On Capacity of Noncoherent MIMO with Asymmetric Link Strengths

Joyson Sebastian, Ayan Sengupta, Suhas N. Diggavi

***Abstract*—We study the generalized degrees of freedom (gDoF) of the block-fading noncoherent MIMO channel with asymmetric distributions of link strengths, and a coherence time of T symbol durations. We first derive the optimal signaling structure for communication over this channel, which is distinct from that for the i.i.d MIMO setting. We prove that for $T=1$, the gDoF is zero for MIMO channels with arbitrary link strength distributions, extending the result for MIMO with i.i.d links. We then show that selecting the statistically best antenna is gDoF-optimal for both Multiple Input Single Output (MISO) and Single Input Multiple Output (SIMO) channels. We also derive the gDoF for the 2×2 MIMO channel with different exponents in the direct and cross links. In this setting, we show that it is always necessary to use both antennas to achieve the optimal gDoF, in contrast to the results for 2×2 MIMO with identical link distributions. We also show that having weaker crosslinks gives gDoF gain compared to the case with identically distriuted links.**

I. Introduction

The capacity of fading MIMO channels when neither the receiver nor the transmitter knows the fading coefficients was first studied by Marzetta and Hochwald [3]. They considered a block fading channel model where the fading gains are i.i.d. Rayleigh distributed and remain constant for T symbol periods. In [6], Zheng and Tse introduced the idea of communication over the Grassmanian manifold for the noncoherent MIMO channel and derived the high SNR behavior of capacity with i.i.d channels.

In this paper, we consider a channel model with asymmetric link distributions, where the link strengths are scaled with different exponents of SNR. In essence, we are moving from the DoF-framework in [3],[6] to the generalized DoF of noncoherent MIMO channels. The channels could have asymmetric link strengths either due to asymmetry in the reflective environment or due to larger antenna spacing. It is also motivated by a fundamental question about the robustness of the results in [3],[6] to changes in the i.i.d channel model.

For our channel model with arbitrary (fading) link strengths, we show in Theorem 1 that the capacity achieving input distribution is of the form LQ where L is lower triangular and Q is independent of L and is unitary isotropically distributed. This is in contrast to the result for the i.i.d setting, which yields a diagonal matrix instead of L multiplying Q [3]. For $T=1$, we show in Theorem 2 that the gDoF is always zero for MIMO of any size. In Theorem 6, we show that the gDoF of the MISO channel can be achieved by only signaling over the (statistically) best transmit antenna. Also for the SIMO case, we show that the gDoF can be achieved by only retaining the signal received by the best receive antenna (Theorem 5).

When the exponents in the SNR-scaling are same for all channels (i.i.d setting), the number of transmit antennas M, required to attain the optimal DoF was shown to be

$$\min\left(\left\lfloor\frac{T}{2}\right\rfloor, N\right)$$

with N receive antennas, in [6]. They showed that increasing the number of transmit antennas beyond this value reduces the DoF. In this paper, we provide evidence that this is not the case when the SNR exponents are different: in Theorem 7 we show that for a 2×2 MIMO with different exponents in direct and cross links, and $T=2$, both transmit antennas are required to achieve the optimal gDoF. We also show that having smaller exponents in cross links lead to gDoF gain of $\frac{2}{T}\gamma_{\text{diff}}$ over the case with same exponents in all links, where γ_{diff} is the difference in the SNR exponents. In showing this, several novel techniques were needed. In particular we would like to highlight the technique used in Lemma 16, where in the optimization problem to find the optimal input distribution for the outerbound, we show that the optimal gDoF can be achieved by a point mass distribution. To arrive at this, we discretized the input distribution without a loss in gDoF, and subsequently used linear programming arguments to show that there exists an optimal distribution with just one mass point.

The rest of this paper is organized as follows: in Section II we set up the system model and notation; Section III presents our main results with the detailed proofs in [5], and Section IV provides some analysis and proof sketches. In some cases, discussion of results in Section III will refer to lemmas and facts detailed in Section IV.

II. System Model and Notation

We consider a block-fading MIMO channel with M transmit and N receive antennas, and a coherence time of T symbol durations. The signal flow (over a blocklength T) is given by:

$$Y = GX + W$$

where X is the $M\times T$ matrix of transmitted symbols with rows $\mathsf{Tran}\left(\overline{X_i}\right)$ corresponding to each transmit antenna ($\overline{X_i}$ being the notation for column vectors and Tran indicates transpose of a matrix); G represents the $N\times M$ channel matrix (which is independently generated every T symbols), and its elements g_{ij} are independent with $g_{ij}\sim\mathcal{CN}\left(0,\rho_{ij}^2\right)=\mathcal{CN}\left(0,\mathsf{SNR}^{\gamma_{ij}}\right)$, where the exponents $\gamma_{ij}>0$ are (constant) parameters of the MIMO channel (for convenience, we also use the notation

This work was supported in part by NSF grants 1514531 and 1314937.

$\rho^2(n)$ to denote the vector of channel strengths to n^{th}receiver antenna); Y represents the $N \times T$ matrix of received symbols, with rows corresponding to each receive antenna; and W is an $N \times T$ noise matrix with elements $w_{ij} \sim$ i.i.d $\mathcal{CN}(0,1)$. The transmit signals have the average power constraint:

$$\frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\mathbb{E}\left[|x_{mt}|^2\right]=1. \tag{1}$$

The logarithm to base 2 is denoted by $\log()$. The notation x^+ indicates $\max(x,0)$, $A^\dagger$ indicates the Hermitian conjugate of a matrix A. We study the generalized degrees of freedom for MIMO given by $\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log(\mathsf{SNR})}$ where $C(\mathsf{SNR})$ is the capacity for a given value of SNR . We use the notation $\doteq$ for relative equality, i.e., we say $f_1(\mathsf{SNR}) \doteq f_2(\mathsf{SNR})$ if

$$\lim_{SNR\to\infty}\frac{f_1(\mathsf{SNR})}{\log(\mathsf{SNR})}=\lim_{SNR\to\infty}\frac{f_2(\mathsf{SNR})}{\log(\mathsf{SNR})}.$$

$\dot{\leq}, \dot{\geq}$ are defined analogously. The script $\mathcal{P}$ is used to indicate an optimization problem and $(\mathcal{P})$ is used to denote the optimal value of the objective function.

III. Main results

Theorem 1. *The capacity of the noncoherent MIMO channel can be achieved with X of the form $X = LQ$ with L being lower triangular and Q being an isotropically distributed unitary matrix, independent of L.*

Proof: The proof is given in subsection IV-B. ∎

This theorem is in contrast to the result for G with i.i.d. Gaussian g_{ij} which yielded the structure of X as $X = DQ$ where D is diagonal [3]. In our system model only W has i.i.d. elements, which ends up restricting the structure to the form LQ.

Theorem 2. *(gDoF of arbitrary noncoherent MIMO for $T = 1$) For any G with $T = 1$, the gDoF is zero.*

Proof: The proof is in Appendix A of [5]. ∎

Note that the above is a generalization of the zero DoF result for MIMO from [2]. In their model, the channels are constant and the power of the i.i.d. noise is scaled, but our results are more general, in the sense it allows the channel coefficients to be scaled with arbitrary exponents.

Theorem 3. *Let the channel matrix G be block diagonal as $G = diag(G_1, \ldots, G_K)$ where G_i are the diagonal blocks of G. Then, the capacity $C(P, diag(G_1, \ldots, G_K))$ of the channel for a power constraint P, can be achieved be splitting the power across the blocks: $C(P, diag(G_1, \ldots, G_K)) = \max_{P_1+\cdots+P_K\leq P}(C(P_1, G_1) + \cdots + C(P_K, G_K))$.*

Proof: This result holds for coherent MIMO and the proof for the noncoherent case is similar. We need to show $C(P, \mathrm{diag}(G_1, G_2)) = \max_{P_1+P_2\leq P}(C(P_1, G_1) + C(P_2, G_2))$, and the rest follows by induction. Let X_{G1}, X_{G2} be the transmitted symbols in G_1 and G_2 segments of the channel. Similarly Y_{G1}, Y_{G2} are the corresponding received symbols. Now $I(X;Y) \leq I(X_{G1};Y_{G1}) + I(X_{G2};Y_{G2})$ follows because $(X_{G2}, Y_{G2}) - X_{G1} - Y_{G1}$, $(X_{G1}, Y_{G1}) - X_{G2} - Y_{G2}$ are Markov chains, and the desired result follows. See Appendix F of [5] for the detailed proof. ∎

Corollary 4. *The gDoF of the noncoherent parallel channel with $G = diag\begin{pmatrix} g_{11} & . & . & g_{MM}\end{pmatrix}$, where $g_{ii} \sim \mathcal{CN}(0, \rho_{ii}^2) = \mathcal{CN}(0, \mathsf{SNR}^{\gamma_{ii}})$, is $\sum_i\left(1-\frac{1}{T}\right)\gamma_{ii}$.*

Theorem 5. *The gDoF of the noncoherent SIMO channel with $G = \mathsf{Tran}\left(\begin{bmatrix} g_{11} & . & . & g_{N1}\end{bmatrix}\right)$, where $g_{i1} \sim \mathcal{CN}(0, \rho_{i1}^2) = \mathcal{CN}(0, \mathsf{SNR}^{\gamma_{i1}})$ is $\left(1-\frac{1}{T}\right)\max_i \gamma_{i1}$, i.e., the gDoF can be achieved by using only the statistically best receive antenna.*

Proof: See Appendix C of [5] for the detailed proof. For the proof, we assume $\max_i \gamma_{i1} = \gamma_{11}$ without loss of generality.

One of the key challenges in the outerbound is to deal with the the received signal of the form $Y = \begin{bmatrix} ag_{11}+w_{11} & w_{12} & . & w_{1T} \\ ag_{21}+w_{21} & w_{22} & . & w_{2T} \\ . & . & . & . \\ ag_{N1}+w_{N1} & w_{N2} & . & w_{NT}\end{bmatrix} Q$ (with Q being isotropic unitary distributed) and evaluate its entropy. Simple bounds based on i.i.d. Gaussians maximizing the entropy of a vector/matrix with given power constraints turn out to be loose. Our main idea in obtaining a tighter outer bound is to perform an LQ transformation so that we get $Y = \begin{bmatrix} \xi_{11} & 0 & 0 & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & 0 & . & 0 \\ . & . & . & 0 & . & . \\ \xi_{N1} & . & . & \xi_{NN} & . & 0\end{bmatrix}\Phi Q$ and then absorb Φ into Q using Fact 10. With this structure, and using properties of unitary isotropically distributed Q, ξ_{11} would behave similar to a term for SISO with gDoF $\left(1-\frac{1}{T}\right)\gamma_{11}$, but other ξ_{i1} conditioned on the first row of the matrix would behave like a SISO with $T = 1$, which has zero gDoF. And rest of the ξ_{ij} would be dominated by noise, and hence, it too will have zero gDoF from Fact 9. More details are in Appendix C of [5]. ∎

Theorem 6. *The gDoF of the noncoherent MISO channel with $G = \begin{bmatrix} g_{11} & . & . & g_{1M}\end{bmatrix}$ is $\left(1-\frac{1}{T}\right)\max_i \gamma_{1i}$, i.e., the gDoF can be achieved by only using the statistically best transmit antenna.*

Proof: In this case Y is a column vector and $h(Y)$ can be evaluated using Lemma 11. Also, we prove that $h(Y|X) \dot{\geq} \mathbb{E}\left[\log\left(1+\sum_{i=1}^{M}\rho_{1i}^2\left\|\overline{X_i}\right\|^2\right)\right]$ using Linear Algebra techniques. With these two lemmas, the gDoF result follows. See Appendix D of [5] for details. ∎

Theorem 7. *For the 2×2 symmetric noncoherent MIMO with $G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22}\end{bmatrix}$, where $g_{11} \sim g_{22} \sim \mathcal{CN}(0, \mathsf{SNR}^{\gamma_{\mathsf{D}}})$ and $g_{12} \sim g_{21} \sim \mathcal{CN}(0, \mathsf{SNR}^{\gamma_{\mathsf{CL}}}), \gamma_D \geq \gamma_{CL}$ ($\gamma_D \geq \gamma_{CL}$ is without loss of generality) the gDoF is given in Table I, and can be achieved by $X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ \eta & c & 0 & . & . & 0\end{bmatrix} Q$, where $\eta \sim \mathcal{CN}\left(0, |b|^2\right)$ independent of the unitary isotropic Q,*

$|a|^2 = \mathsf{SNR}^{\gamma_a}, |b|^2 = \mathsf{SNR}^{\gamma_b}, |c|^2 = \mathsf{SNR}^{\gamma_c}$, *and the values of* $(\gamma_a, \gamma_b, \gamma_c)$ *are as shown in Table I.*

Regime	Solution $(\gamma_a, \gamma_b, \gamma_c)$	gDoF
$T = 2$	$(0, 0, -\gamma_{CL})$	$\gamma_D - \frac{1}{2}\gamma_{CL}$
$T \geq 3$	$(0, 0, 0)$	$2\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{1}{T}\gamma_{CL}\right)$

Table I
GDOF FOR SYMMETRIC 2×2 MIMO WITH
$\gamma_{11} = \gamma_{22} = \gamma_D \geq \gamma_{CL} = \gamma_{12} = \gamma_{21}$

Proof: The proof is in subsection IV-C. From Theorem 1, we have an optimal distribution of the form $X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q$, where Q is unitary isotropically distributed and independent of a, b, c. We first obtain an outerbound as a maximization of the expected value of a function $f\left(|a|^2, |b|^2, |c|^2\right)$. This is using Lemma 14 and Lemma 15 which help to convert the entropy terms $h()$ into expected values. Then in Lemma 16 we prove that the maximization of $\mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right]$ can be achieved with a single mass point of $\left(|a|^2, |b|^2, |c|^2\right)$ for optimal gDoF. Then the gDoF outerbound can be expressed as a piecewise linear optimization problem, which yields the solution as above. ■

Note that the above result shows that we need to use both antennas for achieving the gDoF for $T = 2$, since with only one antenna we can only achieve $\frac{1}{2}\gamma_D$ from Theorem 5. This is in contrast to the result for 2×2 MIMO with i.i.d links, where the optimal DoF could be achieved a using single transmit antenna for $T = 2$, and using both antennas was shown to be sub-optimal. For $T \geq 3$ for a 2×2 MIMO with all exponents γ_D, the DoF is $2\left(1 - \frac{2}{T}\right)\gamma_D$ compared to $2\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{1}{T}\gamma_{CL}\right)$ in our model with crosslink exponents γ_{CL}. Thus having weaker crosslinks gives a gDoF gain of $\frac{2}{T}(\gamma_D - \gamma_{CL})$.Also as $T \to \infty$ the gDoF achieved is $2\gamma_D$, which agrees with the gDoF result for coherent MIMO [1].

IV. Analysis and Proof Sketches

A. Mathematical Preliminaries

We first state some mathematical preliminaries required for the analysis.

Fact 8. *For an exponentially distributed random variable* ξ *with mean* μ_ξ ,$\log(a + b\mu_\xi) - \gamma_E \log(e) \leq \mathbb{E}[\log(a + b\xi)] \leq \log(a + b\mu_\xi)$ *where* γ_E *is Euler-Mascheroni constant. See [4] for proof.*

Fact 9. *For exponential* ξ

$$\mathbb{E}\left[\frac{b}{b + \xi}\right] = \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \Gamma\left(0, \frac{b}{\mu_\xi}\right) \leq \frac{b}{\mu_\xi} \ln\left(1 + \frac{\mu_\xi}{b}\right)$$

where $\Gamma(0, x)$ *is the incomplete gamma function. Note that* $0 \leq x \ln\left(1 + \frac{1}{x}\right) \leq 1$. *See [5] for proof.*

Fact 10. *Let* Q *be isotropic unitary distributed and* Φ *be a unitary matrix distributed according to any distribution independent of* Q, *then* $Q, \Phi Q, Q\Phi$ *all have the same distribution. Moreover* $\Phi Q, Q\Phi$ *are independent of* Φ. *See [3] for details.*

Lemma 11. *Let* $[\xi_1, \xi_2, \ldots, \xi_n]$ *be an arbitrary complex random vector and* Q *be an* $n \times n$ *isotropic unitary distributed matrix, independent of* ξ_i. *Then,* $h([\xi_1, \xi_2, \ldots, \xi_n] Q) = h\left(\sum |\xi_i|^2\right) + (n-1)\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log\left(\frac{\pi^n}{\Gamma(n)}\right)$

Proof: This is proved by using the fact that in radial coordinates, the distribution of $[\xi_1, \xi_2, \ldots, \xi_n] Q$ will depend **only** on the radius. See Appendix G of [5] for more details. ■

B. Properties of transmitted signals that achieve capacity

For any $T \times T$ unitary matrix Φ, we have $Y\Phi^\dagger = GX\Phi^\dagger + W\Phi^\dagger$. Since $w_{ij} \sim$ i.i.d $\mathcal{CN}(0, 1)$, $W\Phi^\dagger$ and W have the same distribution, and hence,

$$p\left(Y\Phi^\dagger | X\Phi^\dagger\right) = p(Y|X). \quad (2)$$

Now $C = \sup_{p(X)} I(X; Y)$ subject to the average power constraint (1), and we have

$$\begin{aligned} I(X; Y) &= \mathbb{E}\left[\log\left(\frac{p(Y|X)}{p(Y)}\right)\right] \\ &= \int dX p(X) \int dY p(Y|X) \\ &\quad \cdot \log\left(\frac{p(Y|X)}{\int d\tilde{X} p\left(\tilde{X}\right) p\left(Y|\tilde{X}\right)}\right) \end{aligned} \quad (3)$$

Lemma 12. *(Invariance of* $I(X; Y)$ *to post-rotations of* X*): Suppose that* X *has a probability density* $p_0(X)$*that generates some mutual information* I_0. *Then, for any unitary matrix* Φ, *the "post-rotated" probability density,* $p_1(X) = p_0\left(X\Phi^\dagger\right)$ *also generates* I_0.

Proof: This can be proved by substituting the post-rotated density $p_1(X)$ into Equation (3), changing the variables of integration, and using $p(Y\Phi|X'\Phi) = p(Y|X')$ from Equation (2). See Appendix B of [5] for details. ■

Lemma 13. *The signal of the form* $X = LQ$ *with* L *lower triangular and* Q *unitary and isotropically distributed (and independent of* L*) achieves the capacity.*

Proof: Let X be a capacity achieving random variable and I_0 be the corresponding mutual information achieved. Now X can be decomposed as $X = L\Phi'$ using the LQ decomposition with L lower triangular and Φ' unitary, but they could be jointly distributed and Φ' may not be unitary isotropically distributed. Let Θ be an isotropically distributed unitary matrix that is statistically independent of L and Φ'. Now use $X_1 = X\Theta$ for signaling and let Y be the corresponding received signal. We have

$$I(X_1; Y|\Theta) = I(X\Theta; Y|\Theta) = I_0$$

using Lemma 12. Now

$$\begin{aligned} I(X_1; Y) + I(\Theta; Y|X_1) &= I(\Theta; Y) + I(X_1; Y|\Theta) \\ I(X_1; Y) + 0 &\overset{(i)}{=} I(\Theta; Y) + I(X_1; Y|\Theta) \\ I(X_1; Y) &\overset{(ii)}{\geq} I(X_1; Y|\Theta) \\ &= I_0 \end{aligned}$$

where (i) follows since $I(\Theta; Y|X_1) = 0$ ($\Theta - X_1 - Y$ is a Markov chain) and (ii) holds since $I(X_1; Y|\Theta) \geq 0$. Hence, without loss of generality, the signal of the form $LQ = L\Phi'\Theta$ with $Q = \Phi'\Theta$ achieves the capacity. Now $Q = \Phi'\Theta$ is also isotropic and independent of Φ' using Fact 10. ■

Next, we focus our attention on computing $h(Y|X)$, which will be necessary in future derivations. Let $Y(n)$ be the n^{th} row of Y. Conditioned on X, the rows of Y are independent Gaussian. Hence:

$$h(Y|X) = \sum_{n=1}^{M} h(Y(n)|X). \tag{4}$$

With $\rho^2(n)$ being the vector of channel strengths to n^{th}receiver antenna, we have:

$$\begin{aligned} K_{Y(n)|X} &= \mathbb{E}\left[Q^\dagger L^\dagger g^\dagger(n) g(n) LQ \middle| LQ\right] + I_T \\ &= Q^\dagger L^\dagger \mathbb{E}\left[g^\dagger(n) g(n)\right] LQ + I_T \\ &= Q^\dagger L^\dagger \text{diag}\left(\rho^2(n)\right) LQ + I_T \end{aligned}$$

where I_T is a $T \times T$ identity matrix. Hence:

$$\begin{aligned} &h(Y(n)|X) \\ &\quad= \mathbb{E}\left[\log\left(\det\left(\pi e K_{Y(n)|X}\right)\right)\right] \\ &\quad= \mathbb{E}\left[\log\left(\det\left(\pi e\left(Q^\dagger L^\dagger \text{diag}\left(\rho^2(n)\right) LQ + I_T\right)\right)\right)\right] \\ &\quad\overset{(i)}{=} \mathbb{E}\left[\log\left(\det\left(\pi e\left(L^\dagger \text{diag}\left(\rho^2(n)\right) L + I_T\right)\right)\right)\right] \end{aligned}$$

where (i) uses the property of determinants to cancel Q and $Q^\dagger$. Also, for $T > M$, using the lower triangular structure of L with $L_{M\times M}$ being the first $M \times M$ submatrix of L, we have:

$$\begin{aligned} &h(Y(n)|X) \\ &= \mathbb{E}\left[\log\left(\det\left(\left(L_{M\times M}^\dagger \text{diag}\left(\rho^2(n)\right) L_{M\times M} + I_M\right)\right)\right)\right] \\ &\quad + (T)\log(\pi e) \end{aligned} \tag{5}$$

C. 2×2 MIMO system

Here we have the structure of the optimal distribution as $X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q$ and we have $G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$, $Y = GX + W$, where W is $2 \times T$ vector with i.i.d $\mathcal{CN}(0,1)$ components. We assume $T \geq 2$ since the case $T = 1$ is handled by Theorem 2.

1) Outer bound: We have:

$$\begin{aligned} &h(Y) \\ &= h\left(G \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q + W\right) \\ &\overset{(i)}{=} h\left(\left(G \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} + W\right) Q\right) \\ &= h\left(\begin{bmatrix} ag_{11} + bg_{12} + w_{11} & cg_{12} + w_{12} & w_{13} & . & w_{1T} \\ ag_{21} + bg_{22} + w_{21} & cg_{22} + w_{22} & w_{23} & . & w_{2T} \end{bmatrix} Q\right) \\ &\overset{(ii)}{=} h\left(\begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} \Phi Q\right) \\ &\overset{(iii)}{=} h\left(\begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} Q\right) \end{aligned}$$

where step (i) used the fact that W and WQ have the same distribution and is independent of Q; in step (ii), ξ_{ij} arises after LQ transformation (using Gram-Schmidt process) $\begin{bmatrix} ag_{11} + bg_{12} + w_{11} & cg_{12} + w_{12} & w_{13} & . & w_{1T} \\ ag_{21} + bg_{22} + w_{21} & cg_{22} + w_{22} & w_{23} & . & w_{2T} \end{bmatrix} = \begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} \Phi$ where Φ is unitary. In step (iii), we absorb Φ into Q using Fact 10. The Gram-Schmidt process for LQ transformation yields ξ_{ij} as given in Equation (6), Equation (7) and Equation (8).

$$\begin{aligned} &|\xi_{11}|^2 \\ &= |ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \end{aligned} \tag{6}$$

$$\begin{aligned} &|\xi_{21}|^2 \\ &= \left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)^{-1} \\ &\quad \cdot \Bigg|(ag_{21} + bg_{22} + w_{21})(ag_{11} + bg_{12} + w_{11})^* \\ &\qquad + (cg_{22} + w_{22})(cg_{12} + w_{12})^* + \sum_{i=3}^{T} w_{2i} w_{1i}^*\Bigg|^2 \end{aligned} \tag{7}$$

$$\begin{aligned} &|\xi_{22}|^2 \\ &= |ag_{21} + bg_{22} + w_{21}|^2 + |cg_{22} + w_{22}|^2 + \sum_{i=3}^{T} |w_{2i}|^2 \\ &\quad - |\xi_{21}|^2 \end{aligned} \tag{8}$$

Now using properties of Q and Lemma 11, it can be shown that

$$\begin{aligned} h(Y) \dot{\leq} h\left(|\xi_{21}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2\right) + h\left(|\xi_{22}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2\right) \\ + (T-2)\mathbb{E}\left[\log\left(|\xi_{22}|^2 |\xi_{11}|^2\right)\right] \end{aligned}$$

The details of the derivation are in [5].

Also using (5),(4) we get

$$\begin{aligned} &h(Y|X) \\ &\quad= \mathbb{E}\left[\log\left(\left(1 + |a|^2 \rho_{11}^2\right)\left(1 + |c|^2 \rho_{12}^2\right) + |b|^2 \rho_{12}^2\right)\right] \\ &\qquad + \mathbb{E}\left[\log\left(\left(1 + |a|^2 \rho_{21}^2\right)\left(1 + |c|^2 \rho_{22}^2\right) + |b|^2 \rho_{22}^2\right)\right] \\ &\qquad + 2T\log(\pi e) \end{aligned} \tag{9}$$

Lemma 14. *$h\left(|\xi_{22}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2\right)$ and $h\left(|\xi_{22}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2, a, b, c\right)$ have the same degrees of freedom and the latter is upperbounded by $\mathbb{E}\left[\log\left(e\mathbb{E}\left[|\xi_{22}|^2 |\xi_{11}|^2 \middle| a, b, c\right]\right)\right]$.*

Proof: See Appendix K of [5]. ■

Lemma 15. *$h\left(|\xi_{21}|^2|\xi_{11}|^2 \middle| |\xi_{11}|^2\right)$ and $h\left(|\xi_{21}|^2|\xi_{11}|^2 \middle| |\xi_{11}|^2, a, b, c\right)$ have the same degrees of freedom and the latter is upperbounded by $\mathbb{E}\left[\log\left(\mathbb{E}\left[|\xi_{21}|^2|\xi_{11}|^2 \middle| a, b, c\right]\right)\right]$.*

Proof: This can be proved similar to the previous lemma. ■

Using these lemmas we can arrive at the bound on $I(X;Y)$ given in Equation (10). The intermediate steps of computation can be found in [5].

$$I(X;Y) \dot{\leq} \mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right] \tag{10}$$

where $f\left(|a|^2, |b|^2, |c|^2\right)$ is given in Equation (11)

$$\begin{aligned} & f\left(|a|^2, |b|^2, |c|^2\right) \\ & = \log\left(\left(|a|^2\rho_{11}^2 + |b|^2\rho_{12}^2 + 1\right)\left(|a|^2\rho_{21}^2 + |b|^2\rho_{22}^2 + 1\right)\right. \\ & \quad \left. + \left(|c|^2\rho_{12}^2 + 1\right)\left(|c|^2\rho_{22}^2 + 1\right)\right) \\ & + (T-1)\log\left(\left(|a|^2\rho_{11}^2 + 1\right)\left(|c|^2\rho_{22}^2 + 1\right)\right. \\ & \quad \left. + |b|^2\left(\rho_{12}^2 + \rho_{22}^2\right) + \left(|a|^2\rho_{21}^2 + 1\right)\left(|c|^2\rho_{12}^2 + 1\right)\right) \\ & - \log\left(\left(1 + |a|^2\rho_{11}^2\right)\left(1 + |c|^2\rho_{12}^2\right) + |b|^2\rho_{12}^2\right) \\ & - \log\left(\left(1 + |a|^2\rho_{21}^2\right)\left(1 + |c|^2\rho_{22}^2\right) + |b|^2\rho_{22}^2\right) \end{aligned} \tag{11}$$

Hence, an outerbound is given by the optimization problem

$$\mathcal{P}_1 : \begin{cases} \underset{\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right] \leq T}{\text{maximize}} \mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right] \quad \text{with} \\ \text{Support}\left(|a|^2, |b|^2, |c|^2\right) = \left(\mathbb{R}^+\right)^3 \end{cases} \tag{12}$$

Lemma 16. *The degrees of freedom achieved in $\mathcal{P}_1$ can be achieved by a point mass distribution, i.e., gDoF$(\mathcal{P}_1)$ = gDoF$(\mathcal{P}_7)$ where $\mathcal{P}_7$ is the following*

$$\mathcal{P}_7 : \begin{cases} \text{maximize} f\left(|a|^2, |b|^2, |c|^2\right) \quad \text{with} \\ |a|^2 \leq T, |b|^2 \leq T, |c|^2 \leq T. \end{cases} \tag{13}$$

Proof: The proof proceeds in several steps:

Step 1: Show that there exists a discretization (over an infinite set) for any distribution of $\left(|a|^2, |b|^2, |c|^2\right)$ that does not incur a loss in gDoF

Step 2: Show that the discretization can be limited to a finite set without incurring a loss in gDoF

Step 3: View the problem as a linear program with 2 constraints, and show that the there is an optimal distribution with just 2 mass points

Step 4: Show that the 2 mass points can be collapsed to a single point using arguments of symmetry

The details are given in Appendix L of [5]. ■

Changing the variables from $\left(|a|^2, |b|^2, |c|^2\right)$ to $(\gamma_a, \gamma_b, \gamma_c)$ with the substitution $|a|^2 = \mathsf{SNR}^{\gamma_a}, |b|^2 = \mathsf{SNR}^{\gamma_b}, |c|^2 = \mathsf{SNR}^{\gamma_c}$, we have

$$\text{gDoF}(\mathcal{P}_1) = \text{gDoF}(\mathcal{P}_7) = (\mathcal{P}_8)$$

where $\mathcal{P}_8$ is the following:

$$\mathcal{P}_8 : \begin{cases} \text{maximize} f_\gamma(\gamma_a, \gamma_b, \gamma_c) \quad \text{with} \\ \gamma_a \leq 0, \gamma_b \leq 0, \gamma_c \leq 0, \end{cases} \tag{14}$$

with $f_\gamma(\gamma_a, \gamma_b, \gamma_c)$ given below in Equation (15).

$$\begin{aligned} & f_\gamma(\gamma_a, \gamma_b, \gamma_c) \\ & = \max\left(\max(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12})^+\right. \\ & \quad \left. + \max(\gamma_a + \gamma_{21}, \gamma_b + \gamma_{22})^+, (\gamma_c + \gamma_{12})^+ + (\gamma_c + \gamma_{22})^+\right) \\ & + (T-1)\max\left((\gamma_a + \gamma_{11})^+ + (\gamma_c + \gamma_{22})^+\right. \\ & \quad \left., \gamma_b + \max(\gamma_{12}, \gamma_{22}), (\gamma_a + \gamma_{21})^+ + (\gamma_c + \gamma_{12})^+\right) \\ & - \max(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}, \gamma_c + \gamma_{12}, \gamma_a + \gamma_c + \gamma_{11} + \gamma_{12})^+ \\ & - \max(\gamma_a + \gamma_{21}, \gamma_b + \gamma_{22}, \gamma_c + \gamma_{22}, \gamma_a + \gamma_c + \gamma_{21} + \gamma_{22})^+. \end{aligned} \tag{15}$$

Now, for the symmetric 2×2 MIMO with $\gamma_{11} = \gamma_{22} = \gamma_D \geq \gamma_{CL} = \gamma_{12} = \gamma_{21}$ the above piecewise-linear problem can be solved using linear programming techniques and using a computer software (like Mathematica) to go through all cornerpoints. More details are in Appendix M of [5]. The solution obtained is given in Table I. The gDoF per symbol is obtained as $\frac{1}{T} \cdot (\mathcal{P}_8)$.

2) Inner bound: For the innerbound, choosing $X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ \eta & c & 0 & . & . & 0 \end{bmatrix} Q$, $\eta \sim \mathcal{CN}\left(0, |b|^2\right)$ (independent of the unitary isotropic Q), with $|a|^2 = \mathsf{SNR}^{\gamma_a}, |b|^2 = \mathsf{SNR}^{\gamma_b}, |c|^2 = \mathsf{SNR}^{\gamma_c}$, and the values of $(\gamma_a, \gamma_b, \gamma_c)$ taken from the Table I, it can be shown the outer bound can be achieved. The calculations are given in Appendix E of [5].

References

[1] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim. Gaussian half-duplex relay networks: Improved constant gap and connections with the assignment problem. *IEEE Transactions on Information Theory*, 60(6):3559–3575, Jun 2014.

[2] Amos Lapidoth and Stefan M Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *Information Theory, IEEE Transactions on*, 49(10):2426–2467, 2003.

[3] Thomas L Marzetta and Bertrand M Hochwald. Capacity of a mobile multiple-antenna communication link in rayleigh flat fading. *Information Theory, IEEE Transactions on*, 45(1):139–157, 1999.

[4] J. Sebastian, C. Karakus, S. N. Diggavi, and I-H. Wang. Rate splitting is approximately optimal for fading gaussian interference channels. *Allerton Conference on Communication, Control and Computing*, October 2015.

[5] J. Sebastian, A. Sengupta, and S. N. Diggavi. On capacity of noncoherent mimo with asymmetric link strengths. *arXiv*, 2017.

[6] Lizhong Zheng and D. N. C. Tse. Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Transactions on Information Theory*, 48(2):359–383, Feb 2002.