Reflections on the Capacity Region of the Multi-Antenna Broadcast Channel

Hanan Weingarten   Yossef Steinberg   Shlomo Shamai (Shitz)

Dep. of Elec. Eng.   Dep. of Elec. Eng.   Dep. of Elec. Eng.

Technion-IIT   Technion-IIT   Technion-IIT

Haifa, Israel 32000,   Haifa, Israel 32000,   Haifa, Israel 32000,

whanan@tx.technion.ac.il   ysteinbe@ee.technion.ac.il   sshlomo@ee.technion.ac.il

Abstract—**We give an overview of the results and techniques used to determine the capacity region of the Multi-Antenna Broadcast Channel [18]. We provide a brief historical perspective on this problem, and finally, indicate recent developments.**

I. INTRODUCTION

The problem of the Gaussian multi-antenna broadcast channel (BC) has been in the limelight of information theoretic research in recent years and turned out to be most relevant in modern communication systems. This short overview is intended to provide a taste of this exiting field and mainly to highlight, the results and techniques, reported in [18], where the capacity region of this channel has been established (This paper is the winner of the 2007 IT Society Best Paper Award). Further, a short historical perspective as well as some recent developments are also mentioned, by pointing out relevant references.

We examine a BC that has one transmitter which sends several independent messages to several users, numbered $k = 1, \ldots, K$, and equipped with receivers which cannot cooperate with one another. Each message is intended for a different user. Therefore, such a channel is used to model the downlink channel in cellular systems, other wireless links and ADSL wire-line links. Independently, there has been a massive interest in multi-antenna channels, mainly in the context of a point to point channel. In such a channel both the transmitter and receiver are equipped with several antennas, thereby obtaining almost a linear increase in the channel capacity, proportional to the minimum number of transmit and receive antennas, without requiring additional bandwidth or transmit power [12]. It is no wonder then that there has been interest in the combination of these field of research, the multi-antenna BC.

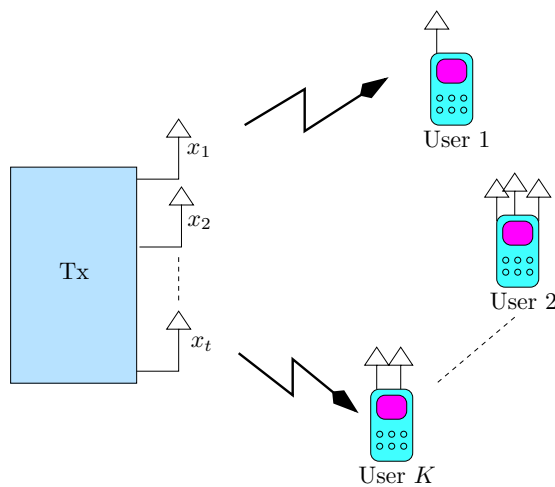In a Gaussian multi-antenna BC, the transmitter is

Fig. 1. A multi-antenna broadcast channel

equipped with t antennas and each of the receivers is equipped with r_i antennas (see Figure 1). Due to a multitude of reflections in wireless links and wire-line coupling in ADSL links, the receivers obtain some linear mixture of the transmitted signals. By using a vector notation to denote the transmitted and received signals at any given time, we can use the following vector equation to define a time sample of the BC

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \qquad k = 1, \ldots, K, \qquad (1)$$

where $\mathbf{y}_k$ is a $r_i \times 1$ vector received by user k, $\mathbf{H}_k$ is an $r_i \times t$ fading matrix, defining the linear mixture, and $\mathbf{x}$ is a $t \times 1$ vector which elements denote the levels being transmitted. $\mathbf{n}_k \sim \mathcal{N}(0, \mathbf{I})$ is a Gaussian noise vector added at the receiver of user k. In addition we assume that there is a power constraint, P, on the input such that $E[\mathbf{x}^T \mathbf{x}] < P$.

We can consider several encoding methods for the

multi-antenna BC. Linear encoding schemes are the simplest to implement yet show the great appeal of using multi-antenna transmitters in a BC setting. As an example, we consider the simple case of two users, each equipped with one antenna and a transmitter equipped with two antennas. The transmitter constructs the following signal:

$$\mathbf{x} = s_1\mathbf{u}_1 + s_2\mathbf{u}_2 \qquad (2)$$

where s_1 and s_2 are scalars carrying independent information streams for the first and second users, respectively, and $\mathbf{u}_1$ and $\mathbf{u}_2$ are 2×1 unitary vectors which form beams which direct the information streams at the direction of the respective users.
At each user we receive the following signal

$$y_1 = \mathbf{h}_1^T\mathbf{x} + n_1 = s_1\mathbf{h}_1^T\mathbf{u}_1 + z_1$$
$$y_2 = \mathbf{h}_2^T\mathbf{x} + n_2 = s_2\mathbf{h}_2^T\mathbf{u}_2 + z_2$$

where $z_1 = s_2\mathbf{h}_1^T\mathbf{u}_2 + n_1$ and $z_2 = s_1\mathbf{h}_2^T\mathbf{u}_1 + n_2$ are the overall interference at each receiver. When zero-forcing beam-forming is considered, the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are chosen in such a manner that $\mathbf{h}_1^T\mathbf{u}_2 = \mathbf{h}_2^T\mathbf{u}_1 = 0$. Thus, the signal intended to one user does not interfere at the receiver of the second user. Alternatively, $\mathbf{u}_1$ and $\mathbf{u}_2$ may be chosen to maximize the signal to interference ratio at the receivers (see, for example, [2], [20]).
Figure 2 is an eye opener and shows the overall transmission rate (sum-rate) that can be obtained when using the beam-forming construction or when using time domain multiple access (TDMA) (i.e. transmitting only to one user at a time). It can be seen that as the SNR increases, the slope (alternatively, multiplexing gain) using beam-forming is twice that obtained when the transmitter directs its signal only to the strongest user. Thus, it is possible to approximately double the communication rates at high SNRs when the transmitter has two antennas. Note that unlike the point to point channel, we did not require that each of the users will be equipped with two antennas.

II. TOWARDS THE CAPACITY REGION

We are now left with the question of what is the best performance which can be obtained in multi-antenna BC, or alternatively, what is the capacity region of this channel. Unlike the point to point channel, we are interested in a capacity region and not a scalar.
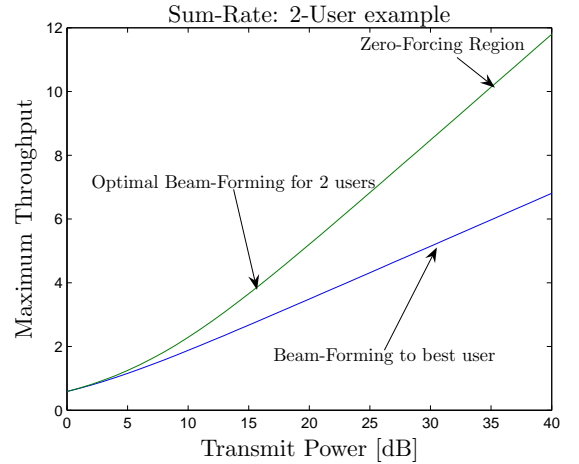

Fig. 2. Linear beam-forming and TDMA in a 2 user channel: $\mathbf{h}_1^T = (1 \ \ 0.5)$, $\mathbf{h}_2^T = (0.2 \ \ 1)$. At high SNRs, zero-forcing beam-forming is best. At mid-range SNRs, the beam-forming vectors are optimize to maximize the sum-rate.

In a BC we have several rates, one for each receiving user. We collect these rates onto a vector and the capacity region is defined as the set of all rate vectors which are achievable with an arbitrarily small probability of error for sufficiently large codeword lengths. The first to address this question for the multi antenna BC were Caire and Shamai [4]. In their paper, they suggested using a non-linear coding scheme known as dirty paper coding (DPC). This coding scheme is based on a result by Costa [5] who investigated a scalar, point to point Gaussian channel, which apart from additive noise also suffers from a Gaussian interference which is known non-causally at the transmitter but not at the receiver. A time sample of this channel is defined by $y = x + s + n$ where x is a power constrained channel input ($Exx < P$), $s \sim \mathcal{N}(0, S)$ is a gaussian interference which is non-causally known at the transmitter but not at the receiver and $n \sim \mathcal{N}(0, N)$ is an additive noise which is not known at either end.
Costa showed that the capacity of this channel is given by $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$. That is, we can obtain the same rate as if the interference, s, did not exist. This is not trivial as the input is power constrained and the interference can not be directly cancelled out. Costa referred to coding in such a scenario as writing on dirty paper, as the interference which is non-causally known at the transmitter may be compared to an ini-

WEINGARTEN, STEINBERG, AND SHAMAI: REFLECTIONS ON THE CAP. REG. OF THE MULTI-ANT. BC

tially dirty paper over which we may wish to deliver a message.

Caire and Shamai suggested using this scheme for transmitting over a BC. To demonstrate this, let us return to the two user example and the linear encoding scheme as in (2). Now assume that we use a standard Gaussian codebook for encoding the signal s_1. Note that when encoding the signal s_2, the transmitter is fully aware of the interference signal $s_1 \mathbf{h}_2^T \mathbf{u}_1$ and may treat it as non-causally known interference at the receiver y_2. Thus, using DPC as costa did, the effect of the interference may be completely eliminated when deciphering s_2 at user 2. On the other hand, user 1 still suffers from interference due to the signal s_2. Thus, the following rates can be obtained:

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{S_1 |\mathbf{h}_1^T \mathbf{u}_1|^2}{1 + S_2 |\mathbf{h}_1^T \mathbf{u}_2|^2} \right)$$
$$R_2 \leq \frac{1}{2} \log \left(1 + S_2 |\mathbf{h}_2^T \mathbf{u}_2|^2 \right)$$
(3)

where $S_1 + S_2 \leq P$ are the powers allotted for the transmission of the two signals. Note that the encoding order may also be reversed. DPC may be used to encode the signal s_1 on top of s_2. The resulting rates will be the same as in the equation above, except the indexes of the users will be reversed.

Figure 3 shows the achievable rate regions using TDMA, linear precoding and DPC with optimized beamforming vectors and power allocations. Evidently, DPC performs bettter than the other options and a natural question is whether the DPC region is the capacity region. It should be noted that unlike the point to point channel, there is no known single letter information theoretic expression for the capacity region for the general BC. See [6] for a short overview. Yet, there are some classes of the BC for which we do have an expression for the capacity region. One such class is the class of stochastically degraded BCs [7]. In this class, the output of one user is in essence the output of the other user after it has been passed through an additional point to point channel. One BC which belongs to this class is the Gaussian BC with a single transmit antenna. In this case, the output of one user is equal to a linear combination of the output of the other user and an independent Gaussian noise. The multi-antenna BC is not degraded in general and that poses a major hardship in characterizing its capacity region, as even a general information theoretic characterization is unavailable, let alone the optimized variables therein.
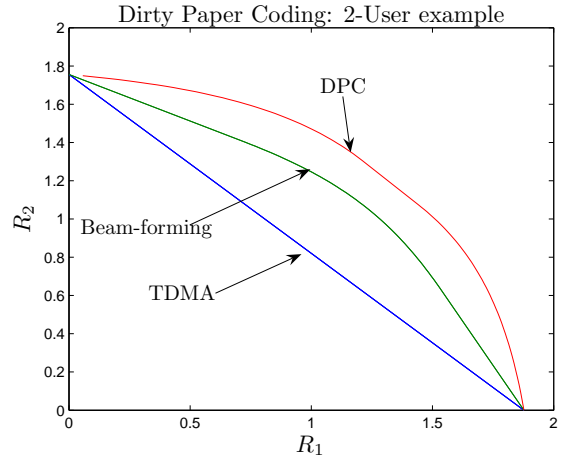

Fig. 3. TDMA, Linear beam-forming, and DPC regions in a 2 user channel: $\mathbf{h}_1^T = (1 \;\; 0.5)$, $\mathbf{h}_2^T = (0.2 \;\; 1)$ and $P = 10$.

Nevertheless, Caire and Shamai [4] were able to show, using the Sato outer bound that in a two user BC with receivers that have a single antenna, the sum-rate points on the boundary of the DPC region coincide with the capacity region. These points are also known as sum-capacity points. That is, the points that lie on the line $R_1 + R_2 = \max(R_1 + R_2)$. This result was later extended to any multi-antenna BC in [16], [14], [21], introducing interesting concepts of duality between Gaussian multi-antenna broadcast and multiple-access channels.

III. THE ENHANCED CHANNEL AND THE CAPACITY REGION

In our work [18], we have shown that all points on the boundary of the DPC region correspond to the capacity region. To simplify the problem, we consider a vector version of the channel where the number of receive antennas at each user is equal to the number of transmit antennas and the fading matrices are identity matrices. However, we now allow the noise covariance matrices to take any form. Therefore, a time sample of the channel takes the following form

$$\mathbf{y}_i = \mathbf{x} + \mathbf{n}_i, \quad i = 1, 2$$
(4)

where $\mathbf{n}_i \sim \mathcal{N}(0, \mathbf{N}_i)$ ($\mathbf{N}_i$ is a positive semi-definite -PSD- matrix). It is not difficult to see that if indeed all matrices $\mathbf{H}_i$ in (1) are square and invertible, the BCs defined by (4) and (1) are equivalent

if $\mathbf{N}_i = \mathbf{H}_i^{-1}\mathbf{H}_i^{-T}$. If the fading matrices are not invertible, it is possible to show that the BC in (1) may be approximated by a sequence of vector BCs (4) in which some of the eigenvalues of $\mathbf{N}_i$ go to infinity [18]. An additional restriction is put on the power constraint. Instead of a total power constraint where we limit the total power transmitted over all antennas, we consider a covariance matrix constraint such that $E\mathbf{x}\mathbf{x}^T \leq \mathbf{S}$ ($\mathbf{S}$ is a PSD matrix). The capacity region under such a power constra~~int may be gener-~~ alized to a broad range of power constraints such as the total power constraint.

In order to transmit over this BC we consider a vector version of DPC [22]. We transmit a superposition of two signals $\mathbf{x} = \mathbf{s}_1 + \mathbf{s}_2$ where $\mathbf{s}_1$ is a codeword taken from a Gaussian codebook and $\mathbf{s}_2$ is a DPC codeword, where $\mathbf{s}_1$ acts as a non-causally known interference. Thus, we obtain the vector version of (3) as follows:

$$R_1 \leq \frac{1}{2}\log\frac{|\mathbf{S}+\mathbf{N}_1|}{|\mathbf{S}_2+\mathbf{N}_1|}$$
$$R_2 \leq \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{N}_2|}{|\mathbf{N}_2|} \tag{5}$$

where $E\mathbf{s}_1\mathbf{s}_1^T = \mathbf{S} - \mathbf{S}_2$ and $E\mathbf{s}_2\mathbf{s}_2^T = \mathbf{S}_2$ are covariance matrices allotted for the messages. We can also switch between the users, reversing the precoding order, thus obtaining the same rates with the user indexes reversed.

Note that if $\mathbf{N}_2 \leq \mathbf{N}_1$ (i.e. $\mathbf{N}_1 - \mathbf{N}_2$ is a PSD), the vector BC is a degraded BC as the output of the first user may be given by $\mathbf{y}_1 = \mathbf{y}_2 + \tilde{\mathbf{n}}$ where $\tilde{\mathbf{n}} \sim \mathcal{N}(0, \mathbf{N}_1 - \mathbf{N}_2)$. Yet, even though there is an information theoretic expression for the capacity region of a degraded BC, proving that (5) gives the capacity region in the vector case is an elusive task (see [13], [15]). The difficulty arises as the central tool, the entropy power inequality as used by Bergmans' in the Gaussian scalar broadcast channel [1], is not tight for Gaussian vectors, unless they posses proportional covariance matrices.

We circumvent this hurdle by the introduction of an *enhanced channel*. A vector BC with noise covariances $\mathbf{N}_2' \leq \mathbf{N}_2$ and $\mathbf{N}_1' \leq \mathbf{N}_1$ is said to be an enhanced version of the vector BC in (4). We showed [18] that for a point on the boundary of the DPC region of a degraded vector BC, denoted by (R_1^*, R_2^*) and obtained by choosing $\mathbf{S}_2 = \mathbf{S}_2^*$, there exists an enhanced and degraded version of the vector BC such that the same rates are obtained in the enhanced
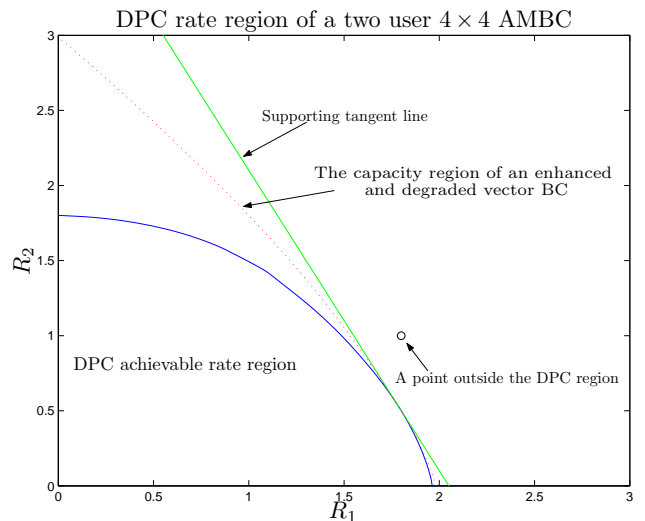

DPC rate region of a two user 4×4 AMBC

Fig. 4. Schematic view of the capacity region proof.

BC using the same power allocation. Furthermore, $\mathbf{S}_2 + \mathbf{N}_2'$ is proportional to $\mathbf{S}_2 + \mathbf{N}_1'$. This proportionality result allowed us to show that indeed (R_1^*, R_2^*) lies on the capacity region of the enhanced vector BC using Bergmans' classical approach [1]. Note that as the capacity region of the enhanced channel must contain that of the original channel (the enhanced channel has less noise), (R_1^*, R_2^*) must also lie on the capacity region of the original vector BC. Furthermore, as this can be done for every point on the boundary of the region in (5), this region must be the capacity region.

We generalized the proof to non-degraded vector BCs by finding a set of enhanced channels for this more general set of BCs. Note that the DPC region must always be convex as time sharing may be used. Therefore, any point that lies outside this region may be separated from the region by a tangent line. We showed [18] that for every tangent line we can find a specific enhanced and degraded vector BC which capacity region is also supported by the same tangent line (see Figure 4). Therefore, we conclude that every point outside the DPC region also lies outside the capacity region of an enhanced and degraded BC. However, as the capacity region of the enhanced channel contains that of the original channel, we conclude that every point outside the DPC region must lie outside the capacity region of the original channel and thus prove that the DPC region (5) is indeed

the capacity region. In [18] this result is extended to any multi-antenna BC (1) and any linear power constraint on the input.

IV. Future Work

Up to this point we assumed that each message has only one intended user. However, we may also consider the case where some of the messages are common to a number of users. The capacity region in this case is still an open problem and some progress has been reported in [17], [19]. Another open problem is the capacity region of fading and compound Gaussian BCs. We assumed that the fading matrix is perfectly known at the transmitter (and receivers). However, in reality, the transmitter may have only a partial knowledge of the fading conditions, giving rise to a fading or a compound BC [8], [9].

Applications and implications of the multiple-antenna broadcast channel in wireless and wire-line communications are reported extensively in recent years. The impact of theoretically motivated state of the art communications technologies for this channel can be substantial as is demonstrated in certain cellular models [10]. For an extended perspective and details see [3], [11].

References

[1] Patrick P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise," *IEEE Trans. on Information Theory*, pp. 279–280, Mar. 1974.

[2] H. Boche, M. Schubert, and E.A. Jorswieck, "Throughput maximization for the multiuser mimo broadcast channel," in *proceedings of Acoustics, Speech abd Signal Processing (ICASSP03)*, April 5–10 2003, pp. 4.808–4.811.

[3] G. Caire, S. Shamai, Y. Steinberg, and H. Weingarten, *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, chapter On Information Theoretic Aspects of MIMO-Broadcast Channels, Cambridge University Press, Cambridge, UK, 2006.

[4] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[5] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, pp. 439–441, May 1983.

[6] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2524 – 2530, Sept. 1998.

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

[8] S. Shamai (Shitz) H. Weingarten and G. Kramer, "On the compound mimo broadcast channel," in *Proceedings of the Information Theory and Applications Workshop (ITA 2007)*,, UCSD, San Diego, CA, JJan. 29 - Feb. 2, 2007.

[9] A. Lapidoth, S. Shamai (Shitz), and M. A. Wigger, "On the capacity of fading mimo broadcast channels with imperfect transmitter side-information," *http://www.arxiv.org/pdf/cs.IT/0605079*, p. arXiv:cs.IT/0605079, May 2006.

[10] B.M. Zaidel S. Shamai (Shitz), O. Somekh, "Downlink multi-cell processing: An information theoretic view," in *Proceedings of the 3rd International Workshop on Signal Processing for Wireless Communications*, London, UK.,, June 13–15, 2005.

[11] S. Shamai, "Reflections on the gaussian broadcast channel: Progress and challenges," in *Plenary address, <http://www.isit2007.org/index.php> 2007 IEEE International Symposium on Information Theory (ISIT2007)*, Nice, France, June 24 – 29, 2007.

[12] I. E. Teletar, "Capacity of multi-anteanna gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[13] D. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds., pp. 87–105. DIMACS, American Mathematical Society, Providence, RI, 2003.

[14] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[15] S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar, and A. Goldsmith, "Capacity bounds for gaussian vector broadcast channels," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds., pp. 107–122. DIMACS, American Mathematical Society, Providence, RI, 2003.

[16] P. Viswanath and D. Tse, "Sum capacity of the vector gaussian channel and uplink-downlink duality," *IEEE Trans. on Information Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[17] H. Weingarten, T. Liu, S. Shamai (Shitz), Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple input multiple output broadcast compound channel," *Submitted to IEEE trans. on Information Theory.*, 2007.

[18] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[19] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "On the capacity region of the multi-antenna broadcast channel with common messages," in *2006 IEEE International Symposium on Information Theory (ISIT2006)*, Seattle, Washington, USA, July 9 – 14 2006, pp. 2195 – 2199.

[20] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Trans. on Signal Processing*, vol. 54, pp. 161–176, Jan. 2006.

[21] W. Yu and J. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[22] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on colored paper," in *proc. of Int. Symp. on Inf. Theory (ISIT2001)*, Washington, DC, June 24 – 29, 2001, p. 309.