Petre Stoica, Yi Jiang,
and Jian Li

# On MIMO Channel Capacity: An Intuitive Discussion

Consider a multiple-input, multiple-output (MIMO) channel with $M_t$ transmit and $M_r$ receive antennas. The received sampled baseband signal can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{1}$$

where $\mathbf{H}$ is the $M_r \times M_t$ matrix whose $(i, j)$th element denotes the complex channel gain between the $j$th transmit and the $i$th receive antennas. We assume that $\mathbf{z}$ is a zero-mean circularly symmetric complex Gaussian noise with covariance matrix $\mathbf{R}_z$. Let $\mathbf{R}_x$ denote the covariance matrix of the input vector $\mathbf{x}$, and let $\mathbf{R}_y$ be the covariance matrix of the output vector $\mathbf{y}$. Assuming that $\mathbf{x}$ and $\mathbf{z}$ are uncorrelated with one another, we have that

$$\mathbf{R}_y \triangleq E[\mathbf{y}\mathbf{y}^*] = \mathbf{H}\mathbf{R}_x\mathbf{H}^* + \mathbf{R}_z, \tag{2}$$

$$\mathbf{R}_{xy} \triangleq E[\mathbf{x}\mathbf{y}^*] = \mathbf{R}_x\mathbf{H}^*. \tag{3}$$

We assume that $\mathbf{H}$, $\mathbf{R}_x$, and $\mathbf{R}_z$ are known to the receiver and, hence, so are $\mathbf{R}_y$ and $\mathbf{R}_{xy}$. Under the above assumptions, the channel capacity is given by (see, e.g., [1], [2])

$$C = \log_2 \frac{|\mathbf{R}_z + \mathbf{H}\mathbf{R}_x\mathbf{H}^*|}{|\mathbf{R}_z|} \quad \text{bits per channel use.} \tag{4}$$

The covariance matrix $\mathbf{R}_x$ in (4) can take on different values depending on the channel state information available to the transmitter (e.g., $\mathbf{R}_x \sim \mathbf{I}$ when the transmitter has no channel information). In what follows, we let $\mathbf{R}_x$ denote any of these possible values of the input covariance matrix, both for the sake of generality and because doing so does not introduce any additional difficulty.

The usual derivation of the capacity formula in (4) relies on the maximization, with respect to the density of $\mathbf{x}$, of the mutual information between $\mathbf{y}$ and $\mathbf{x}$ (which is equal to the difference between the differential entropies of $\mathbf{y}$ and $\mathbf{z}$) [3], [1]. This derivation might not be too enlightening for a new student and as such it might obscure the meaning of the notion of capacity. In this note, we relate the capacity $C$ to the covariance matrix of the linear minimum mean-squared-error (LMMSE) estimate of $\mathbf{x}$. As we will see, doing so provides a clear interpretation of the concept of capacity, as well as of some optimal transceiver design schemes that aim to achieve the channel capacity.

As is well known, the LMMSE estimate of $\mathbf{x}$ and its covariance matrix are given by

$$\hat{\mathbf{x}} = \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{y}, \tag{5}$$

and

$$\begin{aligned}\mathbf{R}_{\text{LMMSE}} &\triangleq E[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^*] \\ &= \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^* \\ &\quad - 2\mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^* + \mathbf{R}_x \\ &= \mathbf{R}_x - \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^*. \end{aligned} \tag{6}$$

It follows from (2), (3), and the matrix inversion lemma (see, e.g., [4], [5]) that

$$\begin{aligned}\mathbf{R}_{\text{LMMSE}}^{-1} &= \mathbf{R}_x^{-1} + \mathbf{R}_x^{-1}\mathbf{R}_{xy} \\ &\quad \times (\mathbf{R}_y - \mathbf{R}_{xy}^*\mathbf{R}_x^{-1}\mathbf{R}_{xy})^{-1} \\ &\quad \times \mathbf{R}_{xy}^*\mathbf{R}_x^{-1} \end{aligned} \tag{7}$$

$$\begin{aligned}&= \mathbf{R}_x^{-1} + \mathbf{H}^* \\ &\quad \times (\mathbf{R}_y - \mathbf{H}\mathbf{R}_x\mathbf{H}^*)^{-1}\mathbf{H} \end{aligned} \tag{8}$$

$$= \mathbf{R}_x^{-1} + \mathbf{H}^*\mathbf{R}_z^{-1}\mathbf{H}. \tag{9}$$

On the other hand, from (4) and a well-known determinantal identity (see, e.g., [4], [5]), we have that

$$\begin{aligned}C &= \log_2 |\mathbf{I} + \mathbf{R}_z^{-1}\mathbf{H}\mathbf{R}_x\mathbf{H}^*| \\ &= \log_2 |\mathbf{I} + \mathbf{H}^*\mathbf{R}_z^{-1}\mathbf{H}\mathbf{R}_x| \end{aligned} \tag{10}$$

$$\begin{aligned} &= \log_2 |\mathbf{R}_x| + \log_2 \\ &\quad |\mathbf{R}_x^{-1} + \mathbf{H}^*\mathbf{R}_z^{-1}\mathbf{H}|. \end{aligned} \tag{11}$$

It follows from (9) and (11) that the channel capacity can be rewritten as

$$C = \log_2 \left( \frac{|\mathbf{R}_x|}{|\mathbf{R}_{\text{LMMSE}}|} \right). \tag{12}$$

The above expression relates, in a simple manner, the channel capacity to the covariance matrix of the LMMSE estimate of $\mathbf{x}$.

## DISCUSSION

The channel capacity formula (12) has an intuitive appeal. We can envisage the LMMSE estimate $\hat{\mathbf{x}}$ as lying (with high probability) in a "small cell" centered around the codeword $\mathbf{x}$. The volume of the cell is proportional to $|\mathbf{R}_{\text{LMMSE}}|$. The volume of the codebook space (in which $\mathbf{x}$ lies with a high probability) is proportional to $|\mathbf{R}_x|$. The ratio $\rho = (|\mathbf{R}_x|)/(|\mathbf{R}_{\text{LMMSE}}|)$ gives the number of cells that can be packed into the codebook space without significant overlapping. The "center" of each such cell, the codeword, can be reliably detected, for instance, using $\hat{\mathbf{x}}$. The conclusion is that one can communicate reliably using a codebook of size $\rho$, which contains $\log_2 \rho$ information bits; this observartion provides an intuitive motivation to the capacity formula in (4) or, equivalently, (12). A similar


1053-5888/05/$20.00©2005IEEE

argument for the single-input, single-output (SISO) channel case can be found in [3, Ch. 10.1].

In general, the matrix $\mathbf{R}_{\text{LMMSE}}$ is not diagonal, i.e., the LMMSE estimates of the elements of $\mathbf{x}$ are correlated. These correlations clearly contain useful information for subsequent decoding procedures. However, in practice, one often encodes and decodes the elements of $\mathbf{x}$ separately and, hence, ignores the aforementioned correlations. This leads to a loss of information. With (12) in mind, we can quantify the capacity loss that results as

$$C_{\text{loss}} = \sum_{k=1}^{M_t} \log_2 [\mathbf{R}_{\text{LMMSE}}]_{kk} - \log_2 |\mathbf{R}_{\text{LMMSE}}|, \quad (13)$$

where $[\mathbf{R}_{\text{LMMSE}}]_{kk}$ denotes the $k$th diagonal element of $\mathbf{R}_{\text{LMMSE}}$. According to the Hadamard inequality [4], for any $K \times K$ positive semidefinite matrix $\mathbf{M}$

$$|\mathbf{M}| \leq \prod_{k=1}^{K} M_{kk}, \quad (14)$$

where the equality holds if and only if $\mathbf{M}$ is diagonal. Hence $C_{\text{loss}} \geq 0$, and there is no capacity loss if and only if $\mathbf{R}_{\text{LMMSE}}$ is diagonal.

Based on the above discussion, we see that: 1) for general MIMO communication channels, the use of the LMMSE estimator followed by separate substream decoding is not a capacity-wise optimal scheme; and 2) if the channel matrix $\mathbf{H}$ is such that $\mathbf{R}_{\text{LMMSE}}$ is a diagonal matrix, then the use of the LMMSE estimator may be the first step of decoupled lossless information processing. If channel state information is available at the transmitter, then the transmitter can use a precoder $\mathbf{P}$ to obtain a virtual channel matrix

$$\mathbf{H}_{\text{vt}} = \mathbf{HP}, \quad (15)$$

which is such that the corresponding $\mathbf{R}_{\text{LMMSE}}$ is diagonal. This explains why all existing optimal linear transceiver designs invariably lead to the diagonalization of the matrix $\mathbf{H}^*\mathbf{H}$ (see e.g., [6], [7] and the references therein). Indeed, if $\mathbf{R}_x$ is diagonal and $\mathbf{R}_z \sim \mathbf{I}$ (as is usually assumed), then it follows from (9) that $\mathbf{H}_{\text{vt}}$ must have orthogonal columns for $\mathbf{R}_{\text{LMMSE}}$ to be diagonal. Then, a precoder $\mathbf{P}$ made from the right singular vectors of $\mathbf{H}$ is the optimal solution. Such a scheme, however, lacks flexibility since the singular value decomposition is unique. If nonlinear processing is allowed, such as the decision feedback equalizer, one can design lossless information transceivers in a more flexible manner, as demonstrated in [8].

## ACKNOWLEGMENTS

This work was supported in part by the National Science Foundation Grant CCR-0104887, the Swedish Science Council (VR), and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT).

## AUTHORS

***Petre Stoica*** is with the Department of Systems and Control, Uppsala University, Uppsala, Sweden. ***Yi Jiang*** and ***Jian Li*** are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA.

## REFERENCES

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

[2] B. Holter, "On the capacity of the MIMO channel: A tutorial introduction," in *Proc. IEEE Norwegian Symp. Signal Processing*, Trondheim, Norway, Oct. 2001, pp. 167–172.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[4] R.A. Horn and C.R. Johnson, *Matrix Analysis*. Cambridge: Cambridge Univ. Press, 1985.

[5] P. Stoica and R.L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[6] A. Scaglione, P. Stoica, S. Barbarossa, G.B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, pp. 1051–1064, May 2002.

[7] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 2381–2401, Sept. 2003.

[8] Y. Jiang and J. Li, "Uniform channel decomposition for MIMO communications," in *Proc. 38th Asilomar Conf. Signals, Systems, Computers*, Asilomar, CA, Nov. 7–10, 2004, pp. 856–860. [SP]