Statistical Distributions that Arise in Constrained Beamforming with Incomplete Channel Information

Bertrand M. Hochwald
Beceem Communications Inc.
Santa Clara, CA 95054, USA
Email: hochwald@beceem.com

Louay M.A. Jalloul
Beceem Communications Inc.
Santa Clara, CA 95054, USA
Email: jalloul@beceem.com

***Abstract*—We examine the statistical distribution of received power when a subscriber terminal with multiple transmit antennas beamforms on the uplink of a wireless orthogonal frequency-division multiplexing system to a basestation with multiple receive antennas. We assume that a time-division duplex terminal has incomplete uplink channel knowledge obtained by measuring the downlink channel during basestation transmissions. The basestation transmissions employ cyclic delay diversity, of which the terminal is unaware. The distribution of uplink power received at the basestation, with and without downlink cyclic delay diversity and with various types of power constraints at the terminal, allow us to make some surprising conclusions about how the power constraints and cyclic delay influence the performance of beamforming. We show that beamforming gains need to be examined critically beyond simple average performance metrics to conclude that they are superior to simpler channel-unaware methods. We also conclude that cyclic delay on the downlink is very beneficial to beamforming on the uplink, even with restrictive types of power constraints.**

I. Introduction

Multiple antenna technology at the transmitter and receiver (MIMO or multiple-input multiple-output) was developed to support the demand for high information data rate services. Fourth-generation mobile broadband standards have employed at least two antennas at a basestation and two antennas at a terminal in an effort to double spectral efficiencies over comparable third-generation products.

MIMO is typically applied to the downlink transmission from the base station (BS) to the mobile station (MS) or terminal. Examples of MIMO schemes include space-time or space-frequency block coding using the Alamouti code [1], spatial-multiplexing with multiple streams, and beamforming techniques that exploit channel knowledge at the transmitter [2].

Traditional terminal transceiver radio architectures use a single transmit antenna on the uplink to the basestation, and two or more receive antennas for downlink reception. Since the basestation can transmit at a power fifty-times higher than the terminal there is an inherent imbalance created between uplink and downlink coverage and data rates [3]. The downlink often enjoys better coverage than the uplink and thus the system becomes uplink limited. Using the second antenna at the terminal in an uplink MIMO transmission can help close the gap in coverage between the two links.

MIMO on the uplink can take two forms: 1) channel-unaware (sometimes referred to as open-loop); 2) channel-aware (closed-loop). Channel-unaware methods include virtualization techniques such as cyclic delay diversity (CDD) Channel-aware transmissions include antenna selection and beamforming.

Time-division duplex systems that employ the same frequency on downlink and uplink such as obtained in the IEEE 802.16e/WiMAX standard [4], [5] and [6], have a big advantage in providing channel information almost "for free" in the sense that the downlink and uplink channels are reciprocal (after calibrating for receive and transmit chain imperfections). Hence any information learned about the channel on the downlink during reception can be turned around and used on the uplink during transmission. Since channel information is usually obtained as part of standard demodulation and decoding of downlink signals, the uplink channel information is often available without special training signals or feedback.

This reciprocity is especially useful for terminals at a cell edge that are barely clinging to the network because of their limited uplink power. Antenna selection is an example of a simple method that exploits the fact that the antenna with the strongest signal on the downlink is also the best to use on the uplink. This has been used to great effect, especially when the basestation uses cyclic delay diversity on the downlink [7].

However, this paper seeks to improve beyond antenna selection by beamforming on the uplink. The channel between the terminal and basestation may be modeled by an $N \times M$ matrix $\mathbf{H}(f)$, where N is the number of antennas at the basestation, M is the number of antennas at the terminal, and f is the carrier frequency. When the channel is frequency-flat, we may drop the dependence on f.

If the terminal has full knowledge of the channel $\mathbf{H}$ then the terminal can perform the singular value decomposition $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}$ and beamform by applying the first column of $\mathbf{V}$. The result is receive power at the basestation that is proportional to σ_1^2, the largest singular value, and $\sigma_1^2 \approx (\sqrt{N} + \sqrt{M})^2$ for large M and N [8]. Increasing either N or M has a beneficial effect.

However, it is unlikely that the terminal learns $\mathbf{H}$ in its entirety. In the IEEE 802.16e/WiMAX standard, the basestation transmits five-millisecond "frames" that begin with a "preamble" which contains pilot signals known to the terminal.

These pilots are typically transmitted from either one of the N antennas available to the basestation, or with cyclic delay diversity using all N antennas. The terminal is obliged to learn what it can about $\mathbf{H}$ from these pilots, and typically does not know if cyclic delay is employed at the basestation.

The terminal generally must form its uplink beam using the incomplete knowledge it can glean about $\mathbf{H}$ from the downlink preamble. Because the terminal has M antennas, it has an $M \times 1$ downlink channel to work with to form its uplink beam. In the case where the basestation transmits from only one antenna, only the first row of $\mathbf{H}$ is learned.

It is assumed that the beam transmitted by the terminal on the uplink is simply the conjugate of the channel obtained on the downlink. However, several different types of constraints are employed to normalize the transmitter output power, some of which are easily realizable in practice while others are more difficult. An example of a constraint that is easy to realize is a per-antenna power constraint at the terminal; an example not easy to realize, but mathematically convenient, is the sum-power constraint across all M antennas. The constraints we consider include:

- Average power constraint: The expected transmitted power of an OFDM symbol summed over all antennas is constrained. The instantaneous symbol, tone (subcarrier), and antenna powers are unconstrained.
- Per-symbol constraint: The instantaneous symbol power summed over all tones and over all antennas is constrained. The instantaneous power of a tone or antenna is unconstrained.
- Per-tone constraint: The instantaneous power per tone summed over all antennas is constrained. The instantaneous power per antenna is unconstrained.
- Per-symbol, per-antenna constraint: The instantaneous power per symbol and per antenna is constrained. The instantaneous power per tone is unconstrained. This constraint is generally realizable.
- Per-tone and per-antenna constraint: The instantaneous power per tone and per antenna is constrained. This constraint is also realizable and the strictest of all.

These constraints are listed (approximately) from least to most restrictive, with the first three difficult to implement because they do not enforce a power constraint on a per-antenna basis. Nevertheless, all the constraints are analyzed separately to compare and contrast performance. Although the average performance of the beamformer decreases with more restrictive constraints, other performance metrics used to derive the beamformer performance can improve. In fact, we show that uplink beamforming with realizable power constraints in the presence of downlink CDD yields to analysis and performs surprisingly well. We now proceed with the system model.

II. The System Model

The basestation is equipped with N antennas and the terminal is equipped with M antennas. The uplink and downlink both use orthogonal frequency division multiplexing (OFDM) as the modulated signal. OFDM lends itself to simple receiver processing in the presence of the multipath delay spread typical of wireless communication channels [2]. Due to the nature of the OFDM signal, the transmitter and receiver use the fast Fourier transform (FFT) as the main processing unit for transmission and reception. Most receivers and transmitters do their processing in the frequency domain, and convert to a time-domain signal using the inverse FFT just before transmitting a signal, and convert to the frequency domain using the FFT just after receiving a signal. The number of frequency-domain subcarriers (sometimes also called "tones") in an OFDM symbol is denoted by F. This is generally the same as the FFT size.

The output of the FFT at the basestation receiver is then

$$\mathbf{y}(f) = \mathbf{H}(f)\mathbf{s}(f) + \mathbf{n}(f), \quad f = 0, 1, \cdots, F-1 \tag{1}$$

where $\mathbf{y}(f)$ is an $N \times 1$ complex vector representing the FFT output at the fth subcarrier, and

$$\mathbf{H}(f) = \begin{bmatrix}\mathbf{h}_1(f) & \mathbf{h}_2(f) & \cdots & \mathbf{h}_N(f)\end{bmatrix}^T \tag{2}$$

is the channel between the terminal and the basestation. The superscript T denotes transpose. We assume a flat channel: $\mathbf{H}(f) = \mathbf{H}$ comprises elements h_{nm}, $n = 1, 2, \cdots, N$, $m = 1, 2, \cdots, M$ that are independent identically distributed (i.i.d.) zero-mean complex Gaussian circularly symmetric with unit variance, denoted $\mathcal{CN}(0, 1)$. The vector $\mathbf{h}_n = [h_{nm}]_{m=1}^{M}$ is an $M \times 1$ vector, $\mathbf{n}(f)$ is an $N \times 1$ complex vector that denotes the contribution of thermal noise and interference. The signal $\mathbf{s}(f)$ is a $M \times 1$ vector transmitted by the terminal:

$$\mathbf{s}(f) = \mathbf{w}(f)b(f), \qquad f = 0, 1, \ldots, F-1 \tag{3}$$

where $\mathbf{w}(f)$ is a $M \times 1$ "beam" used to transmit the complex symbol $b(f)$. We assume that $E|b(f)|^2 = \frac{1}{F}$, and therefore $\sum_{f=0}^{F-1} E|b(f)|^2 = 1$. We defer the discussion of the power constraint on $\mathbf{w}(f)$ until the next section.

To form the beam, the terminal needs to use channel information derived from the preamble. The output of the FFT at the MS receiver for the preamble is

$$\tilde{\mathbf{y}}(f) = \tilde{\mathbf{h}}(f)x(f) + \tilde{\mathbf{n}}(f), \qquad f = 0, 1, \ldots, F-1 \tag{4}$$

where $\tilde{\mathbf{h}}(f) = \begin{bmatrix}\tilde{h}_1(f) & \cdots & \tilde{h}_M(f)\end{bmatrix}^T$ is a $M \times 1$ vector which denotes the downlink channel as seen by the terminal, $x(f)$ is the transmitted preamble (known to the terminal) and $\tilde{\mathbf{n}}(f)$ is the complex noise-plus-interference. The channel $\tilde{\mathbf{h}}(f)$ may be written as

$$\tilde{\mathbf{h}}(f) = \mathbf{H}^T\mathbf{d}(f) \tag{5}$$

where $\mathbf{H}$ is defined in (2) and $\mathbf{d}(f)$ is an $N \times 1$ complex vector that denotes the frequency-selective antenna virtualization technique used by the basestation during the preamble. Antenna virtualization is a method applied at the transmitter wherein the N transmitter antennas are virtualized to a single transmit antenna from the receiver perspective. One form of antenna virtualization is *cyclic delay diversity* (CDD) [9]. In CDD the signals at antennas $n = 2, \ldots, N$ are phase-shifted with respect to the first antenna as a function of the subcarrier index. One way of achieving CDD is setting

$$\begin{aligned}\mathbf{d}(f) &= \begin{bmatrix} d_1(f) & d_2(f) & \cdots & d_N(f) \end{bmatrix}^T \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{-j2\pi \frac{f}{F}\tau_2} & \ldots & e^{-j2\pi \frac{f}{F}\tau_N} \end{bmatrix}^T \end{aligned} \quad (6)$$

where $\tau_2, \ldots, \tau_N$, denote the integer cyclic time-delays applied to the basestation antennas (relative to the first antenna). The first antenna is assumed to have no phase shift. CDD has the effect of converting spatial-selectivity to frequency-selectivity. The terminal estimates the frequency-selective channel without needing to know $\mathbf{d}(f)$.

Nevertheless, instead of using CDD, the BS may transmit from a single antenna, say, the first, and then $\mathbf{d}(f)$ is given by

$$\mathbf{d}(f) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \quad (7)$$

In this case the channel as seen by the MS is simply $\tilde{\mathbf{h}}(f) = \mathbf{h}_1$.

In either case, $E\|\tilde{\mathbf{h}}(f)\|^2 = M,\ f = 0, \ldots, F-1$. Equation (5) implicitly assumes that the downlink and uplink channels are reciprocal. We ignore any calibration needed to achieve reciprocity. We also assume that the estimate of $\tilde{\mathbf{h}}(f)$ at the terminal is noise-free.

The performance metric we follow is the total received signal power at the basestation, defined as

$$P = \frac{1}{F} \sum_{f=0}^{F-1} \|\mathbf{H}\mathbf{w}(f)\|^2. \quad (8)$$

Note that P has been averaged over the statistics of the data symbol $b(f)$. This metric is a proxy for the channel capacity on a symbol

$$C = \sum_{f=0}^{F-1} \log_2 \left(1 + \frac{\rho}{F} \|\mathbf{H}\mathbf{w}(f)\|^2\right), \quad (9)$$

where $\frac{\rho}{F}$ is the average SNR per tone (subcarrier). For small ρ, $C \approx (\log_2 e) \frac{\rho}{F} \sum_{f=0}^{F-1} \|\mathbf{H}\mathbf{w}(f)\|^2$.

The power P in (8) is a random variable whose statistics depend on the channel, the beamforming algorithm, and the type of power constraints employed at the terminal. We are generally interested in outage events where the beamformer is characterized by its cumulative distribution function (cdf),

$$\eta = \Pr\{P < P_\eta\}, \qquad 0 \le \eta \le 1. \quad (10)$$

Generally, values of η of interest in a wireless network are between 0.01 and 0.1. The value P_η may then be considered as a power (or capacity) that we would like the channel to have with reliability $1 - \eta$. When comparing beamformers, the larger the P_η for a given η, the better the beamformer.

To compute (10) and analyze the behavior of P_η as a function of η, we derive the distribution of P wherever possible. But even having the distribution does not guarantee that we can compute P_η in closed form. We therefore also consider the following properties to aid our analysis.

Mean: The average value of P is defined as

$$\bar{P} = E[P] = \frac{1}{F} \sum_{f=0}^{F-1} E\|\mathbf{H}\mathbf{w}(f)\|^2. \quad (11)$$

The average power $\bar{P}$ is directly proportional to the signal-to-noise ratio (SNR) typically used to characterize the performance of a communication link between a transmitter and a receiver. For a given beamformer, the increase of $\bar{P}$ over the average received power of a single-antenna transmission technique is often called the *beamforming gain*.

Variance: The variance of P is denoted

$$\sigma_P^2 = E(P - \bar{P})^2. \quad (12)$$

The variance is a quantity that captures the amount of channel hardening due to the transmission method at the terminal [10].

Tail: The tail exponent of the probability density function (pdf) of P is defined as

$$\alpha = \lim_{P_\eta \to 0} \frac{\log \Pr(P < P_\eta)}{\log P_\eta}. \quad (13)$$

Equivalently, α is the slope of the curve of $\log \Pr(P < P_\eta)$ versus $\log P_\eta$ for small P_η. We abuse terminology and sometimes simply refer to α as the tail of P.

Characteristics of a "good" beamformer (that lead to a large P_η for a given η) include some combination of large $\bar{P}$, small σ_P^2, and large α.

III. Constrained Beamforming Solutions

The beams $\mathbf{w}(f)$ that we examine are all scaled versions of $\tilde{\mathbf{h}}^*(f)$, where the scaling is determined by the type of power constraint. The beams are particularly simple to implement because they are readily derived from $\tilde{\mathbf{h}}(f)$, the channel measured by the terminal. For each beam, an expression for P is derived. This expression is analyzed for its statistical distribution, and specifically its mean, variance and tail, wherever possible, and these are used to deduce properties of P_η.

A. Summary

The results for channel-unaware transmission techniques are summarized in Table I, and the results for beamforming in Table II. The following notations are used:

- $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix
- Superscript $*$ denotes the conjugate transpose
- χ^2_{2N} denotes a chi-squared distributed random variable with $2N$ degrees of freedom.
- $\gamma_N = \frac{\Gamma\left(N+\frac{1}{2}\right)}{\sqrt{N}\Gamma(N)} < 1$, where $\Gamma(\cdot)$ is the Gamma function [11].
- $\mathcal{R}_M$ denotes a sum of M i.i.d. Rayleigh random variables.
- $\mathbf{R}$ is an $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m = \sqrt{\frac{N}{F \sum_{n=1}^{N} |h_{nm}|^2}} \quad (14)$$

TABLE I

PROBABILITY DENSITY FUNCTION, MEAN, VARIANCE, AND TAIL EXPONENT OF THE RECEIVED POWER P FOR CHANNEL-UNAWARE TRANSMISSION

Transmission	pdf of P, $\bar{P}$, σ_P^2,α
1 TX	$\frac{1}{2}\chi^2_{2N}$ N N N
Uplink CDD	$\frac{1}{2M}\chi^2_{2NM}$ N $\frac{N}{M}$ NM

- $\mathbf{R}(f)$ is an $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m(f) = \frac{1}{|\tilde{h}_m(f)|} \tag{15}$$

We now discuss these tables in some detail, beginning with Table I. Most of the derivations are given only in outline or omitted altogether because of space constraints.

B. Single Transmit Antenna (Channel-unaware)

In this case the terminal transmits from a single antenna which we assume to be antenna 1. No channel knowledge is used to influence the transmission. The weight vector is therefore simply

$$\mathbf{w}(f) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T. \tag{16}$$

Thus, $\mathbf{H}\mathbf{w}(f) = \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N1} \end{bmatrix}^T$, therefore the total received power at the BS becomes

$$P = \frac{1}{F}\sum_{f=0}^{F-1}\sum_{n=1}^{N}|h_{n1}|^2 = \sum_{n=1}^{N}|h_{n1}|^2. \tag{17}$$

Since $h_{n1}, n = 1, 2, \cdots, N$ are i.i.d. $\mathcal{CN}(0, 1)$, equation (17) shows that P is a $\frac{1}{2}$-chi-square distributed random variable with $2N$ degrees of freedom, which we write as

$$P \sim \frac{1}{2}\chi^2_{2N} \tag{18}$$

The pdf of P is thus [12]

$$f_P(p) = \frac{1}{\Gamma(N)}p^{N-1}e^{-p}, \quad p \geq 0. \tag{19}$$

where $\Gamma(N)$ is the gamma function defined as

$$\Gamma(x) = \int_0^\infty dt\, t^{x-1}e^{-t}, \quad x > 0 \tag{20}$$

In our case, N is an integer and therefore $\Gamma(N) = (N-1)!$. The mean value of P is $\bar{P} = N$ and the variance of P is $\sigma_P^2 = N$. The tail exponent of P can be obtained from (19),

$$\begin{aligned} \Pr(P < P_\eta) &= \int_0^{P_\eta} dt \frac{1}{\Gamma(N)}t^{N-1}e^{-t} \\ &\approx \frac{1}{\Gamma(N+1)}P_\eta^N \end{aligned} \tag{21}$$

for small P_η, which leads to tail exponent

$$\alpha = N.$$

C. Uplink CDD (Channel-unaware)

In this case, the terminal uses CDD to transmit on the uplink, and the weight vector is

$$\mathbf{w}(f) = \frac{1}{\sqrt{M}}\begin{bmatrix} 1 & e^{-j2\pi\frac{f}{F}\tau_2} & \ldots & e^{-j2\pi\frac{f}{F}\tau_M} \end{bmatrix}^T$$

which has the same total power as in (16).

The total received power at the BS is

$$P = \text{tr}\left(\mathbf{H}\left[\frac{1}{F}\sum_{f=0}^{F-1}\mathbf{w}(f)\mathbf{w}^*(f)\right]\mathbf{H}^*\right) \tag{22}$$

It can be shown that $\frac{1}{F}\sum_{f=0}^{F-1}\mathbf{w}(f)\mathbf{w}^*(f) = \frac{1}{M}\mathbf{I}$, where $\mathbf{I}$ is the $M \times M$ identity matrix and thus P in equation (22) becomes

$$P = \frac{1}{M}\text{tr}\,(\mathbf{H}\mathbf{H}^*) = \frac{1}{M}\sum_{n=1}^{N}\|\mathbf{h}_n\|^2 \sim \frac{1}{2M}\chi^2_{2MN}. \tag{23}$$

Simple computations yield $\bar{P} = N$, $\sigma_P^2 = N/M$, and $\alpha = NM$. Even though $\bar{P}$ for uplink CDD is identical to the single antenna transmission, σ_P^2 for uplink CDD is reduced by a factor of M. The tail exponent of the pdf of P for uplink CDD is M times larger than the single antenna transmission.

As we see in Table I, although uplink CDD does not affect the mean of P, (23) has better variance and tail properties than (18).

Next, we define the beamformer that we subject to different types of power constraints.

D. Beamforming with Average Power Constraint

The terminal forms the beam with its M transmitter antennas using only its knowledge of the equivalent channel $\tilde{\mathbf{h}}(f)$ on the downlink. With only an average power constraint, it simply conjugates and normalizes the equivalent channel to obtain

$$\mathbf{w}^T(f) = \frac{1}{\sqrt{M}}\tilde{\mathbf{h}}^*(f). \tag{24}$$

This beam satisfies the average power constraint

$$E\|\mathbf{w}(f)\|^2 = 1. \tag{25}$$

The total received power at the BS is then given by

$$\begin{aligned} P &= \frac{1}{F}\sum_{f=0}^{F-1}\text{tr}\,(\mathbf{H}\mathbf{w}(f)\mathbf{w}^*(f)\mathbf{H}^*) \\ &= \frac{1}{FM}\sum_{f=0}^{F-1}\text{tr}\left(\mathbf{H}\tilde{\mathbf{h}}^{*T}(f)\tilde{\mathbf{h}}^T(f)\mathbf{H}^*\right) \end{aligned} \tag{26}$$

To continue the analysis, we now need to separate into the two cases, without and with downlink CDD.

TABLE II

PROBABILITY DENSITY FUNCTION, MEAN, VARIANCE, AND TAIL EXPONENT OF THE RECEIVED POWER P FOR UPLINK BEAMFORMING. THE TABLE IS ORDERED FROM LEAST TO MOST RESTRICTIVE TYPE OF POWER CONSTRAINT. THE LAST TWO ROWS (PER-ANTENNA CONSTRAINT) ARE CONSIDERED THE MOST EASILY REALIZABLE.

TX Constraint	Density of P, $\bar{P}$, σ_P^2,α without Downlink CDD	Density of P, $\bar{P}$,σ_P^2,α with Downlink CDD
Average	$\frac{1}{2M}\chi^2_{2M}(1)\left[\frac{1}{2}\chi^2_{2M}(1)+\frac{1}{2}\chi^2_{2(N-1)}\right]$ $N+M$ $\frac{N^2+4M^2+5NM+3N+5M+2}{M}$ $\min\left(\frac{N+M-1}{2}, M\right)$	$\frac{1}{NM}\mathrm{tr}\left(\mathbf{HH}^*\mathbf{HH}^*\right)$ $N+M$ $2\left(\frac{2N^2+2M^2+5NM+1}{NM}\right)$ $\frac{NM}{2}$
Per-Symbol	$\frac{1}{2}\chi^2_{2(N+M-1)}$ $N+M-1$ $N+M-1$ $N+M-1$	$\frac{\mathrm{tr}\left(\mathbf{HH}^*\mathbf{HH}^*\right)}{\mathrm{tr}\left(\mathbf{HH}^*\right)}$ $\frac{NM(N+M)}{NM+1}$ $\frac{NM\left[(N+M)^2(N^2M^2+3MN+4)+2(NM+1)^3\right]}{(NM+1)^2(NM+2)(NM+3)}$ NM
Per-Tone	$\frac{1}{2}\chi^2_{2(N+M-1)}$ $N+M-1$ $N+M-1$ $N+M-1$	$\frac{1}{F}\mathrm{tr}\left(\sum_{f=0}^{F-1}\frac{\mathbf{d}^{*T}(f)\mathbf{d}^T(f)}{\mathrm{tr}\left(\mathbf{H}^T\mathbf{d}(f)\mathbf{d}^*(f)\mathbf{H}^{*T}\right)}\mathbf{HH}^*\mathbf{HH}^*\right)$ $N+M-1$ $< N+M-1$ NM
Per-Symbol Per-Ant	$\frac{1}{M}\mathcal{R}^2_M+\frac{1}{2}\chi^2_{2(N-1)}$ $N+\frac{\pi}{4}(M-1)$ $N+\frac{M-1}{8M}\left[\pi(8M-12)+\pi^2(3-2M)+8\right]$ $N+M-1$	$\frac{F}{NM}\mathrm{tr}\left(\mathbf{HRH}^*\mathbf{HRH}^*\right)$ $N+\gamma_N^2(M-1)$ $2\gamma_N^2(1-\gamma_N^2)(2M-3)\left(1-\frac{1}{M}\right)+\frac{N(4M+N-3)}{M(N+1)}$ NM
Per-Tone Per-Ant	$\frac{1}{M}\mathcal{R}^2_M+\frac{1}{2}\chi^2_{2(N-1)}$ $N+\frac{\pi}{4}(M-1)$ $N+\frac{M-1}{8M}\left[\pi(8M-12)+\pi^2(3-2M)+8\right]$ $N+M-1$	$\frac{1}{FM}\sum_{f=0}^{F-1}\mathrm{tr}\left(\mathbf{HR}(f)\mathbf{H}^*\mathbf{d}^{*T}(f)\mathbf{d}^T(f)\mathbf{HR}(f)\mathbf{H}^*\right)$ $N+\frac{\pi}{4}(M-1)$ $< N+\frac{M-1}{8M}\left[\pi(8M-12)+\pi^2(3-2M)+8\right]$ NM

1) Without Downlink CDD: In this case $\tilde{\mathbf{h}}(f)=\mathbf{h}_1$, and thus $\mathbf{w}^T(f)=\frac{1}{\sqrt{M}}\mathbf{h}_1^*$ which is independent of frequency index f, and P becomes

$$\begin{aligned} P &= \frac{1}{M}\left\|\mathbf{H}\mathbf{h}_1^{*T}\right\|^2=\frac{1}{M}\sum_{n=1}^{N}\left|\mathbf{h}_n^T\mathbf{h}_1^{*T}\right|^2 \\ &= \frac{1}{M}\left(\|\mathbf{h}_1\|^4+\sum_{n=2}^{N}|\mathbf{h}_n^T\mathbf{h}_1^{*T}|^2\right). \end{aligned} \tag{27}$$

It can be shown that P is distributed as

$$P\sim\frac{1}{2M}\chi^2_{2M}(1)\left(\frac{1}{2}\chi^2_{2M}(1)+\frac{1}{2}\chi^2_{2(N-1)}\right). \tag{28}$$

Observe that the term $\frac{1}{2}\chi^2_{2M}(1)$ appears twice in (28) which denotes the *same* random variable, however the $\frac{1}{2}\chi^2_{2(N-1)}$ denotes an independent random variable.

2) With Downlink CDD: Substituting for $\tilde{\mathbf{h}}(f)$ from (5) results in

$$P=\frac{1}{FM}\mathrm{tr}\left(\mathbf{HH}^*\left[\sum_{f=0}^{F-1}\mathbf{d}^{*T}(f)\mathbf{d}^T(f)\right]\mathbf{HH}^*\right) \tag{29}$$

It can be shown that

$$\sum_{f=0}^{F-1}\mathbf{d}(f)\mathbf{d}^*(f)=\frac{F}{N}\mathbf{I}, \tag{30}$$

where $\mathbf{I}$ is the $N\times N$ identity matrix. Thus, P becomes

$$P=\frac{1}{NM}\mathrm{tr}\left(\mathbf{HH}^*\mathbf{HH}^*\right). \tag{31}$$

This is the trace of the square of a Wishart matrix [12].

The mean, variance and the tail of P (with and without CDD) are provided in Table II (see row marked "Average"). We omit the detailed derivations.

E. Per-Symbol Constraint

In this case the beam is

$$\mathbf{w}^T(f)=\sqrt{F}\frac{\tilde{\mathbf{h}}^*(f)}{\sqrt{\sum_{f=0}^{F-1}\left\|\tilde{\mathbf{h}}(f)\right\|^2}} \tag{32}$$

which satisfies the constraint

$$\sum_{f=0}^{F-1}\|\mathbf{w}(f)\|^2=F. \tag{33}$$

The total received power at the BS is then

$$P=\frac{1}{\sum_{f=0}^{F-1}\left\|\tilde{\mathbf{h}}(f)\right\|^2}\sum_{f=0}^{F-1}\mathrm{tr}\left(\mathbf{H}\tilde{\mathbf{h}}^{*T}(f)\tilde{\mathbf{h}}^T(f)\mathbf{H}^*\right). \tag{34}$$

1) Without Downlink CDD: Since $\tilde{\mathbf{h}}(f)=\mathbf{h}_1$, the weight vector becomes $\mathbf{w}^T(f)=\frac{\mathbf{h}_1^*}{\|\mathbf{h}_1\|}$ and thus P becomes

$$\begin{aligned} P &= \left\|\mathbf{H}\frac{\mathbf{h}_1^{*T}}{\|\mathbf{h}_1\|}\right\|^2=\sum_{n=1}^{N}\left|\mathbf{h}_n^T\frac{\mathbf{h}_1^{*T}}{\|\mathbf{h}_1\|}\right|^2 \\ &= \|\mathbf{h}_1\|^2+\sum_{n=2}^{N}\left|\mathbf{h}_n^T\frac{\mathbf{h}_1^{*T}}{\|\mathbf{h}_1\|}\right|^2, \end{aligned} \tag{35}$$

Analysis shows that P is distributed as

$$P \sim \frac{1}{2}\chi^2_{2(N+M-1)}. \tag{36}$$

2) With Downlink CDD: Substituting for $\tilde{\mathbf{h}}(f)$ from (5) and using (30) yields

$$P = \frac{\text{tr}\left(\mathbf{HH}^*\mathbf{HH}^*\right)}{\text{tr}\left(\mathbf{HH}^*\right)}. \tag{37}$$

The mean, variance and tail exponent of P are shown in Table II. The derivation is omitted.

F. Per-Tone Constraint

In this case the beam is

$$\mathbf{w}(f)^T = \frac{\tilde{\mathbf{h}}^*(f)}{\|\tilde{\mathbf{h}}(f)\|} \tag{38}$$

which satisfies

$$\|\mathbf{w}(f)\|^2 = 1, \quad f = 0, 1, \cdots, F-1. \tag{39}$$

1) Without Downlink CDD: The weight vector is $\mathbf{w}^T(f) = \frac{\mathbf{h}_1^*}{\|\mathbf{h}_1\|}$, which is the same weight vector as the per-symbol power constraint case and thus P is given in (35).

2) With Downlink CDD: Substituting for $\tilde{\mathbf{h}}(f)$ from (5), yields

$$P = \frac{1}{F}\text{tr}\left(\sum_{f=0}^{F-1}\frac{\mathbf{d}^{*T}(f)\mathbf{d}^T(f)}{\text{tr}\left(\mathbf{H}^T\mathbf{d}(f)\mathbf{d}^*(f)\mathbf{H}^{*T}\right)}\mathbf{HH}^*\mathbf{HH}^*\right) \tag{40}$$

The mean and tail exponent of P are provided in Table II. Only an upper bound on the variance of P is provided.

G. Per-Symbol, Per-Antenna Constraint

In this case the beam is

$$\mathbf{w}^T(f) = \sqrt{\frac{F}{M}}\tilde{\mathbf{h}}^*(f)\mathbf{R} \tag{41}$$

where $\mathbf{R}$ is a $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m = \frac{1}{\sqrt{\sum_{f=0}^{F-1}\left|\tilde{h}_m(f)\right|^2}}. \tag{42}$$

The weight vector in this case satisfies the constraint

$$\sum_{f=0}^{F-1}|w_m(f)|^2 = \frac{F}{M}, \quad m = 1, 2, \cdots, M. \tag{43}$$

The total received power at the BS is

$$P = \frac{1}{M}\sum_{f=0}^{F-1}\text{tr}\left(\mathbf{HR}\tilde{\mathbf{h}}^{*T}(f)\tilde{\mathbf{h}}^T(f)\mathbf{RH}^*\right) \tag{44}$$

1) Without Downlink CDD: In this case, $\tilde{\mathbf{h}}(f) = \mathbf{h}_1$ and $\mathbf{R}$ becomes a $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m = \frac{1}{\sqrt{F}\,|h_{1m}|} \tag{45}$$

Thus, P is given by

$$P = \frac{1}{M}\left\|\begin{matrix}\sum_{m=1}^M |h_{1m}| \\ \sum_{m=1}^M h_{2m}e^{-j\phi_m} \\ \vdots \\ \sum_{m=1}^M h_{Nm}e^{-j\phi_m}\end{matrix}\right\|^2 \tag{46}$$

where $h_{1m} = |h_{1m}|\,e^{j\phi_m}$. We conclude that P is distributed as

$$P \sim \frac{1}{M}\left(\left[\sum_{m=1}^{M}\sqrt{\frac{1}{2}\chi_2^2(m)}\right]^2 + \sum_{n=1}^{N-1}\frac{M}{2}\chi_2^2(n)\right) \tag{47}$$

The term $\sqrt{\frac{1}{2}\chi_2^2(m)}$ is Rayleigh distributed. Let $\mathcal{R}_M$ denote the sum of M i.i.d. Rayleigh random variables. Then P is distributed as

$$P \sim \frac{1}{M}\mathcal{R}_M^2 + \frac{1}{2}\chi^2_{2(N-1)} \tag{48}$$

2) With Downlink CDD: Using (30), $\mathbf{R}$ becomes an $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m = \sqrt{\frac{N}{F\sum_{n=1}^N |h_{nm}|^2}}. \tag{49}$$

Thus, P becomes

$$P = \frac{1}{M}\text{tr}\left(\mathbf{HRH}^*\left[\sum_{f=0}^{F-1}\mathbf{d}^{*T}(f)\mathbf{d}^T(f)\right]\mathbf{HRH}^*\right). \tag{50}$$

Using the result from (30), P can be rewritten as

$$P = \frac{F}{NM}\text{tr}\left(\mathbf{HRH}^*\mathbf{HRH}^*\right) \tag{51}$$

The mean, variance and the tail exponent of P are provided in Table II.

H. Per-Tone, Per-Antenna Constraint

In this case the beam is

$$\mathbf{w}^T(f) = \sqrt{\frac{1}{M}}\tilde{\mathbf{h}}^*(f)\mathbf{R}(f) \tag{52}$$

where $\mathbf{R}(f)$ is a $M \times M$ diagonal matrix whose mth diagonal entry is

$$r_m(f) = \frac{1}{|\tilde{h}_m(f)|} = \frac{1}{|\sum_{n=1}^N h_{nm}d_n(f)|} \tag{53}$$

where $d_n(f) = \frac{1}{\sqrt{N}}e^{-j2\pi\frac{f}{F}\tau_n}$ as defined in (6). The weight vector in this case satisfies the following constraint

$$|w_m(f)|^2 = \frac{1}{M}, m = 1, 2, \cdots, M,\ f = 0, 1, \cdots, F-1 \tag{54}$$

and

$$P = \frac{1}{FM}\sum_{f=0}^{F-1}\text{tr}\left(\mathbf{HR}(f)\tilde{\mathbf{h}}^{*T}(f)\tilde{\mathbf{h}}^T(f)\mathbf{R}(f)\mathbf{H}^*\right).$$

1) Without Downlink CDD: This is identical to the per-symbol per antenna constraint and thus P is the same as in (46).

2) With Downlink CDD: The total received power becomes

$$= \frac{1}{FM} \sum_{f=0}^{F-1} \mathrm{tr}\left(\mathbf{H}\mathbf{R}(f)\mathbf{H}^* \mathbf{d}^{*T}(f) \mathbf{d}^T(f) \mathbf{H}\mathbf{R}(f)\mathbf{H}^* \right).$$

The mean and the tail exponent of the pdf of P are provided in Table II. Only an upper bound on the variance of P is provided.

IV. Discussion

Table II provides a way to compare the effects of the various constraints on beamformer performance. The last two rows constrain the transmitter power on a per-antenna basis. We believe these constraints are the most practical.

The formulas for variance in Table II are not necessarily easy to digest. Figure 1 shows the mean plus/minus one standard deviation of P for $N = 4$ as a function of M. The standard deviation σ_P is shown as the bar height around $\bar{P}$. For a single transmit antenna the mean and variance of P are independent of M. For uplink CDD, the mean is the same as the single transmit antenna, however as M increases the variance of P decreases. However, all beamformers show an increase in the mean value of P as M increases. Also shown in the figure is that downlink CDD reduces the variance of P for all uplink beamformers.

Figure 2 shows the cumulative distribution function of P for $N = 4$ and $M = 2$; equivalently η as a function of P_η. Included also is the cumulative distribution obtained if the terminal knows $\mathbf{H}$ perfectly and can beamform using the sum-power constraint. We view this scenario as not realizable but useful as a reference case. Figure 3 shows the same plots but for $M = N = 4$. We see that the beamforming results all move significantly to the right (improving their performance) with the larger M.

To draw some conclusions, we examine $P_{\eta=0.01}$ in Figure 3 for $N = M = 4$ for the various uplink transmission schemes. It is perhaps counterintuitive that the worst performer of all is the beamformer obtained with an average (least restrictive type of) power constraint. In particular, the beamformer without downlink CDD is worse than even the simple single-antenna transmission. Its performance is poor at $\eta = 0.01$ even though its average $N + M = 8$ is the best according to Table II, and it only starts to perform well if we look at $\eta \approx 0.3$, which corresponds to 30% outage; not a practical operating point. Its poor performance can be attributed to its large variance and its small tail exponent given in the table as $\alpha = 3.5$. Many of the competing schemes have a better variance and tail exponent, even if they have a worse average. For example, uplink CDD (channel-unaware) is at least 5 dB better, even though its average received power is only $N = 4$, because its tail is $\alpha = 16$. We are led to wonder if we can improve the tail of the beamformer without compromising its average too much.

As the trends in Table II show, a more restrictive type of power constraint accomplishes this by reducing the variance and increasing the tail exponent, at a small expense to $\bar{P}$. In fact, enforcing the per-symbol per-antenna power constraint is very beneficial to performance. Figure 3 denotes the curve as "Per Sym Per Ant, no CDD", and it is described in Section III-G. Its performance is uniformly better than uplink CDD throughout the range of η. The average $\bar{P}$ has been reduced from $N + M = 8$ to $N + (\pi/4)(M - 1) = 6.4$, but the tail has been improved from $\alpha = 3.5$ to $\alpha = 7$. While this is commendable, the performance advantage over uplink CDD is small, especially at $\eta = 0.01$, not readily justifying its use.

However, another 1-2 dB in performance is obtained when the downlink employs CDD since $\bar{P}$ increases slightly but, more importantly, the tail improves to $\alpha = MN = 16$. We have therefore seen that a careful analysis of beamforming is needed to justify its use on the uplink, but downlink CDD can help ensure that uplink beamforming is advantageous over simpler channel-unaware methods.

References

[1] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Jour. Select. Areas in Commun.*, vol. 16, pp. 1451–1458, October 1998.

[2] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. New York: Cambridge University Press, 2003.

[3] L. Jalloul and S. P. Alex, "Coverage Analysis for IEEE 802.16e/WiMAX Systems," *IEEE Trans. Wireless Comm.*, vol. 7, pp. 4627–4634, November 2008.

[4] IEEE 802.16e Technical Working Group, *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*. IEEE, 2006.

[5] WiMAX Forum, *Mobile System Profile*. Release 1.0 Approved Specification, 2008.

[6] 3GPP, *TS 36.211 Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation*. Third Generation Partnership Project, May 2009.

[7] L. Jalloul, N. Czink, B. Hochwald, and A. Paulraj, "Why downlink cyclic delay diversity helops uplink mimo transmission schemes," in *IEEE Vehicular Technology Conference*, (Barcelona, Spain), April 2009.

[8] A. Edelman, *Eigenvalues and condition numbers of random matrices*. PhD Thesis, MIT, Dept. of Mathematics, 1989.

[9] J. C. Y. Li and N. Sollenberg, "Transmitter diversity for ofdm systems and its impacts on high-rate data wireless networks," *IEEE Jour. Select. Areas in Commun.*, vol. 17, pp. 1233–1243, July 1999.

[10] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Info. Th.*, vol. 50, pp. 1893–1909, September 2004.

[11] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products, 5th ed.* San Diego: Academic Press, 1994.

[12] A. Gupta and D. Nagar, *Matrix Variate Distributions*. Boca Raton: Chapman & Hall/CRC, 2000.

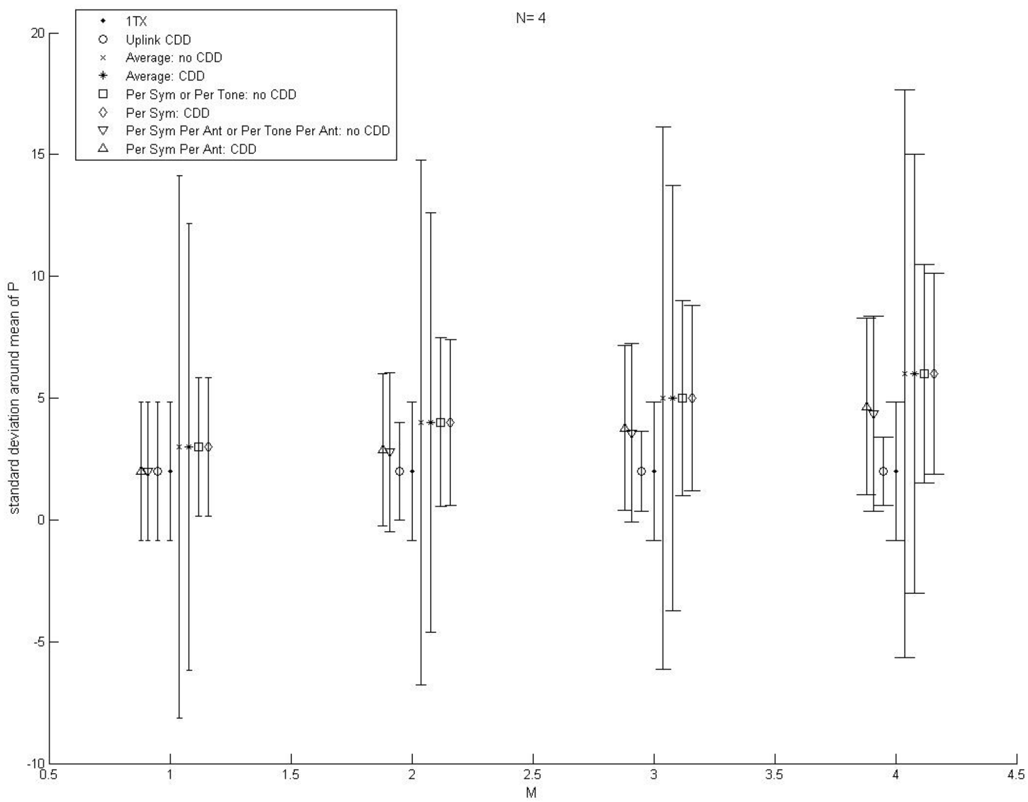

Fig. 1. Mean of P for $N = 4$ as a function of M, shown with "confidence intervals" that extend plus/minus one standard deviation. The means of the channel-unaware schemes stay constant, while all the beamformers improve with M. A combination of large mean and small variance generally makes for best performance. Any portions of the intervals that extend below $P = 0$ are not physically meaningful.

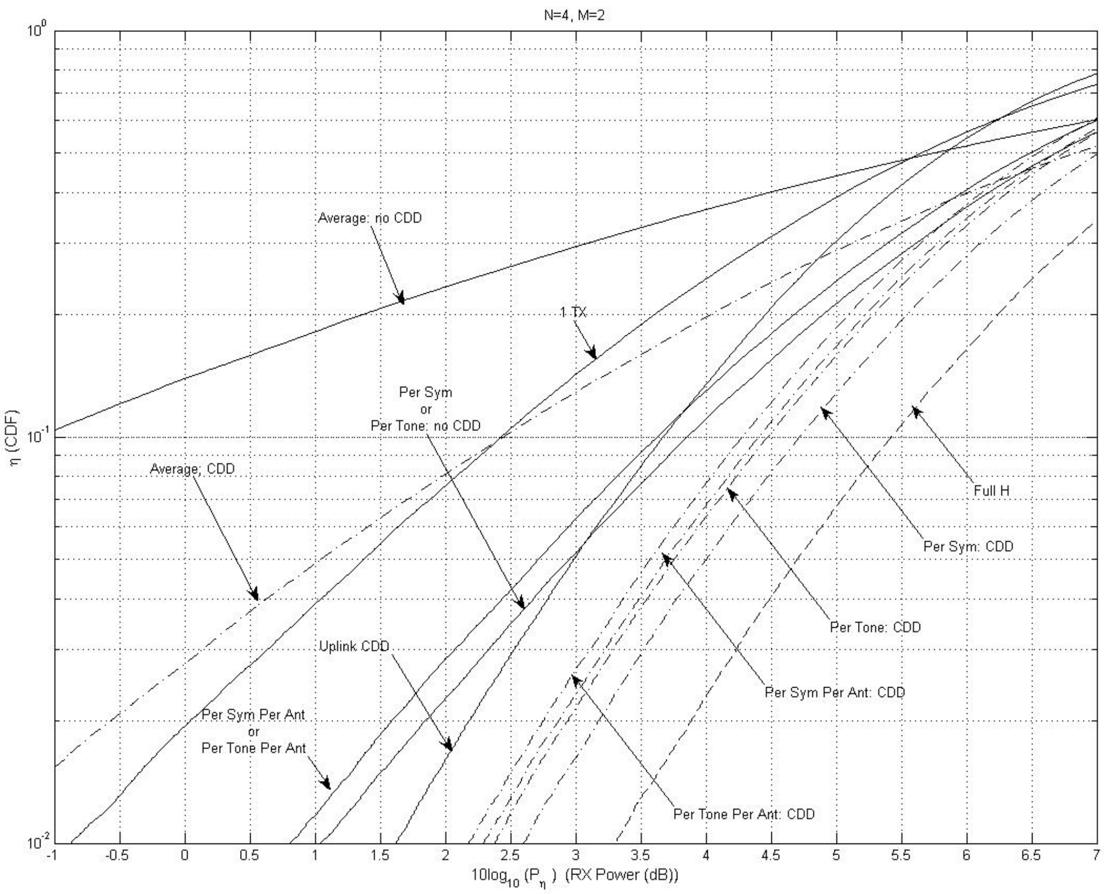

Fig. 2. Cumulative distribution function of P (η as a function of P_η) for $N = 4$ and $M = 2$ for both channel-unaware and beamforming schemes with the various types of power constraints. The solid lines are for no downlink CDD and the dashed lines are with downlink CDD. The right-most dashed curve is for full knowledge of $\mathbf{H}$. The means, variances, and tails of the distributions can all be found in Tables I and II.

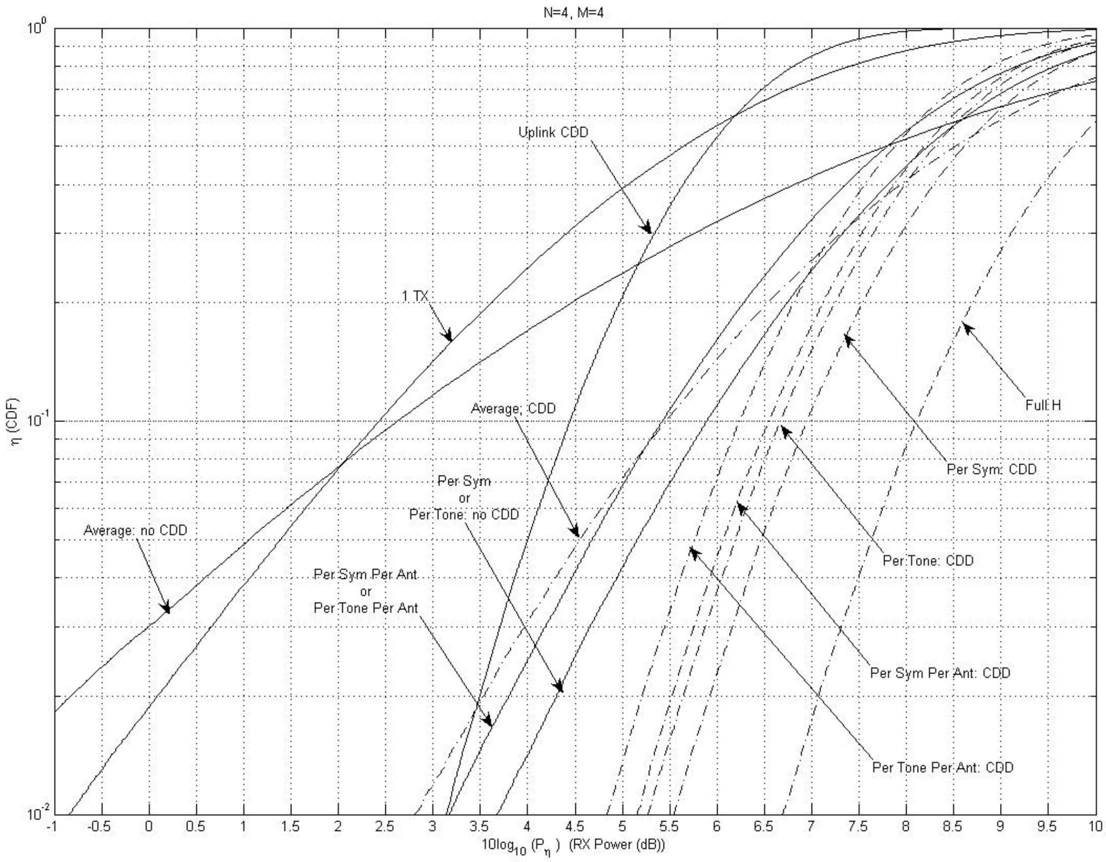

Fig. 3. Similar to Figure 2, except for $N = M = 4$. Observe that the beamformers have moved significantly to the right (improved) compared with Figure 2, and the gap with the channel-unaware techniques has increased.