Progress Report

Tianpei Chen
Department of Electrical and Computer Engineering
McGill University
Montreal, Quebec, Canada

Abstract

The abstract goes here.

Index Terms

IEEEtran, journal, LaTeX, paper, template.

I. System Model

Consider an uncoded complex Large-Scale MIMO (LS-MIMO) uplink spatial multiplexing (SM) system with N_t users, each is equipped with one transmit antenna. The number of receive antennas at the Base Station (BS) is N_r. Typically LS-MIMO systems have hundreds of antennas at the BS,

Bit sequences, which are modulated to complex symbols, are transmitted by the users over a flat fading channel. The discrete time model of the system is given by:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ is the received symbol vector, $\mathbf{s} \in \mathbb{C}^{N_t \times 1}$ is the transmitted symbol vector, with components that are mutually independent and taken from a finite signal constellation alphabet $\mathbb{O}$ (e.g., BPSK, 4-QAM, 16-QAM, 64-QAM), $|\mathbb{O}| = M$. The transmitted symbol vectors $\mathbf{s} \in \mathbb{O}^{N_t}$, satisfy $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_t} E_s$, where E_s denotes the symbol average energy, $\mathbb{E}[\cdot]$ denotes the expectation operation, $\mathbf{I}_{N_t}$ denotes identity matrix of size $N_t \times N_t$. Furthermore $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$

denotes the Rayleigh fading channel propagation matrix. $\mathbf{H}_{ij}$ denotes the component of $\mathbf{H}$ at ith row and jth column, representing the channel response from ith receive antenna to the jth transmit antenna. Each component is independent identically distributed (i.i.d) circularly symmetric complex Gaussian (CSCG) random variable with unit variance, denoted by $\mathbf{H}_{ij} \sim \mathbb{C}N(0,1)$. Finally, $\mathbf{n} \in \mathbb{C}^{N_r}$ is the additive white Gaussian noise (AWGN) vector with zero mean components and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_{N_r}N_0$, where N_0 denotes the noise power spectrum density, and hence $\frac{E_s}{N_0}$ is the symbol signal to noise ratio (SNR).

The task of a MIMO detector is to estimate the transmit symbol vector $\mathbf{s}$, based on the knowledge of receive symbol vector $\mathbf{y}$ and channel state information (CSI) $\mathbf{H}$. Then a constellation demapper maps each component of $\hat{\mathbf{s}}$ to the corresponding bit sequences. The resulting N_t substreams are finally multiplexed to obtain the reconstructed information sequence.

The optimal (in a sense of lowest average frame error probability) Maximum Likelihood Detector (MLD) for MIMO system is given by

$$\hat{\mathbf{s}}^{ML} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{O}^{N_t}} ||\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}||^2, \tag{2}$$

where $||\cdot||$ denotes the 2-norm operation. From (2), the solution of MLD is the $\hat{\mathbf{s}}$ that can generate the minimum Euclidean distance between vector $\mathbf{y}$ and $\mathbf{H}\hat{\mathbf{s}}$.

Consider an alternative representation of MLD principle, let $\mathbf{H}_1 \in \mathbb{C}^{N_r \times N}$ denotes a sub matrix composed of N columns from $\mathbf{H}$, where $1 \le N \le N_t$, let $\mathbf{s}_1 \in \mathbb{C}^N$ denote the sub symbol vector whose components are transmitted from the users corresponding to $\mathbf{H}_1$. Similarly, let $\mathbf{H}_2 \in \mathbb{C}^{N_r \times (N_t - N)}$ denotes the sub matrix composed of the remaining columns from $\mathbf{H}$ and $\mathbf{s}_2 \in \mathbb{C}^{(N_t - N)}$ is the sub symbol vector whose components are transmitted by the users corresponding to $\mathbf{H}_2$. Thus (1) can be rewritten as

$$\mathbf{y} = \mathbf{H}_1\mathbf{s}_1 + \mathbf{H}_2\mathbf{s}_2 + \mathbf{n}. \tag{3}$$

Let $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$ denote the estimations of $\mathbf{s}_1, \mathbf{s}_2$, we have $\hat{\mathbf{s}}_1 \in [\hat{\mathbf{s}}_1^1, \hat{\mathbf{s}}_1^2, \cdots \hat{\mathbf{s}}_1^K], , K = M^N$ and $\hat{\mathbf{s}}_2 \in [\hat{\mathbf{s}}_2^1, \hat{\mathbf{s}}_2^2, \cdots \hat{\mathbf{s}}_2^Q], Q = M^{N_t - N}$, MLD in (2) can be rewritten as

$$[\tilde{k}, \tilde{q}] = \arg \min_{k \in [1,2\ldots K]} \min_{q \in [1,2,\ldots,Q]} ||\mathbf{y} - \mathbf{H}_1\hat{\mathbf{s}}_1^k - \mathbf{H}_2\hat{\mathbf{s}}_2^q||^2, \mathbf{s}_1^{ML} = \hat{\mathbf{s}}_1^{\tilde{k}}, \mathbf{s}_2^{ML} = \hat{\mathbf{s}}_2^{\tilde{q}}, \tag{4}$$

(4) can be divided into three steps, first define

$$\mathbf{y}^k = \mathbf{y} - \mathbf{H}_1\hat{\mathbf{s}}_1^k \quad k \in [1, 2, \ldots, K], \tag{5}$$

Then solve

$$\hat{\mathbf{x}}_2^k = \arg \min_{\hat{\mathbf{s}}_2^q \in [\hat{\mathbf{s}}_2^1, \hat{\mathbf{s}}_2^3 \ldots, \hat{\mathbf{s}}_2^Q]} ||\mathbf{y}^k - \mathbf{H}_2\hat{\mathbf{s}}_2^q||^2, \tag{6}$$

$$\tilde{k} = \arg \min_{k \in [1,2,\ldots,K]} ||\mathbf{y}^k - \mathbf{H}_2\hat{\mathbf{x}}_2^k||^2, \tag{7}$$

Finally we have $\mathbf{s}_1^{ML} = \hat{\mathbf{s}}_1^{\tilde{k}}, \mathbf{s}_2^{ML} = \hat{\mathbf{x}}_2^{\tilde{k}} = \hat{\mathbf{s}}_2^{\tilde{q}}$.

II. Minimum Achievable Diversity based Channel Partition

A. Diversity Maximization Selection

Based on the alternative form of MLD in (4)-(7), General Parallel Interference Cancellation (GPIC) algorithm first generate a list of symbol vector candidates and then choose the candidate with the minimum Euclidean distance as the solution. To elaborate further, GPIC detectors first generate a list of symbol vector candidate by exhaustive search for all the possible $\hat{\mathbf{s}}_1$ as shown in (5), then rather than performing exhaustive search in (6), GPIC detectors exploit the low complexity sub optimal linear detectors (i.e., zeros forcing (ZF) and minimum mean square error (MMSE)) and their successive interference cancellation (SIC) counterparts to get the estimation of $\mathbf{s}_2^k$, $k = [1, 2, \ldots, K]$. For sake of simplicity, hereinafter we use LD/SIC. Then the best candidate in the list with the minimum Euclidean distance is chosen as the solution.

Performance analysis of GPIC algorithm is provided in [1], here we briefly conclude the main results in [1]. At the receiving side, the frame error (detection error) is defined as there is at least one erroneous symbol in the estimation of transmitted symbol vector. Let P_e denotes the average frame error probability. The diversity order, denoted by d is defined as the slope of the P_e in log-scale in high SNR region, which is given by [2]

$$d = -\lim_{SNR \to \infty} \frac{\log P_e}{\log(SNR)}, \tag{8}$$

Let P_e^{ML} denotes the average frame error probability of MLD, The average frame error probability of GPIC algorithm $P_{et} = \mathbb{E}_{\mathbf{H},\mathbf{s}}[Pr(\hat{\mathbf{s}} \neq \mathbf{s}|\mathbf{H}, \mathbf{s})]$ and $P_{e2} = \mathbb{E}_{\mathbf{H}_2,\mathbf{s}_2}(Pr(\hat{\mathbf{s}}_2^{k_1} \neq \mathbf{s}_2|\mathbf{H}_2, \mathbf{s}_2))$, where $P_r(A)$ denotes the probability that a event A occurs, $\hat{\mathbf{s}}_1^{k_1} = \mathbf{s}_1$, therefore P_{e2} is the average frame error probability of the LD/SICs for the sub system in which the interference from the users that transmitted $\mathbf{s}_1$ are perfectly cancelled from the observation $\mathbf{y}$. given by

$$\mathbf{y}^{k_1} = \mathbf{y} - \mathbf{H}_1 \hat{\mathbf{s}}_1^{k_1}, \tag{9}$$

because $\hat{\mathbf{s}}_1^{k1} = \mathbf{s}_1$, we have

$$\mathbf{y}^{k1} = \mathbf{H}_2\mathbf{s}_2 + \mathbf{H}_1(\mathbf{s}_1 - \hat{\mathbf{s}}_1^{k1}) + \mathbf{n} = \mathbf{H}_2\mathbf{s}_2 + \mathbf{n}, \tag{10}$$

Based on [1], P_{et} is bounded by

$$\max(P_e^{ML}, P_{e2}) \leq P_{et} \leq P_e^{ML} + P_{e2}, \tag{11}$$

in (11), the diversity order of MLD is N_r. Let d_t denote the overall diversity order of GPIC detectors and d_2 denote the diversity order of LD/SICs with the interferences from $\mathbf{s}_1$ perfectly cancelled, based on (8), we have

$$\lim_{SNR\to\infty} P_e^{ML} \propto SNR^{-N_r}, \tag{12}$$

$$\lim_{SNR\to\infty} P_{et} \propto SNR^{-d_t}, \tag{13}$$

$$\lim_{SNR\to\infty} P_{e2} \propto SNR^{-d_2}, \tag{14}$$

now consider the following conditions

- $d_2 > N_r$,

  $$\lim_{SNR\to\infty} \frac{P_{e2}}{P_e^{ML}} = SNR^{N_r - d_2} = 0, \tag{15}$$

  based on (11), we have

  $$\lim_{SNR\to\infty} \frac{P_{et}}{P_e^{ML}} = 1, \tag{16}$$

  thus $d_t = N_r$.
- $d_2 = N_r$, we have

  $$\lim_{SNR\to\infty} P_{et} \propto SNR^{-N_r}, \tag{17}$$

thus $d_t = N_r$.

- $d_2 < N_r$,

$$\lim_{SNR\to\infty} \frac{P_e^{ML}}{P_{e2}} = SNR^{d_2 - N_r} = 0, \tag{18}$$

based on (11), we have

$$\lim_{SNR\to\infty} P_{et} = P_{e2} \propto SNR^{-d_2}, \tag{19}$$

thus $d_t = d_2$.

Therefore when $d_2 \geq N_r$, GPIC algorithm can achieve ML performance asymptotically, when $d_2 < N_r$, the diversity order that GPIC can achieve at high SNR region is equal to that of LD/SICs. In [1], the authors employ diversity maximization selection (DMS) scheme for channel partition, which is originally proposed in [3]. The optimal diversity order can be guaranteed with low complexity for conventional small MIMO systems.

To elaborate further, for a given number of antennas chosen at the channel partition stage N, there are $N_u = \binom{N_t}{N}$ possible combinations of $[\mathbf{H}_1, \mathbf{H}_2]$. According to DMS principle, the maximum diversity order of LD/SIC is achieved by choosing the subset $\mathbf{H}_2$ that has the strongest weakest substream, in a sense of post processing SNR. Therefore, when using LD/SIC detectors as the sub optimal detector, the subset $\mathbf{H}_2^{opt}$ selected based on DMS principle should be [3]

$$\mathbf{H}_2^{opt} = \mathbf{H}_2^p, \tag{20}$$

$$p = \arg \min_{j=1,2,\ldots,N_u} \theta_j, \tag{21}$$

$$\theta_j = \max_{k=1,2,\ldots,N_t-N} ((\mathbf{H}_2^j)^H \mathbf{H}_2^j + SNR^{-1}\mathbf{I})^{-1}_{kk}, \quad \mathbf{H}_2^j \in [\mathbf{H}_2^1, \mathbf{H}_2^2, \ldots, \mathbf{H}_2^{N_u}]; \tag{22}$$

Where $\mathbf{A}_{kk}$ denotes the kth diagonal component of matrix $\mathbf{A}$. The diversity order of LD/SIC detector after DMS channel partition process is given by $d_{2DMS} = (N+1)(N_r - N_t + N + 1)$. In [1], the authors derive the minimum number of antennas N_{min} at channel partition stage that can guarantee $d_{2DMS} \geq N_r$, which is given by

$$N_{min} = \lceil \sqrt{\frac{(N_r - N_t)^2}{4} + N_r} - \frac{N_r - N_t}{2} - 1 \rceil, \tag{23}$$

where $\lceil \alpha \rceil$ denotes the minimum integer no less than α.

B. Minimum Achievable Diversity: How many transmit antennas do we need to chose in LS-MIMO?

In LS-MIMO V-BLAS systems, where the substreams are transmitted independently, the maximum diversity order is N_r, which is extremely large under the condition that there are tens to hundreds of receive antennas. On the one hand, diversity order is achieved at high SNR region, however, in practical, such a high diversity order is not necessary, because the SNR region of interest is the range in which the bit error rate (BER) or symbol error rate (SER) can be 10^{-5} to 10^{-7}. On the other hand, the computational complexity to guarantee ML performance by DMS channel partition is excessive in LS-MIMO, for example, when $N_r = N_t$, based on (23), the $N_{min} = \lceil \sqrt{N_r} - 1 \rceil$, the number of the symbol vector candidates in the list is $M^{N_{min}}$.

Therefore, here we consider minimum achievable diversity (MAD) principle, which can reduce the detection complexity by sacrificing redundant asymptotic diversity gain in LS-MIMO. Let d_{2MAD} denote the diversity order achieved by the suboptimal detectors after the MAD principle based channel partition, $\tilde{N}_{min}$ denote the minimum number of antennas chosen by MAD principle. Our goal is to guarantee the overall diversity is no less than a given minimum diversity while keep $\tilde{N}_{min}$ small. Let $1 \leq g < N_r$ denote the given minimum achievable diversity order, based on (19) and the corresponding analysis in section II-A, we have

$$\begin{aligned} &if \quad d_{2MAD} < N_r \quad and \quad d_{2MAD} \geq g \\ &d_t = d_{2MAD} \geq g. \end{aligned} \tag{24}$$

Therefore we can guarantee that the overall diversity order of GPIC algorithm is no less than g, in order to derive the minimum number of antennas required based on MAD principle, we define function

$$f(N) = d_{2MAD} - g = (N+1)(N_r - N_t + N + 1) - g, \tag{25}$$

where $N \in [1, 2, \ldots, N_t]$, our goal is to find minimum N, that satisfy $f(N) \geq 0$. The two zero points of quadratic function $f(N)$ are

$$N_1 = -\sqrt{\frac{(N_r - N_t)^2}{4} + g} - \frac{N_r - N_t}{2} - 1, \tag{26}$$

$$N_2 = \sqrt{\frac{(N_r - N_t)^2}{4} + g} - \frac{N_r - N_t}{2} - 1 \tag{27}$$

when $N \leq N_1$ or $N \geq N_2$, $f(N) \geq 0$, obviously $N_1 < 0$, there is no feasible N exits that can make the former condition satisfied. Thus, we have

$$\tilde{N}_{min} = \lceil \sqrt{\frac{(N_r - N_t)^2}{4} + g} - \frac{N_r - N_t}{2} - 1 \rceil \tag{28}$$

C. *Computer Simulations*

In this section, computer simulation results for the bit error rate (BER) performances of MAD principle based GPIC algorithm are presented. As a comparison, Sel-MMSE-OSIC algorithm in [1] is also considered, denoted by maximum diversity selection (MDS) scheme. The software testbed is built by C, compiled by GCC compiler version 4.9.2 on 64 bit Debian (release 8.2) Linux systems, the experiments are performed on two desktops, one with Intel I5-4th generation CPU, quad cores, 3.2GHz clock rate and the other one with Intel I7-5th generation CPU, six cores, 3.5GHz clock rate.

We consider uncoded complex spatial multiplexing LS-MIMO systems, For each receive SNR point, the BER are calculated with minimum 10^5 independent channel realizations achieved and minimum 300 symbol errors accumulated. The modulation schemes considered are rectangular M-QAM(4-QAM and 16-QAM) with Gray code labeling. In each channel realization, N_t mutually independent randomly generated bit sequences are modulated to complex symbols. N_t complex symbols are transmitted over randomly generated Rayleigh flat fading channel, each channel matrix component is CSCG random variable with unit variance. The receive symbol vector is corrupted by AWGN.

For sake of simplicity, we use MAD-n to present the results for minimum achievable diversity selection, where n denotes the minimum achievable diversity. The minimum number of antennas selected by channel partition for a given minimum achievable diversity g is given by (28). Maximum diversity selection (MDS) is used to present the Sel-MMSE-OSIC that can achieve

ML diversity asymptotically. The minimum number of antennas selected in MDS is given by (23).

First, we consider 16×16 system with 4-QAM modulation, Table.I shows the information of the schemes considered, the corresponding results are presented in Fig.1,

TABLE I
16×16 SYSTEM

Scheme	number of antennas selected (N)	Asymptotic diversity
MDS	3	16
MAD-9	2	9
MAD-4	1	4

As shown in Fig.1, compared with MDS which can achieve ML performance asymptotically, MAD-9 has a small performance loss. While MAD-4's performance is inferior compared to MDS. At BER 10^{-5}, the MAD-9 is about 0.8dB worse than MDS, MAD-4 is about 1.8dB worse than MDS, let $\kappa(\cdot)$ denote the number of symbol vector candidates in the list generated by GPIC algorithm [1], from Table.I, we have $\frac{\kappa(MAD-9)}{\kappa(DMS)} = 1/4$ and $\frac{\kappa(MAD-4)}{\kappa(DMS)} = 1/16$.

The results for 16×16 system with 16-QAM modulation are shown in Fig.2, the information of the schemes considered is as the same as that in Talbe.I. From Fig.2, we can see that similar to the case of 4-QAM system, MAD-9 has a very small performance loss comparing to DMS, at BER 2×10^{-5}, MAD-9 is about 1.1dB worse than MDS, while MAD-4 has about 3dB performance loss comparing to DMS.

Furthermore, compared with 4-QAM systems, the reduction of the number of list candidates provided by MAD scheme is more significant in 16-QAM systems, $\frac{\kappa(MAD-9)}{\kappa(DMS)} = 1/16$ and $\frac{\kappa(MAD-4)}{\kappa(DMS)} = 1/256$.

Now we consider the performance of MAD in large systems, Fig.3 shows the results for 32×32 system with 4-QAM, Table.II shows the information of the schemes. Due to the time limitation, we just provide the results for MAD. As we can see in Fig.3, although the asymptotic diversity of MAD-16 is much higher than that of MAD-9, at the BER region of interest, their performance difference is negligible. At BER=3×10^{-4}, the performance of MAD-9 is about

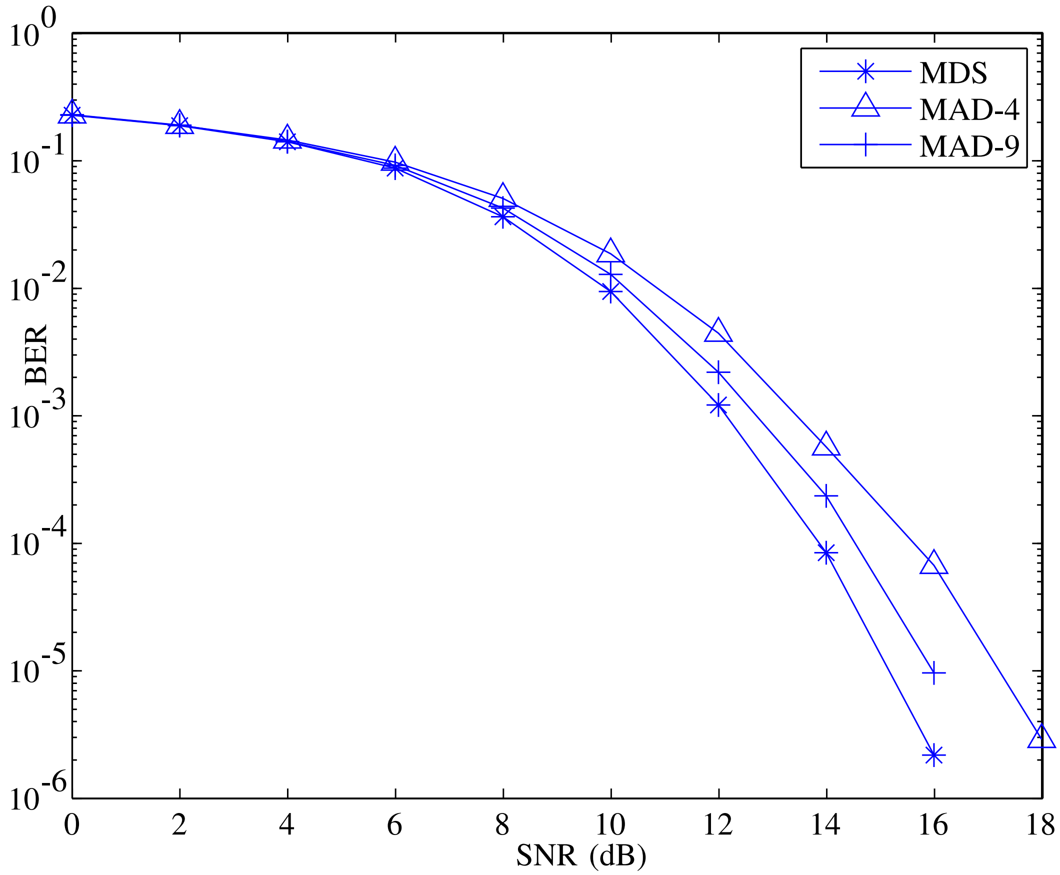

Fig. 1. 16×16 system 4-QAM

TABLE II
32×32 SYSTEM 4-QAM

Scheme	number of antennas selected (N)	Asymptotic diversity
MAD-4	1	4
MAD-9	2	9
MAD-16	3	16

0.3dB worse comparing to MAD-16, at BER=3×10^{-5}, the performance loss of MAD-9 is about 0.5dB comparing to MAD-16.

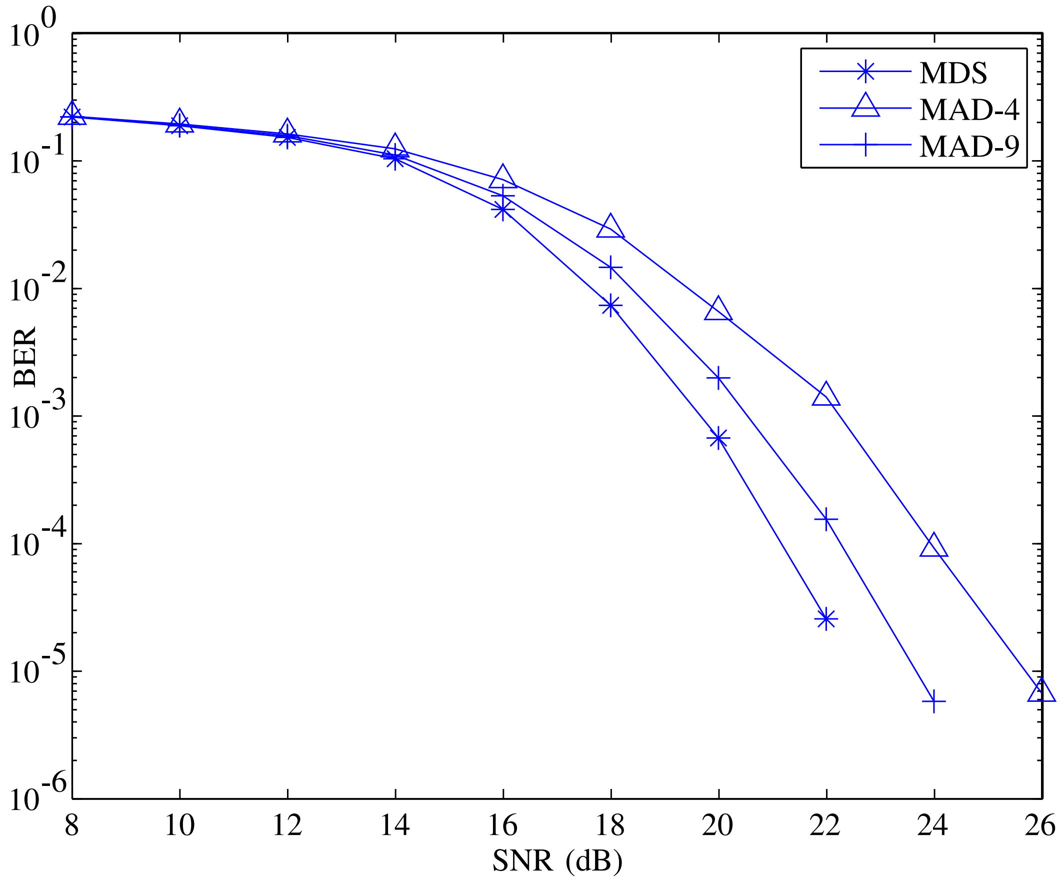

Fig. 2. 16×16 system 16-QAM

III. Theoretical Analysis of Channel Hardening Phenomenon

A. Preliminary

Orthogonality deficiency ϕ_{od} measures the orthogonality of a matrix [4], which is defined by

$$\phi_{od} = 1 - \frac{det(\mathbf{W})}{\prod_{i=1}^{N_t} ||\mathbf{h}_i||^2}, \tag{29}$$

where $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ denotes Wishart matrix, $\mathbf{h}_i$ denotes the i th column of $\mathbf{H}$, $det(\cdot)$ denotes determinant operation, $||\cdot||$ denotes 2-norm operation. Based on Hadamard's inequality $\prod_{i=1}^{N_t} ||\mathbf{h}_i|| \geq det(\mathbf{H})$, we have $0 \leq \phi_{od} \leq 1$, if $\mathbf{H}$ is singular, $\phi_{od} = 1$, if $\mathbf{H}$ is orthogonal, then $\phi_{od} = 0$.

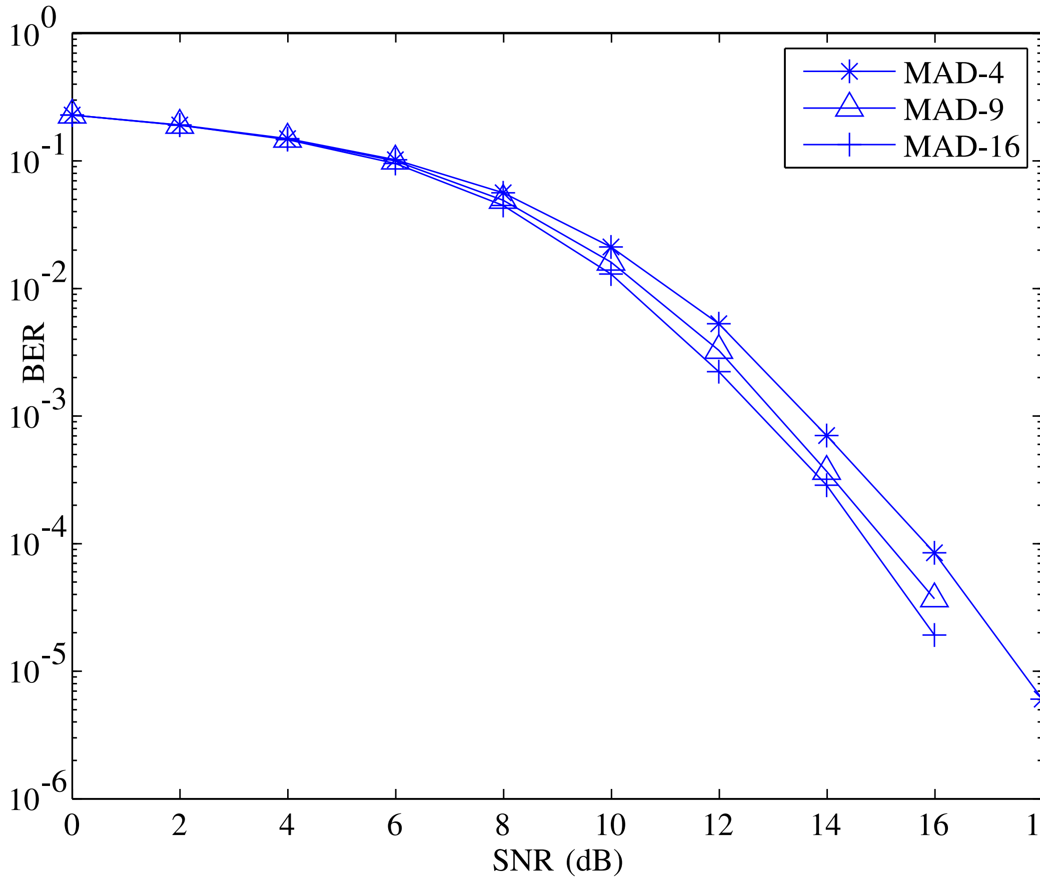

Fig. 3. 32×32 system 4-QAM

$||\mathbf{h}_i||^2 = \sum_{j=1}^{N_r} |\mathbf{H}_{ji}|^2$, where $|\cdot|$ denotes magnitude operation. $\mathbf{H}_{ij} \sim \mathbb{C}N(0,1)$, where $\mathbb{C}N(0,\sigma^2)$ denotes CSCG distribution with variance σ^2, thus $|\mathbf{H}_{ji}| \sim Rayleigh(1/\sqrt{2})$, where $Rayleigh(\sigma)$ denotes the Rayleigh distribution with shape parameter σ, therefore $||\mathbf{h}_i||^2 \sim Gamma(N_r, 1)$ [5]. $Gamma(k, \theta)$ denotes Gamma distribution, with k degrees of freedom and scale parameter θ.

For sake of simplicity, we define orthogonality measure ϕ_{om}, which is given by

$$\phi_{om} = \frac{det(\mathbf{W})}{\prod_{i=1}^{N_t} ||\mathbf{h}_i||^2}, \tag{30}$$

$0 \leq \phi_{om} \leq 1$, $\forall \mathbf{H}$, if ϕ_{om} is closer to 1, $\mathbf{H}$ is more closer to an orthogonal matrix.

B. Logarithmic Expectation of Orthogonality Measure

Based on (30), The logarithm of ϕ_{om} can be written as

$$\ln(\phi_{om}) = \ln(det(\mathbf{W})) - \sum_{i=1}^{N_t} \ln(||\mathbf{h}_i||^2), \tag{31}$$

Taking expectation of (31), we have

$$\mathbb{E}[\ln(\phi_{om})] = \mathbb{E}[\ln(det(\mathbf{W}))] - \sum_{i=1}^{N_t} \mathbb{E}[\ln(||\mathbf{h}_i||^2)]. \tag{32}$$

First we consider the first summand on the right hand of (32), $\mathbb{E}[\ln(det(\mathbf{W}))]$. Let $\mathbb{C}W_m(n, \Sigma)$ denote complex Wishart distribution, which is the joint distribution of sample covariance matrix from multivariate complex Gaussian random variable [6]. n is the degree of freedom and $\Sigma \in \mathbb{C}^{m\times m}$ is the covariance matrix. Let Π_i denotes the ith row of $\mathbf{H}$. Π_i is a N_t-variate complex Gaussian random variable, so does Π_i^H. Therefore $\mathbf{W} = \mathbf{H}^H\mathbf{H} = \sum_{i=1}^{N_r} \Pi_i^H \Pi_i \sim \mathbb{C}W_{N_t}(N_r, \mathbf{I}_{N_t})$. The logarithmic expectation of $det(\mathbf{W})$ is

$$\mathbb{E}[\ln(det(\mathbf{W}))] = \sum_{i=1}^{N_t} \psi(N_r - i + 1), \tag{33}$$

where $\psi(n)$ denotes Digamma function, which is given by [5]

$$\psi(n) = \frac{\Gamma'(n)}{\Gamma(n)}, \tag{34}$$

$\Gamma(n)$ denotes Gamma function [5].

Proof. : see Appendix A.

Now we consider the second summand on the right hand of (32), $\sum_{i=1}^{N_t} \mathbb{E}[\ln(||\mathbf{h}_i||^2)]$. $||\mathbf{h}_i||^2 \sim Gamma(N_r, 1)$, the logarithmic expectation of a Gamma random variable $\gamma \sim Gamma(n, \theta)$ can be written as:

$$\mathbb{E}[\ln(\gamma)] = \psi(n) + \ln(\theta), \tag{35}$$

Thus (35), we have

$$\sum_{i=1}^{N_t} \mathbb{E}[\ln(||\mathbf{h}_i||^2)] = \sum_{i=1}^{N_t} \psi(N_r), \tag{36}$$

Proof. : see Appendix B.

Based on (32), (33) and (36), the logarithmic expectation of orthogonality measure ϕ_{om} is

$$\mathbb{E}[\ln(\phi_{om})] = \sum_{i=1}^{N_t}[\psi(N_r - i + 1) - \psi(N_r)], \tag{37}$$

C. Probability Density Function of Orthogonality Measure

In this section, we derive the probability distribution of ϕ_{om}. First, let's consider an alternative form of (30), do QR factorization to $\mathbf{H}$, $\mathbf{H} = \mathbf{QR}$ [7], where $\mathbf{Q} \in \mathbb{C}^{N_r \times N_t}$ is a unitary matrix and $\mathbf{R} \in \mathbb{C}^{N_t \times N_t}$ is an upper triangular matrix. Thus $\mathbf{W} = \mathbf{H}^H\mathbf{H} = \mathbf{R}^H\mathbf{R}$. Let $r_{ji}, j \leq i$ denote the component of $\mathbf{R}$ at jth row and ith column, r_{ji}^* denotes the conjugate of r_{ji}, we have

$$det(\mathbf{W}) = det(\mathbf{R}^H\mathbf{R}) = det(\mathbf{R}^H)det(\mathbf{R}) = \prod_{i=1}^{N_t} r_{ii}^* \prod_{i=1}^{N_t} r_{ii} = \prod_{i=1}^{N_t} |r_{ii}|^2, \tag{38}$$

Furthermore, let $\mathbf{R}_i$ denote the ith column of $\mathbf{R}$, $\mathbf{W}_{ii}$ denote the ith diagonal component of $\mathbf{W}$, we have

$$\mathbf{W}_{ii} = ||\mathbf{h}_i||^2 = ||\mathbf{R}_i||^2 = \sum_{j<i} |r_{ji}|^2 + |r_{ii}|^2, \tag{39}$$

based on (38) and (39), (30) can be rewritten as

$$\phi_{om} = \prod_{i=1}^{N_t}\left(\frac{|r_{ii}|^2}{|r_{ii}|^2 + \sum_{j<i}|r_{ji}|^2}\right). \tag{40}$$

Based on [8], given $\mathbf{W} \sim \mathbb{C}W_{N_t}(N_r, \mathbf{I}_{N_t})$ and $\mathbf{W} = \mathbf{R}^H\mathbf{R}$, r_{ji} are mutually independent distributed. Furthermore, for $1 \leq i \leq N_t$, $|r_{ii}|^2 \sim Gamma(N_r - i + 1, 1)$, for $1 \leq j < i \leq N_t$, $r_{ji} \sim \mathbb{C}N(0, 1)$. Because $|r_{ji}| \sim Rayleigh(1/\sqrt{2})$, $\sum_{j<i}|r_{ji}|^2 \sim Gamma(i-1, 1)$ $(i \geq 2)$. Define

$$\alpha_i = |r_{ii}|^2 \sim Gamma(k_1^i, 1), \quad k_1^i = N_r - i + 1, \quad i = 1, 2 \ldots N_t \tag{41}$$

$$\beta_i = \begin{cases} 0 \quad i = 1 \\ \sum_{j<i}|r_{ji}|^2 \sim Gamma(k_2^i, 1), \quad k_2^i = i - 1, \quad i = 2 \ldots N_t \end{cases} \tag{42}$$

(40) can be rewritten as

$$\phi_{om} = \prod_{i=1}^{N_t} \frac{\alpha_i}{\alpha_i + \beta_i}, \tag{43}$$

α_i and β_i ($i \geq 2$) are independent Gamma random variables, from [9], if $X \sim Gamma(k_1, \theta)$ and $Y \sim Gamma(k_2, \theta)$, then $\frac{X}{X+Y} \sim Beta(k_1, k_2)$, where $Beta(k_1, k_2)$ denotes Beta distribution with shape parameters $k_1, k_2 > 0$. Consider (43), if $i = 1$, $\beta_i = 0$, therefore $\frac{\alpha_1}{\alpha_1+\beta_1} = 1$. Thus, we have $\frac{\alpha_i}{\alpha_i+\beta_i} \sim Beta(k_1^i, k_2^i)$ when $i \geq 2$. Define $\eta_i = \frac{\alpha_i}{\alpha_i+\beta_i}, i = 1, 2, \ldots, N_t$, $\eta_1 = 1$, and when $i \geq 2$, η_i are mutually independent Beta random variables. Therefore (40) can be rewritten as

$$\phi_{om} = \prod_{i=1}^{N_t} \eta_i = \prod_{i=2}^{N_t} \eta_i, \tag{44}$$

Thus ϕ_{om} is the product of N_t independent Beta random variables, the p.d.f of ϕ_{om} is given by

$$f_{\phi_{om}}(\rho) = \sum_{\mathbf{j}} \{[\prod_{i=2}^{N_t} c(k_1^i, k_2^i, j_i)] U(\rho|\mathbf{k}_1 + \mathbf{j})\}, \tag{45}$$

where $\sum_{\mathbf{j}} = \sum_{j_2} \sum_{j_3} \ldots \sum_{j_{N_t}}$, $j_i \in [0, 1, \ldots, k_2^i - 1]$. $c(k_1^i, k_2^i, j_i) = (-1)^{j_i} \binom{k_2^i - 1}{j_i} [(k_1^i + j_i)\mathbb{B}(k_1^i, k_2^i)]^{-1}$, $\mathbb{B}(\alpha, \beta)$ denotes Beta function with parameters α and β. $\mathbf{k}_1 + \mathbf{j} = [k_1^2 + j_1, k_1^3 + j_2, \ldots, k_1^{N_t} + j_{N_t}]$,

$$U(\rho|\mathbf{k}_1 + \mathbf{j}) = \rho^{-1} \prod_{i=2}^{N_t} (k_1^i + j_i) \sum_{i=2}^{N_t} [\rho^{k_1^i + j_i} / \prod_{j \neq i}^{N_t} (k_1^j + j_j - k_1^i - j_i)], \tag{46}$$

Proof. : see Appendix C.

If a Beta random variable $\nu \sim Beta(k_1, k_2)$, then $\mathbb{E}[\ln(\nu)] = \psi(k_1) - \psi(k_1 + k_2)$ [5], take logarithmic expectation of (44), we have

$$\mathbb{E}[\ln(\phi_{om})] = \sum_{i=2}^{N_t} \mathbb{E}[\ln(\eta_i)] = \sum_{i=2}^{N_t} [\psi(k_1^i) - \psi(k_1^i + k_2^i)] = \sum_{i=2}^{N_t} [\psi(N_r - i + 1) - \psi(N_r)] = \sum_{i=1}^{N_t} [\psi(N_r - i + 1) - \psi(N_r)], \tag{47}$$

which is consistent with (37).

D. Computer Simulations

Computer simulations are made to demonstrate the correctness of the results in this section. The experiments are performed by Matlab, on a desktop with I5-4th generation CPU, quad cores, 3.2GHz clock rate.

Different sizes of channel matrices are considered, with $5 \leq N_r \leq 100$ and $5 \leq N_t \leq N_r$. The theoretical logarithmic expectation of ϕ_{om}, denoted by $\mathbb{E}[\ln(\phi_{om})]_t$ are calculated based on (37), the result is shown in Fig.4. The empirical estimation of the logarithmic expectation of ϕ_{om}, denoted by $\mathbb{E}[\ln(\phi_{om})]_{em}$, are calculated based on taking average over 10^5 independent channel realizations of each size of channel matrix. In each realization, the channel matrix are generated randomly, each component of the channel matrix is CSCG random variable with unit variance. The result is shown in Fig.5.

The variance between $\mathbb{E}[\ln(\phi_{om})]_{em}$ and $\mathbb{E}[\ln(\phi_{om})]_t$, denoted by υ, is also calculated by

$$\upsilon = \frac{1}{m}\sum_{i=1}^{m}(\mathbb{E}[\ln(\phi_{om})]_{em}^{i} - \mathbb{E}[\ln(\phi_{om})]_{t}^{i})^2, \tag{48}$$

where m is the number of the different sizes of channel matrix considered. By simulation $\upsilon = 8.0116 \times 10^{-4}$, which demonstrates the correctness of the result in (37).

IV. CONCLUSION

The conclusion goes here.

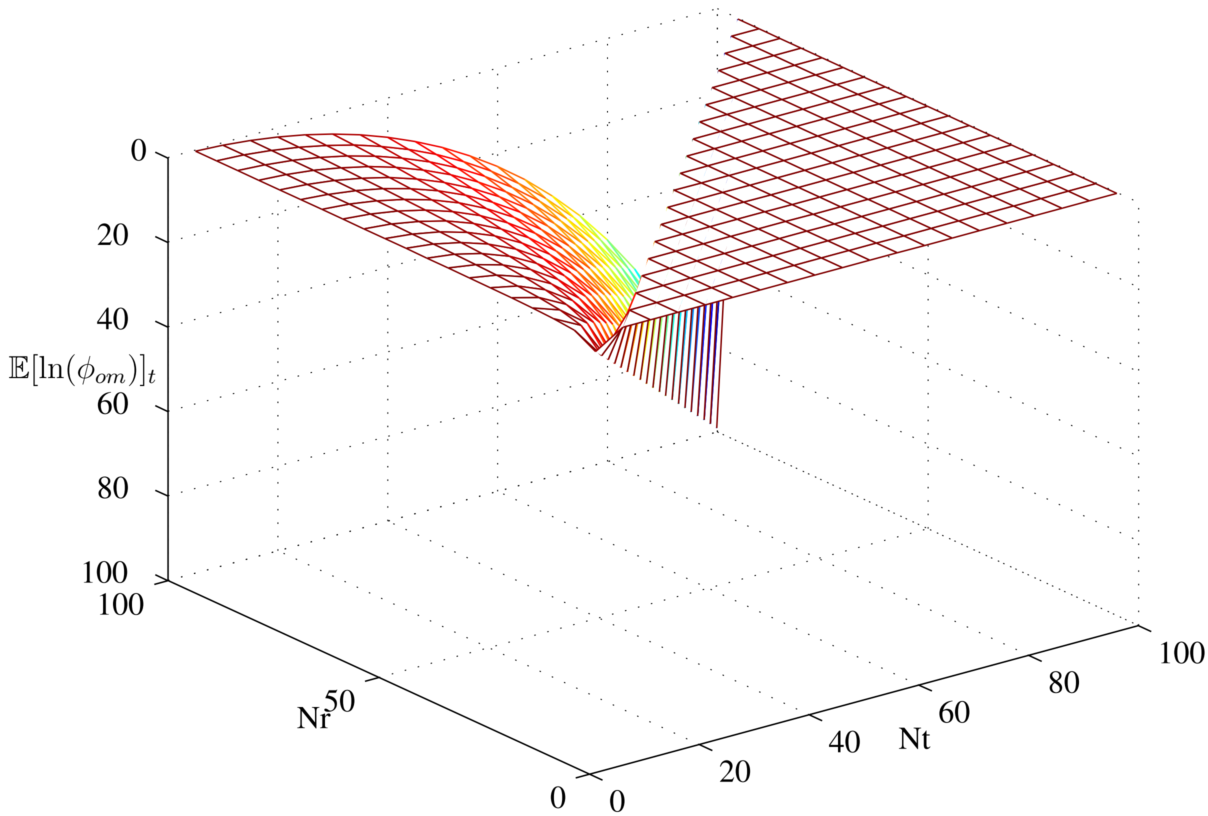

Fig. 4. Theoretical logarithmic expectation of ϕ_{om}

Appendix A

Let $\mathbf{A} \in \mathbb{C}^{m \times m}$, $A \sim \mathbb{C}W_m(n, \boldsymbol{\Sigma})$, According to the definition of complex Wishart matrix, it is obvious $\mathbf{A}$ is Hermition positive definite matrix (i.e., $\mathbf{A} = \mathbf{A}^H > 0$). Define $etr(\mathbf{A}) = e^{tr(\mathbf{A})}$, $tr(\mathbf{A}) = \mathbf{A}_{11} + \mathbf{A}_{22} + \cdots + \mathbf{A}_{mm}$.

The p.d.f. of $\mathbf{A}$ can be written as [8]:

$$f(\mathbf{A}) = \{\tilde{\Gamma}_m(n) det(\boldsymbol{\Sigma})^n\}^{-1} det(\mathbf{A})^{n-m} etr(-\boldsymbol{\Sigma}^{-1}\mathbf{A}), \tag{49}$$

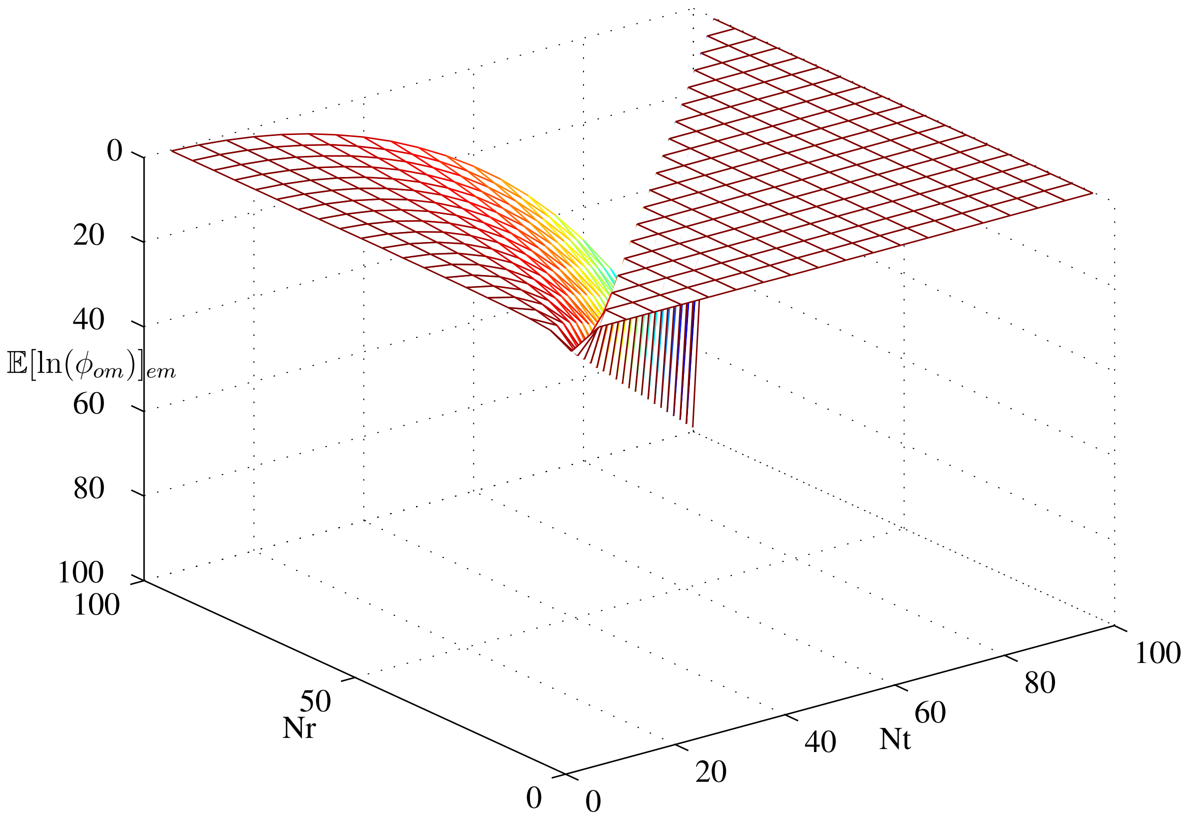

Fig. 5. Empirical estimation of the logarithmic expectation of ϕ_{om}

where $\tilde{\Gamma}_m(\beta)$ denotes multivariate complex Gamma function defined by [8]:

$$\tilde{\Gamma}_m(\beta) = \pi^{\frac{m(m-1)}{2}} \prod_{i=1}^{m} \Gamma(\beta - i + 1) \quad Re(\beta) > m - 1. \tag{50}$$

Furthermore, from [8], we have

$$\tilde{\Gamma}_m(\beta) = \int_{\mathbf{X}=\mathbf{X}^H>0} etr(-\mathbf{X}) det(\mathbf{X})^{\beta-m} d\mathbf{X} \quad Re(\beta) > m - 1. \tag{51}$$

We derive logarithmic expectation of $det(\mathbf{A})$

$$
\begin{aligned}
\mathbb{E}[\ln(det(\mathbf{A}))] &= \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A})) f(\mathbf{A}) d\mathbf{A} \\
&= \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A})) \{\tilde{\Gamma}_m(n) det(\mathbf{\Sigma})^n\}^{-1} det(\mathbf{A})^{n-m} etr(-\mathbf{\Sigma}^{-1}\mathbf{A}) d\mathbf{A} \\
&= \frac{det(\mathbf{\Sigma})^{-n}}{\tilde{\Gamma}_m(n)} \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A})) det(\mathbf{A})^{n-m} etr(-\mathbf{\Sigma}^{-1}\mathbf{A}) d\mathbf{A}, \qquad (52)
\end{aligned}
$$

if $\mathbf{\Sigma} = \mathbf{I}$, (52) can be written as

$$
\mathbb{E}[\ln(det(\mathbf{A}))] = \frac{1}{\tilde{\Gamma}_m(n)} \int_{\mathbf{A}=\mathbf{A}^H>0} \ln(det(\mathbf{A})) det(\mathbf{A})^{n-m} etr(-\mathbf{A}) d\mathbf{A}. \qquad (53)
$$

Because $\frac{d}{dn}[det(\mathbf{A})]^{n-m} = \ln(det(\mathbf{A})) det(\mathbf{A})^{n-m}$, (53) can be rewritten as

$$
\mathbf{E}[\ln(det(\mathbf{A}))] = \frac{1}{\tilde{\Gamma}_m(n)} \frac{d}{dn} \int_{\mathbf{A}=\mathbf{A}^H>0} etr(-\mathbf{A}) det(\mathbf{A})^{n-m} d\mathbf{A}, \qquad (54)
$$

Based on (51), in (54), we have

$$
\tilde{\Gamma}'_m(n) = \frac{d}{dn} \int_{\mathbf{A}=\mathbf{A}^H>0} etr(-\mathbf{A}) det(\mathbf{A})^{n-m} d\mathbf{A}, \qquad (55)
$$

Therefore (54) can be rewritten as

$$
\mathbf{E}[\ln(\mathbf{A})] = \frac{\tilde{\Gamma}'_m(n)}{\tilde{\Gamma}_m(n)}. \qquad (56)
$$

Based on (50), we have

$$
\tilde{\Gamma}'_m(n) = \pi^{\frac{m(m-1)}{2}} \sum_{i=1}^{m} [\Gamma'(n-i+1) \prod_{j \neq i}^{m} \Gamma(n-j+1)], \qquad (57)
$$

Using (50) and (57), we have

$$
\begin{aligned}
&\frac{\tilde{\Gamma}'_m(n)}{\tilde{\Gamma}_m(n)} = \frac{\sum_{i=1}^{m} [\Gamma'(n-i+1) \prod_{j \neq i}^{m} \Gamma(n-j+1)]}{\prod_{k=1}^{m} \Gamma(n-k+1)} \\
&= \sum_{i=1}^{m} [\frac{\Gamma'(n-i+1) \prod_{j \neq i}^{m} \Gamma(n-j+1)}{\prod_{k=1}^{m} \Gamma(n-k+1)}] = \sum_{i=1}^{m} \frac{\Gamma'(n-i+1)}{\Gamma(n-i+1)}, \qquad (58)
\end{aligned}
$$

Therefore (56) can be rewritten as

$$\mathbf{E}[\ln(det(\mathbf{A}))] = \sum_{i=1}^{m} \psi(n-i+1), \tag{59}$$

where $\psi(n) = \frac{\Gamma'(n)}{\Gamma(n)}$ denotes Digamma function.

APPENDIX B

If $x \sim Gamma(n, \theta)$, with shape parameter k and scale parameter θ, $x > 0$, $\Gamma(k)$ denotes Gamma function, the density function of Gamma distribution is

$$f(x,k,\theta) = \frac{x^{k-1}e^{-x/\theta}}{\Gamma(k)\theta^k}. \tag{60}$$

where $\Gamma(n)$ satisfies [5]

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx, \tag{61}$$

Thus the logarithmic expectation of x can be written as

$$\mathbf{E}[\ln(x)] = \frac{1}{\Gamma(k)} \int_0^\infty \ln(x)x^{k-1}e^{-x/\theta}\theta^{-k}dx, \tag{62}$$

define $z = x/\theta$, (62) can be rewritten as

$$\mathbb{E}[\ln(x)] = \ln(\theta)\frac{1}{\Gamma(k)} \int_0^\infty z^{k-1}e^{-z}dz + \frac{1}{\Gamma(k)} \int_0^\infty \ln(z)z^{k-1}e^{-z}dz, \tag{63}$$

Based on (61), (63) can be rewritten as

$$\mathbf{E}[\ln(x)] = \ln(\theta) + \frac{1}{\Gamma(k)} \int_0^\infty \ln(z)z^{k-1}e^{-z}dz. \tag{64}$$

Because $\frac{d(z^{k-1})}{dk} = \ln(z)z^{k-1}$, (64) can be rewritten as

$$\mathbf{E}[\ln(x)] = \ln(\theta) + \frac{1}{\Gamma(k)}\frac{d}{dk} \int_0^\infty z^{k-1}e^{-z}dz, \tag{65}$$

Based on (61), we have

$$\Gamma'(k) = \frac{d}{dk} \int_0^\infty z^{k-1}e^{-z}dz, \tag{66}$$

Thus (65) can be rewritten as

$$\mathbf{E}(\ln(x)) = \ln(\theta) + \frac{\Gamma'(k)}{\Gamma(k)} = \ln(\theta) + \psi(k), \tag{67}$$

where $\psi(k)$ denotes Digamma function.

APPENDIX C

Define $x = \prod_{i=1}^{n} x_i$, $x_1, x_2, \ldots, x_n$ are independent Beta random variables, where $x_i \sim Beta(k_1^i, k_2^i)$, the p.d.f. of x_i is given by

$$f_{x_i}(\rho) = \frac{1}{\mathbb{B}(k_1^i, k_2^i)} \rho^{k_1^i - 1} (1-\rho)^{k_2^i - 1}, \tag{68}$$

where $\mathbb{B}(k_1^i, k_2^i)$ denotes Beta function with parameters k_1^i and k_2^i. Define $y_i = -\ln(x_i) = g(x_i)$, Based on Jacobian transformation, we have

$$f_{y_i}(\rho) = |\frac{dy_i}{dx_i}|^{-1} f_{x_i}(g^{-1}(\rho)) = \frac{1}{\mathbb{B}(k_1^i, k_2^i)} e^{-k_1^i \rho} (1 - e^{-\rho})^{k_2^i - 1}. \tag{69}$$

Based on Taylor series, (69) can be alternatively expressed as [10]

$$f_{y_i}(\rho) = \sum_{j_i=0}^{k_2^i - 1} c(k_1^i, k_2^i, j_i)(k_1^i + j_i) e^{-(k_1^i + j_i)\rho}, \tag{70}$$

where $c(k_1^i, k_2^i, j_i) = (-1)^{j_i} \binom{k_2^i - 1}{j_i} [(k_1^i + j_i)\mathbb{B}(k_1^i, k_2^i)]^{-1}$, $j_i \in [0, 1, \ldots, k_2^i - 1]$, From (70), one can conclude that $f_{y_i}(\rho)$ is a weighted summation of the p.d.f. of exponential distributions.

If $\tau_1, \tau_2, \ldots, \tau_n$ are independent exponential random variables, where $\tau_i \sim exp(t_i)$, the p.d.f. of τ_i is given by

$$f_{\tau_i}(\rho) = t_i e^{-t_i \rho}, \quad \rho \geq 0, \tag{71}$$

define $\tau = \sum_{i=1}^{n} \tau_i$, by induction, the p.d.f of τ is given by [10]

$$f_{\tau}(\rho) = f(\rho|\mathbf{t}) = \prod_{i=1}^{n} t_i \sum_{i=1}^{n} [e^{-t_i \rho} / \prod_{j \neq i}^{j=n} (t_j - t_i)], \tag{72}$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_n]$. Define $y = \sum_{i=1}^{n} y_i = -\ln(\prod_{i=1}^{n} x_i) = -\ln(x)$. Based on (70) and

(72), the p.d.f. of y is given by [10]

$$f_y(\rho) = \sum_{\mathbf{j}} \{[\prod_{i=1}^{n} c(k_1^i, k_2^i, j_i)] f(\rho|\mathbf{k_1} + \mathbf{j})\}, \tag{73}$$

where $\sum_{\mathbf{j}} = \sum_{j_1} \sum_{j_2} \ldots \sum_{j_n}$, $j_i \in [0, 1, \ldots, k_2^i - 1]$, $\mathbf{k_1} + \mathbf{j} = [k_1^1 + j_1, k_1^2 + j_2 \ldots k_1^n + j_n]$. Based on (73) and $x = e^{-y} = G(y)$, using Jacobian transformation, the p.d.f. of x is given by

$$f_x(\rho) = f_y(G^{-1}(\rho))|\frac{dx}{dy}|^{-1} = f_y(-\ln(\rho))\rho^{-1} = \sum_{\mathbf{j}} \{[\prod_{i=1}^{n} c(k_1^i, k_2^i, j_i)] f(-\ln(\rho)|\mathbf{k}_1 + \mathbf{j})\rho^{-1}\}, \tag{74}$$

define

$$U(\rho|\mathbf{k}_1 + \mathbf{j}) = f(-\ln(\rho)|\mathbf{k}_1 + \mathbf{j})\rho^{-1}, \tag{75}$$

Thus (74) can be rewritten as

$$f_x(\rho) = \sum_{\mathbf{j}} \{[\prod_{i=1}^{n} c(k_1^i, k_2^i, j_i)] U(\rho|\mathbf{k}_1 + \mathbf{j})\}. \tag{76}$$

References

[1] D. Radji and H. Leib, “Interference cancellation based detection for v-blast with diversity maximizing channel partition,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 6, pp. 1000–1015, 2009.

[2] L. Zheng and D. N. Tse, “Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels,” *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1073–1096, 2003.

[3] H. Zhang, H. Dai, Q. Zhou, and B. L. Hughes, “On the diversity order of spatial multiplexing systems with transmit antenna selection: A geometrical approach,” *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5297–5311, 2006.

[4] X. Ma and W. Zhang, “Performance analysis for MIMO systems with lattice-reduction aided linear equalization,” *Communications, IEEE Transactions on*, vol. 56, no. 2, pp. 309–318, 2008.

[5] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[6] N. Goodman, “Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction),” *Annals of mathematical statistics*, pp. 152–177, 1963.

[7] D. S. Watkins, *Fundamentals of matrix computations*. John Wiley & Sons, 2004, vol. 64.

[8] D. K. Nagar and A. K. Gupta, “Expectations of functions of complex Wishart matrix,” *Acta applicandae mathematicae*, vol. 113, no. 3, pp. 265–288, 2011.

[9] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC Press, 2004.

[10] R. Bhargava and C. Khatri, “The distribution of product of independent beta random variables with application to multivariate analysis,” *Annals of the Institute of Statistical Mathematics*, vol. 33, no. 1, pp. 287–296, 1981.