.

C. Optimum Data Block Length

As shown in Fig.6, the average rate $AR(N_D)$ is maximized by an optimum data block length N_D, which is defined as N_o in (25). It can be observed that N_o for MIMO with $\mathbf{\Lambda}_o(n)$, SIMO, MISO, and SISO is 22, 6, 6, and 2, respectively. This decreasing order can be intuitively explained via the virtual path partitioning in Fig. 2. In case of MIMO, the paths are partitioned by both the transmit and receive beams and, therefore, each virtual channel coefficient in beamspace captures fewer paths with smaller Doppler spread, compared to those in SIMO, MISO, and SISO cases[7]. The smaller Doppler spread results in slower decay of temporal correlation in virtual domain, which yields better channel prediction and hence longer N_o, as implied by (13). In short, MIMO has finer resolution in beamspace, which effectively slows down the temporal channel variation. For the same reason, the N_o for SIMO or MISO is longer than that for SISO due to the finer path partitioning. The idea of reducing the temporal channel variation by beamforming has been reported in [32, 33]. In this work, the benefit of the slowed variation is accomplished by extending the data block length, which improves the rate by reducing the training update frequency.

The N_o is also affected by the number of antennas. Fig.8 shows the average rate when the maximum number of antennas at one side is 7 for MIMO, SIMO, and MISO. The N_o for MIMO, SIMO, MISO, and SISO is found to be 15, 6, 5, and 2, respectively. Compared with Fig.6, the smaller antenna number decreases the average rate in each case and significantly reduces N_o for MIMO. This is because fewer antennas enlarge the size of each virtual angular bin $S_{q,p}$ as well as the associated path Doppler spread, which reduces the temporal correlation and hence N_o.

The fading rate affects N_o as well. Fig.9 shows the average rate for $f_{\max}T_s = 10^{-1}$. Compared

[7]For instance, the paths in $S_{2,2}$ are fewer than those in $S_{R,2}$, $S_{T,2}$ and those in the whole $\theta_R \times \theta_T$ domain in Fig. 2.

with Fig.6, the higher fading rate significantly reduces N_o, which[8] is 8, 3, and 1 for MIMO, SIMO, and MISO, since the training signal has to be sent more frequently to track the channel variation. The higher fading rate also impairs the average rate in each case due to the larger channel estimation error.

D. Accuracy of Virtual Channel Model

As described in $\mathbf{A3}$ and $\mathbf{A4}$ in Section IV, the virtual channel coefficients are modeled as independent Gaussian entries. To assess the modeling accuracy, the rate performance is rigorously simulated based on the physical model (2), which does not impose any assumptions on the statistics of virtual coefficients. The physical-model-based statistics are used to optimize the transceiver components in the same way as described in Section IV. As shown in Fig.10, the good agreement between the performances of the two verifies the accuracy of virtual channel model. In contrast to the physical model, the virtual model provides insights into the system design: the capacity-optimum signaling corresponds to transmitting independent streams from different virtual angles, and the optimum data block length is determined by the temporal correlation in beamspace.

VII. Conclusions

This work proposes a statistically adaptive spatial multiplexing scheme for correlated time-varying MIMO channels based on the virtual channel representation. With the knowledge of channel statistics, the transmitter adjusts the power and rate for each data stream in each symbol period, and the data block length is further optimized to maximize the average rate. Major results for the performance of instantaneous and average rates are summarized below.

- For each antenna configuration, the instantaneous rate $R_o(n)$ decreases for larger time index n, while the average rate $AR(N_D)$ is usually a hill-shape function implying an optimum N_D;
- The rate-maximizing power-shaping matrix $\mathbf{\Lambda}_o(n)$ makes a judicious diversity-multiplexing trade-off by exciting the optimum number of data streams, which significantly improves the rate;

[8]The average rate is always zero in case of SISO.

- Compared with the statistics-based design, the design with CSIT on average improves $R_o(n)$ and $AR(N_D)$ by 35% and 24%, respectively. Therefore, the former might be a good complexity-performance tradeoff by avoiding the instantaneous feedback of CSIT, which will both reduce the resource and suffer from the imperfections in practice;

Interesting results on the behavior of optimum data block length N_o are summarized below.

- N_o increases according to the order SISO $<$ $\{$SIMO, MISO$\}$ $<$ MIMO and increases for more antennas in each case. This is because each virtual channel coefficient captures fewer paths with smaller Doppler spread due to the finer path partitioning in beamspace. The smaller spread effectively slows down the channel variation in beamspace and hence yields a longer N_o;
- Higher fading rate factor $f_{\max}T_s$ yields smaller N_o and reduces both $R_o(n)$ and $AR(N_D)$;

References

[1] I.E. Telatar, "Capacity of multi-antenna gaussian channels," Tech. Rep., AT&T Bell Labs., 1995.

[2] G.J. Foschini, M.J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," Wirless Personal Commun., vol. 6, no. 3, pp. 311-335, Mar. 1998.

[3] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," IEEE Trans. on wireless Communications, vol. 4, no. 2, pp. 757-764, March 2005.

[4] N. Khaled, G. Leus, C. Desset, and H. De Man, "A robust joint linear precoder and decoder MMSE design for slowly time-varying MIMO channels," in Proc. IEEE ICASSP, pp. 485-488, May 2004.

[5] Y. Ko and C. Tepedelenlioglu, "Space-time block coded rate-adaptive modulation with uncertain SNR feedback," in Proc. IEEE Conf. on Signals, Systems and Computers, pp. 1032-1036, Nov. 2003.

[6] S. Zhou and G.B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," IEEE Trans. on Wireless Commun., vol. 3, no. 5, pp. 1626-1636, Sep. 2004.

[7] S. Zhou and G.B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" IEEE Trans. on wireless Communications, vol. 3, no. 4, pp. 1285-1294, July 2004.

[8] M. Baissas and A.M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," IEEE Trans. on Signal Processing, pp. 2037-2049, August 2002.

[9] M. Dong and L. Tong, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," IEEE Trans. on Signal Processing, vol. 52, no. 5, pp. 1403-1417, May 2004.

[10] S. Ohno and G.B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," IEEE Trans. Wireless Communications, vol. 1, pp. 712-720, Oct. 2002.

[11] A. Boariu, "Effect of delayed commands on error probability and throughput capacity in a multiple access system," in Proc. IEEE VTC, Orlando, USA, pp. 262-265, Oct. 2003.

[12] A.E. Ekpenyong and Y.F. Huang, "Markov channel-based feedback schemes for adaptive modulation systems," in Proc. IEEE GLOBECOM, pp. 1091-1095, Nov. 2004.

[13] C.J. Jakes, Microwave Mobile Communications, Wiley, 1974.

[14] J. D. Parsons, The Mobile Radio Propagation Channel, John Wiley & Sons, 2000.

[15] A.M. Sayeed, "Deconstructing multi-antenna fading channels," IEEE Trans. Signal Processing, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.

[16] A.M. Sayeed, "A virtual representation for time- and frequency-selective correlated MIMO channels," in Proc. IEEE ICASSP, vol. 4, pp. 648-651, 2003.

[17] Y. Zhou and A.M. Sayeed, "Experimental study of MIMO channel statistics and capacity via virtual channel representation," submitted to IEEE Trans. Wireless Communications, June 2005.

[18] S. Haykin, Adatptive Filter Theory. Prentice Hall: New Jersey, 1996.

[19] J. Kotecha and A.M. Sayeed, "Optimal estimation of correlated MIMO channels," IEEE Trans. on Signal Processing, vol. 52, no. 2, pp. 546-557, Feb. 2004.

[20] V.V. Veeravalli, Y. Liang, and A.M. Sayeed, "Correlated MIMO rayleigh fading channels: capacity, optimal signaling, and asymptotics," IEEE Trans. on Info. Theo., vol. 51, no. 6, pp. 2058-2072, June 2005.

[21] J. Kotecha and A.M. Sayeed, "Canonical statistical models for correlated MIMO fading channels and capacity analysis," submitted to IEEE Trans. on Wireless Communications.

[22] R.A. Horn, Matrix Analysis. Cambridge University Press, 1985.

[23] H.V. Poor and S.V. Verdu, ”Probability of error in MMSE multiuser detection,” IEEE Trans. on Info. Theo., vol. 43, pp. 858-871, May 1997.

[24] A. Scaglione, P. Stoica, S. Barbarossa, G.B. Giannakis, and H. Sampath, ”Optimal Designs for Space-time Linear Precoder and Decoder,” IEEE Trans. on SP, vol. 50, pp. 1051-1064, May 2002.

[25] A. Scaglione, G.B. Giannakis, and S. Barbarossa, ”Redundant filterband precoder and equalizers Part I: unification and optimal designs” IEEE Trans. on SP, vol. 47, pp. 1988-2005, July 1999.

[26] D.P. Palomar, M.A. Lagunas, and J.M. Cioffi, ”Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints,” IEEE Trans. on SP, vol. 52, pp. 1179-1197, May 2004.

[27] A. Goldsmith and S.G. Chua, ”Variable-rate variable-power MQAM for fading channels,” IEEE Trans. on Communications, vol. 45, pp. 1218-1230, Oct. 1997.

[28] D.N.C. Tse and O. Zeitouni, ”Linear multiuser receivers in random environments,” IEEE Trans. on Info. Theo., vol. 46, pp. 171-188, Jan. 2000.

[29] H. Gao, P.J. Smith, and M.V. Clark, ”Theoretical reliability of MMSE linear diversity combining in rayleigh-fading additive interference channels,” IEEE Trans. on Commun., vol. 46, pp. 666-672, May 1998.

[30] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, ”Space-time block codes from orthogonal designs,” IEEE Trans. on Info. Theo., vol. 45, pp. 1456-1467, July 1999.

[31] V. Tarokh, N. Seshadri, and A.R. Calderbank, ”Space-time codes for high data rate wireless communication: Performance criterion and code construction” IEEE Trans. on Info. Theo., vol. 44, pp. 744-765, Mar 1998.

[32] D. Chizhik, ”Slowing the time-fluctuating MIMO channel by beam forming,” IEEE Trans. on Wireless Communications, vol. 3, pp. 1554-1565, Sep. 2004.

[33] R.G. Vaughan, ”Angular partitioning to yield equal Doppler contributions” IEEE Trans. on Veh. Technol., vol. 48, pp. 1437-1442, Sep. 1999.

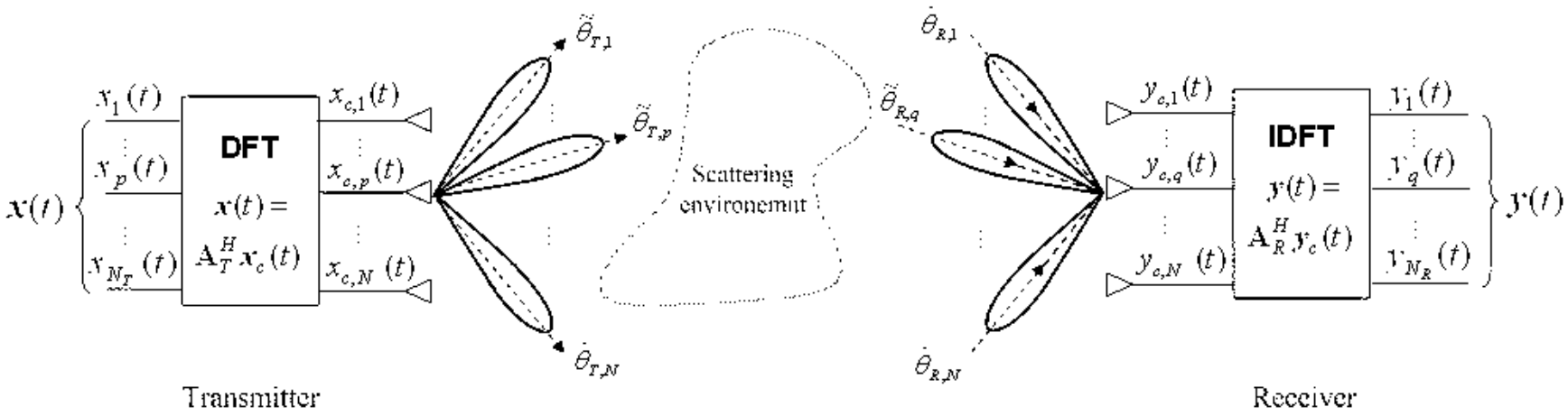

Fig. 1: Illustration of signaling in virtual angle domain.

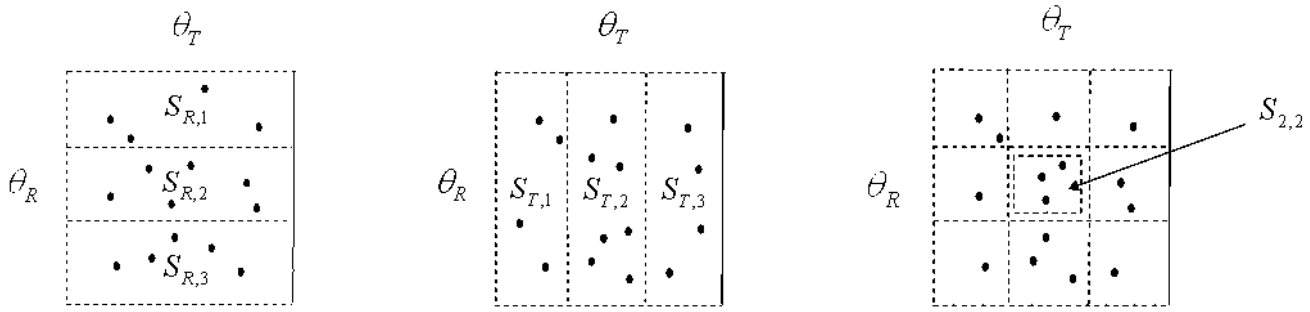

Fig. 2: Illustration of virtual path partitioning $(N_T = N_R = 3)$.

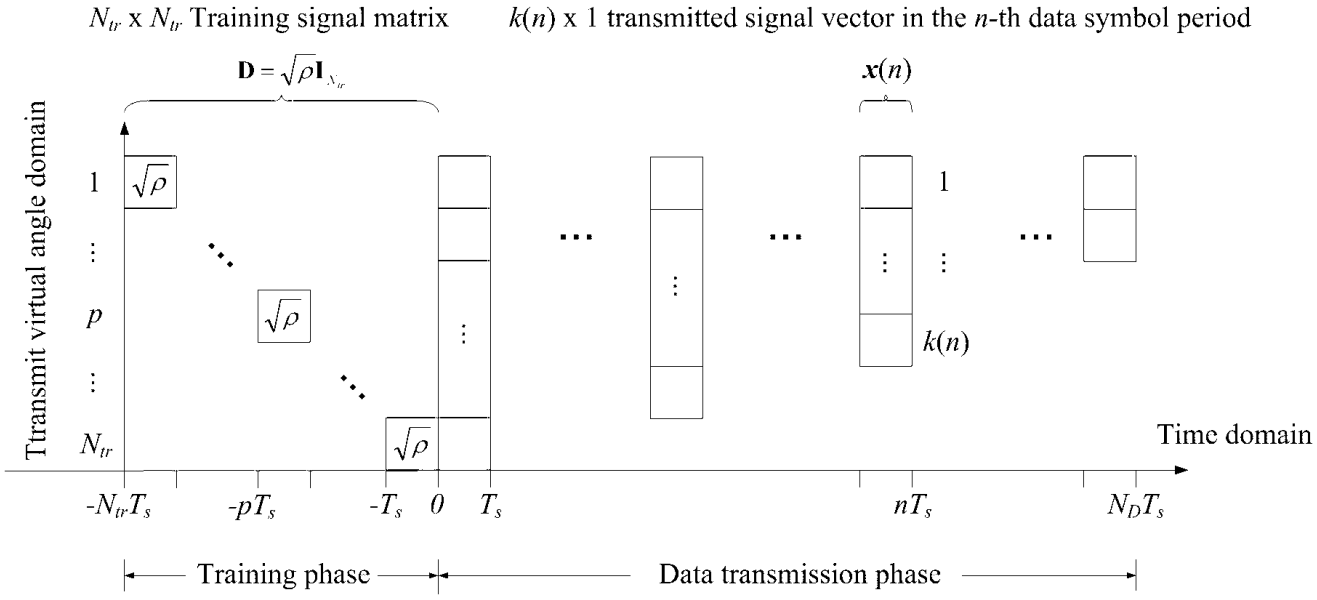

Fig. 3: Format of transmitted signal in the first packet.

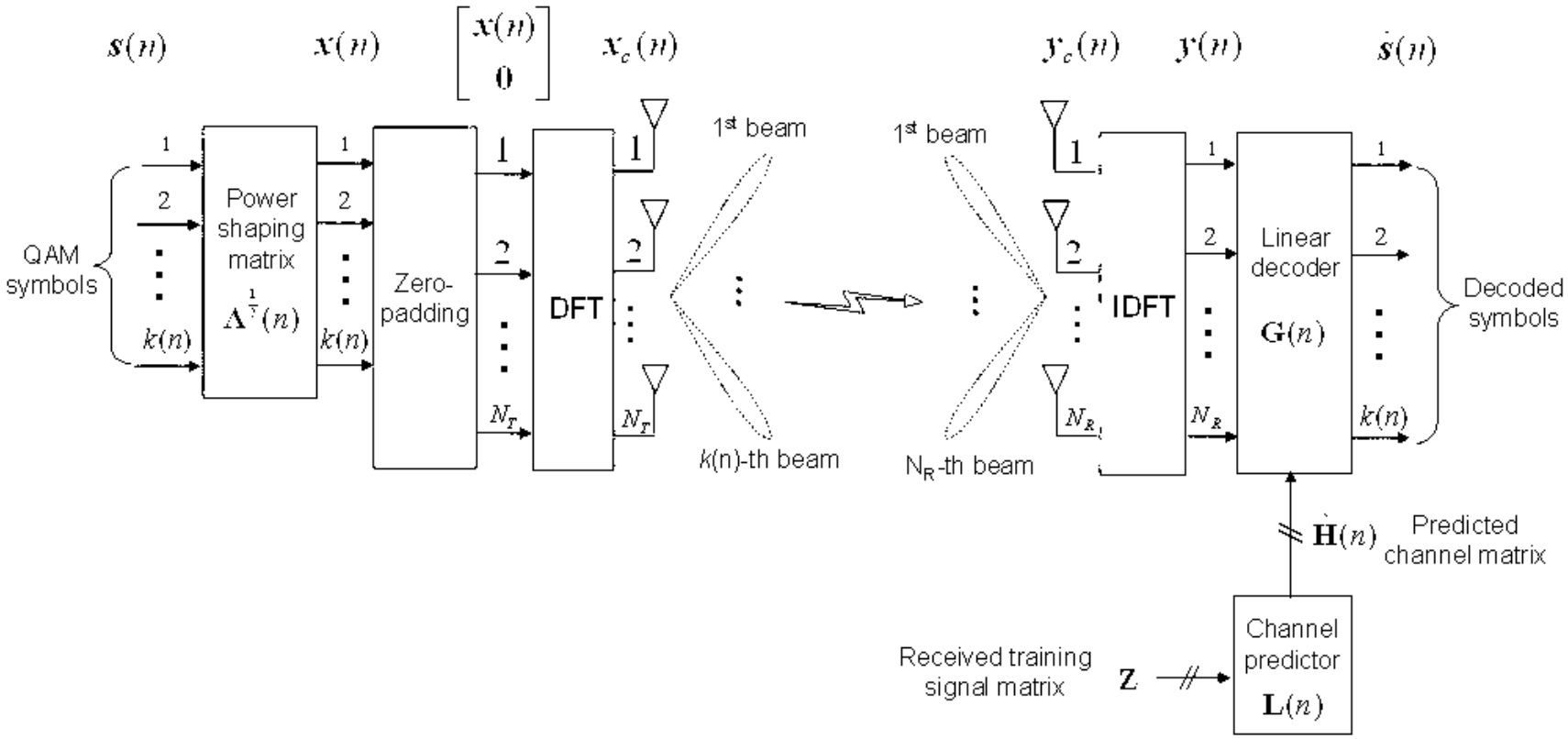

Fig. 4: Block diagram of the proposed system.

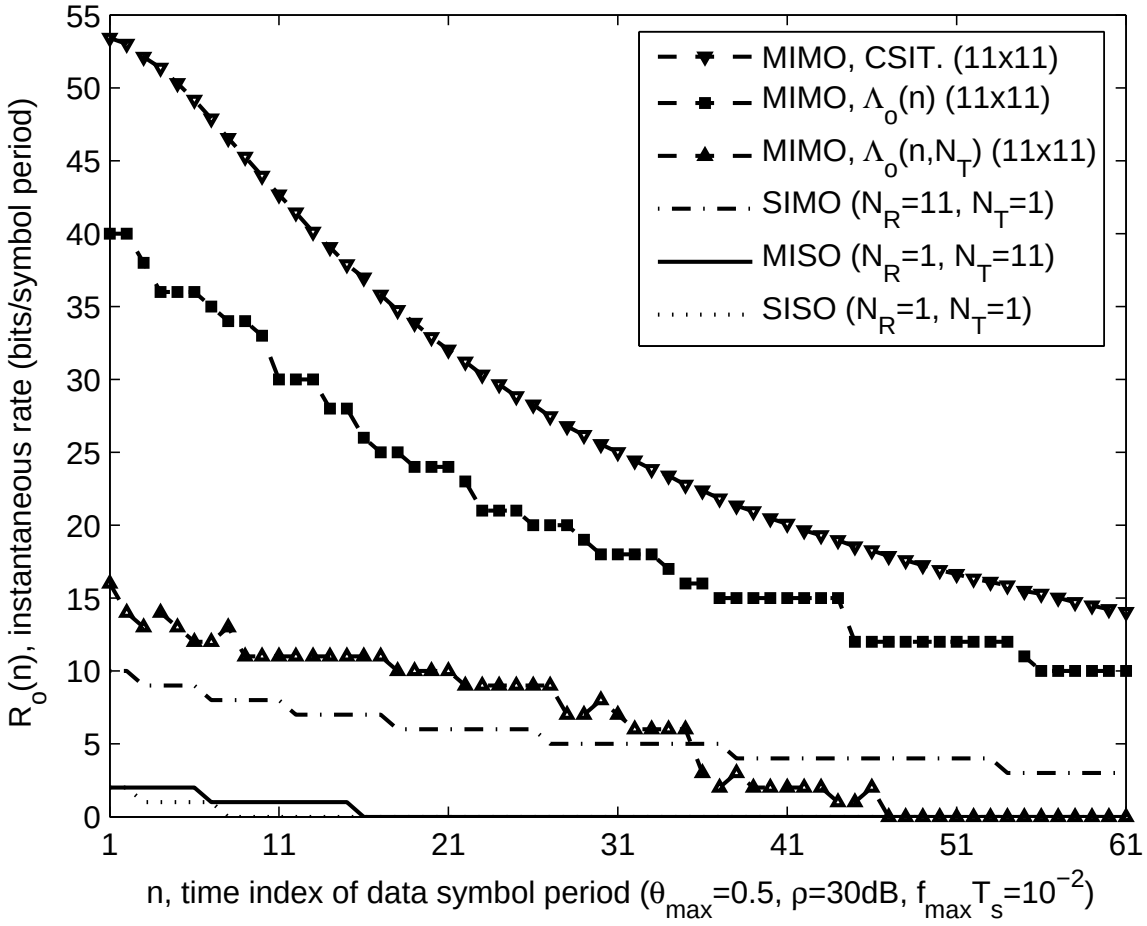

Fig. 5: Instantaneous rate in each data symbol period for angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-2}$.

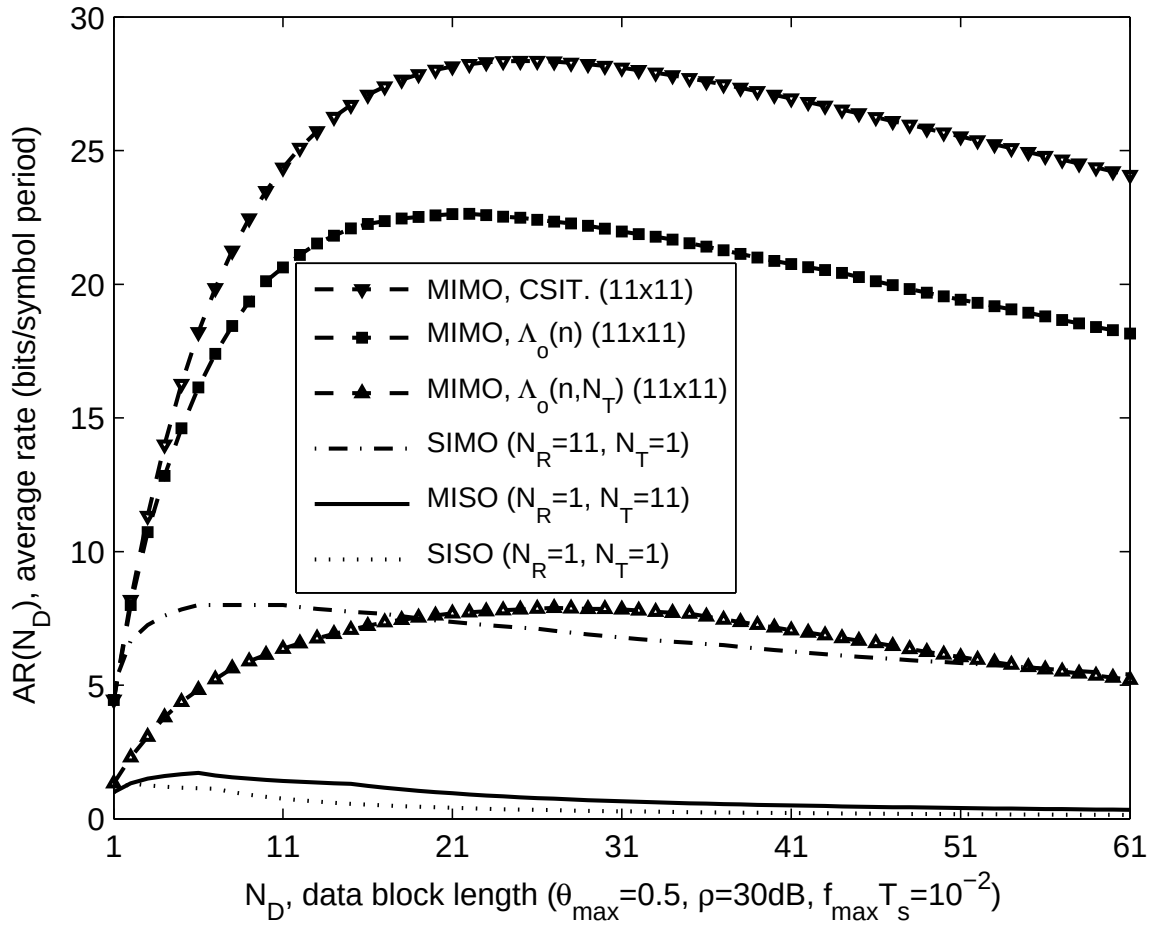

Fig. 6: Average rate in each packet for angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-2}$.

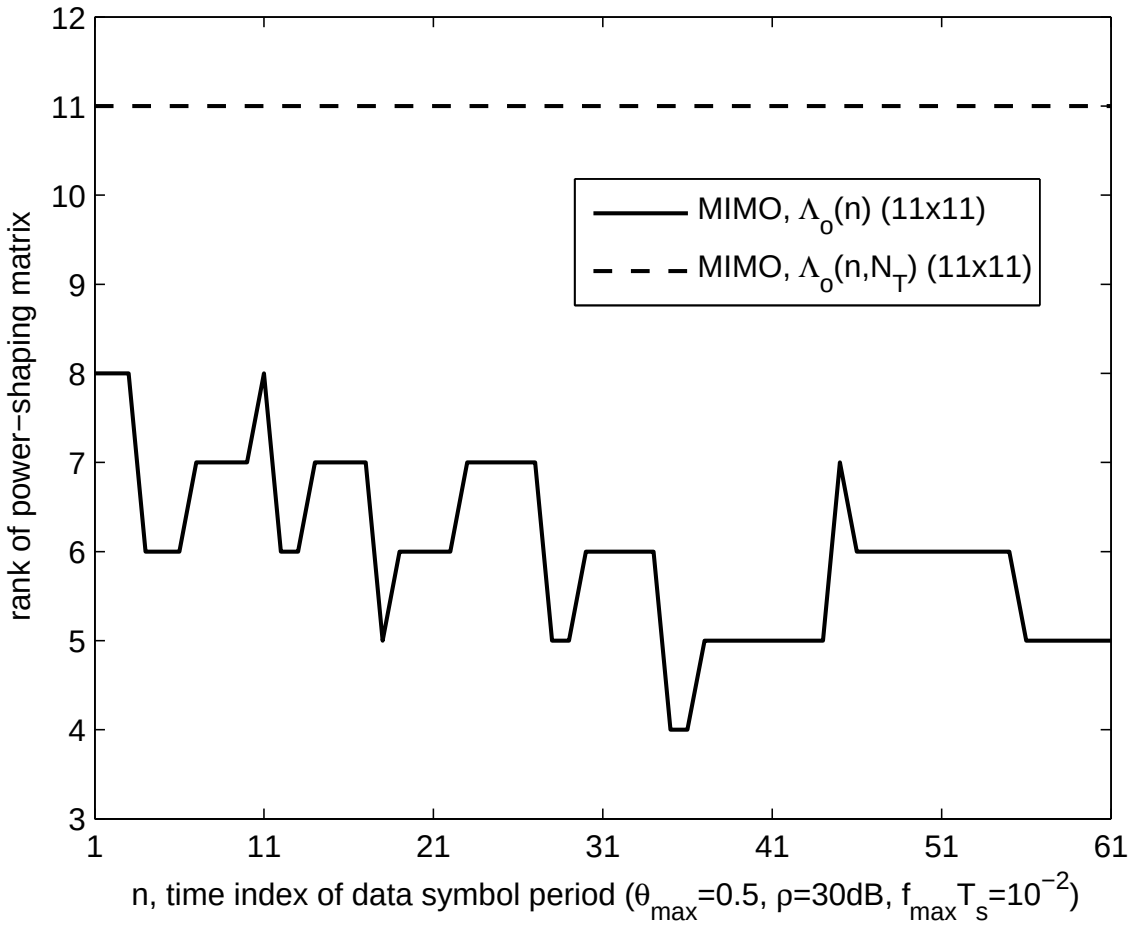

Fig. 7: Ranks of power-shaping matrices for angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-2}$.

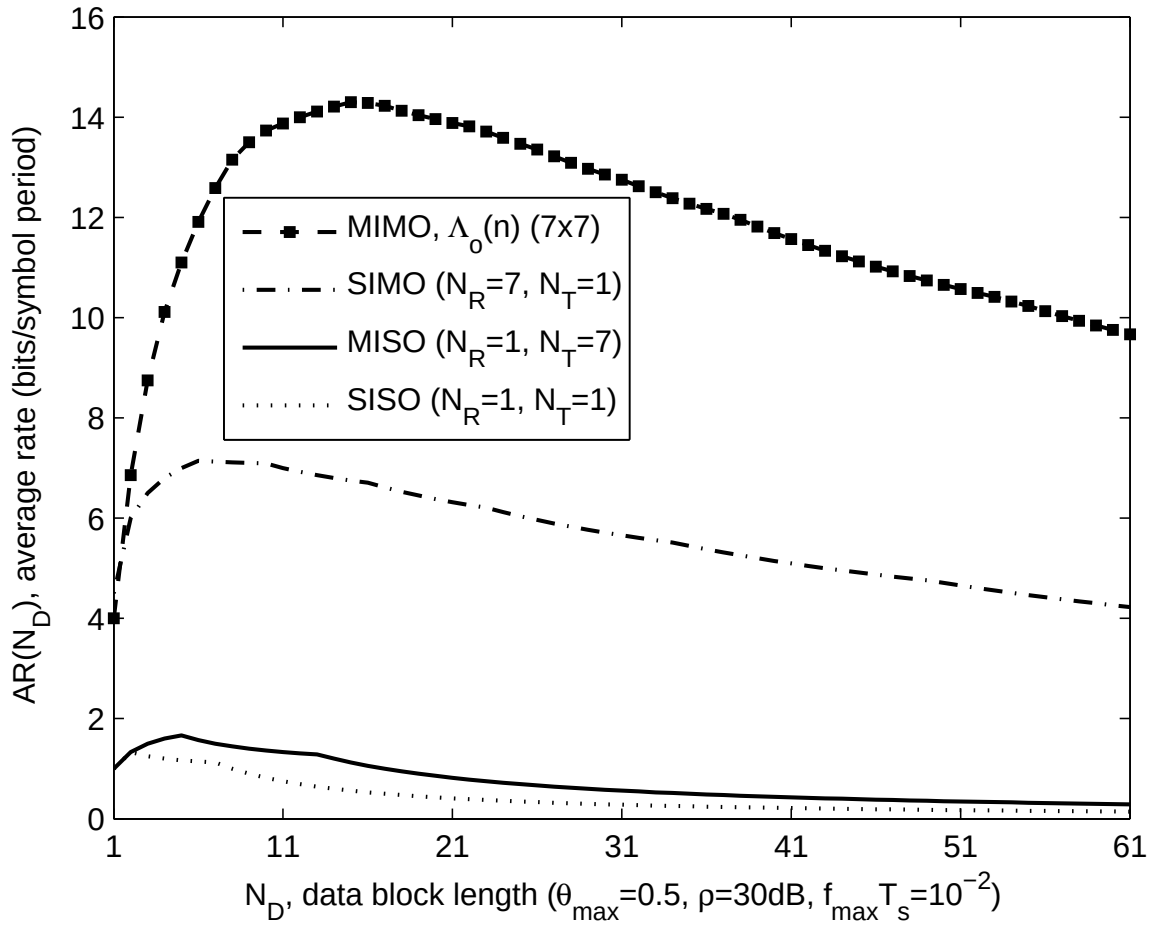

Fig. 8: Average rate in each packet for angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-2}$.

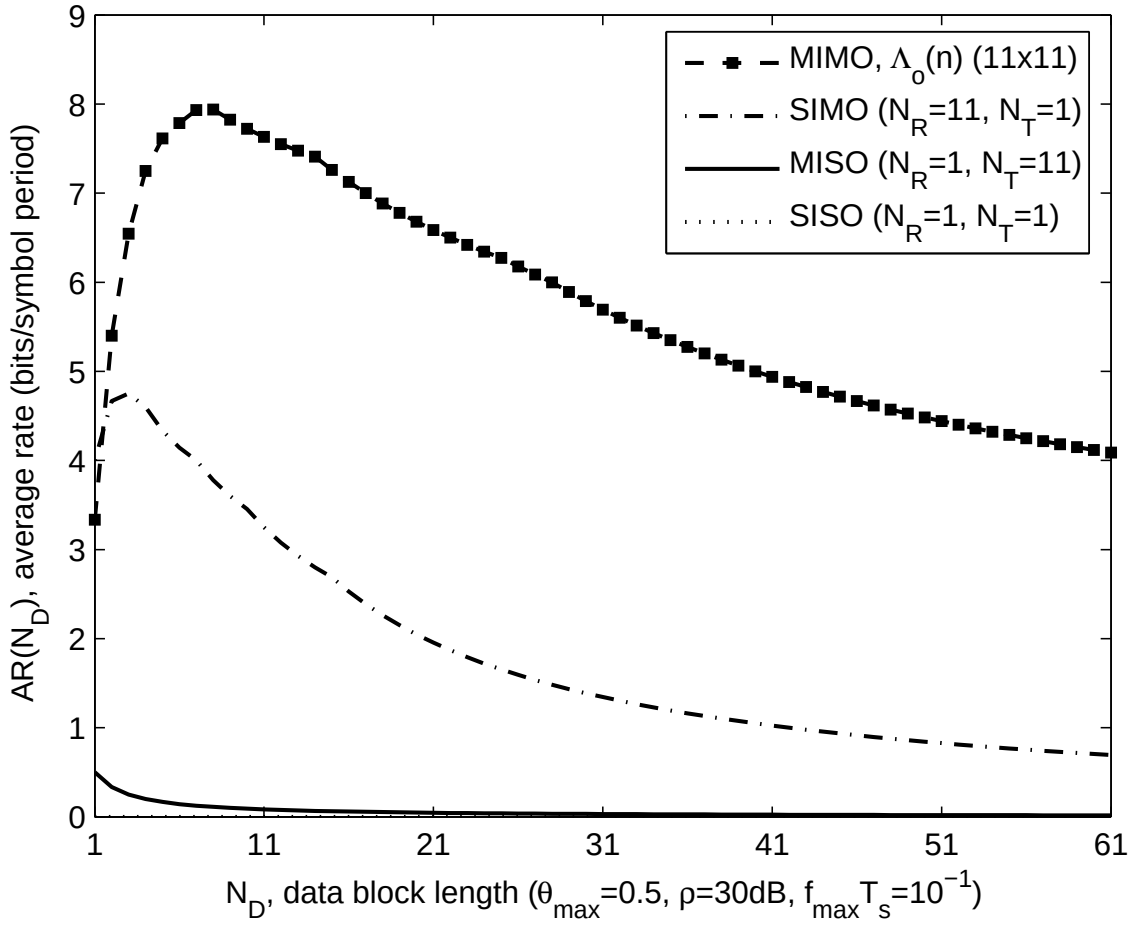

Fig. 9: Average rate in each packet for angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-1}$.

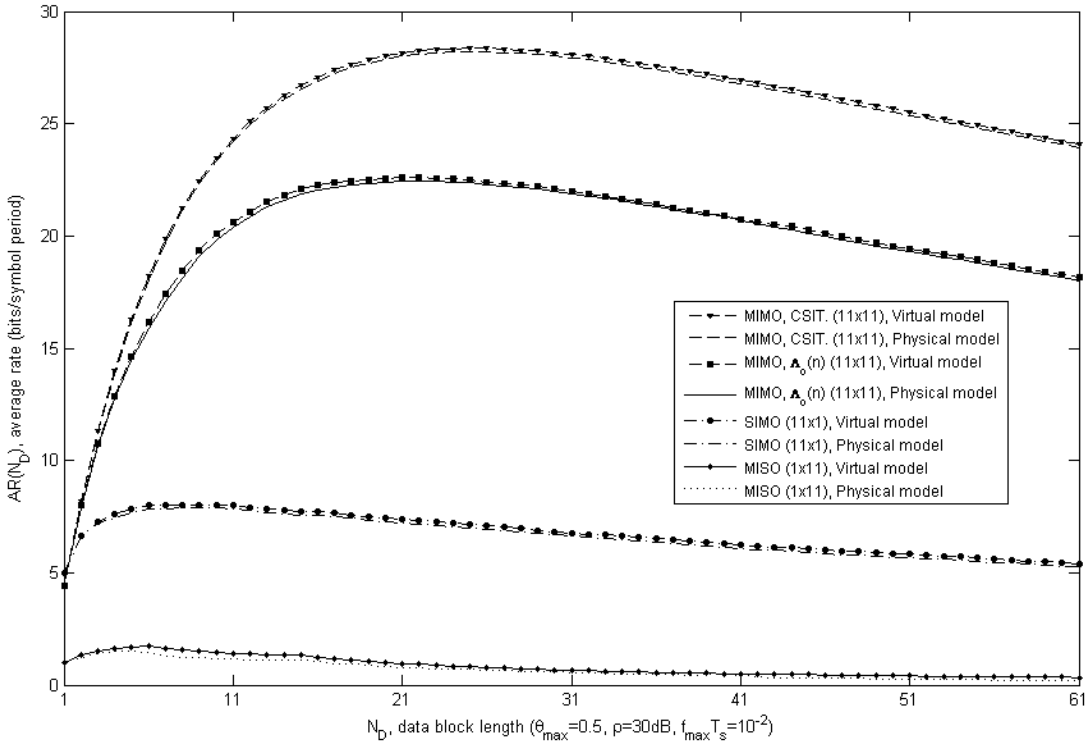

Fig. 10: Average rate comparison for virtual and physical models, angular spread $\theta_{\max} = 0.5$, transmit SNR $\rho = 30$dB, fading rate factor $f_{\max}T_s = 10^{-2}$.