Signaling over arbitrarily permuted parallel channels

REFERENCES

[1] K. A. S. Abdel-Ghaffar and J. H. Weber, "Generalized iterative decoding for linear block codes on the binary erasure channel," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 66–70.

[2] K. A. S. Abdel-Ghaffar and J. H. Weber, "Complete enumeration of stopping sets of full-rank parity-check matrices of Hamming codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3196–3201, Sep. 2007.

[3] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1115–1131, Jun. 2004.

[4] C. Di, D. Proietti, İ. E. Telatar, T. J. Richardson, and R. L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1570–1579, Jun. 2002.

[5] T. Etzion, "On the stopping redundancy of Reed–Muller codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4867–4879, Nov. 2006.

[6] J. Han and P. H. Siegel, "Improved upper bounds on stopping redundancy," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 90–104, Jan. 2007.

[7] H. D. L. Hollmann and L. M. G. M. Tolhuizen, "On parity-check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 823–828, Feb. 2007.

[8] R. Ikegaya, K. Kasai, T. Shibuya, and K. Sakaniwa, "Asymptotic weight and stopping set distributions for detailedly represented irregular LDPC code ensembles," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./ Jul. 2004, p. 208.

[9] N. Kashyap and A. Vardy, "Stopping sets in codes from designs," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun. / Jul. 2003, p. 122.

[10] C. Kelley, D. Sridhara, J. Xu, and J. Rosenthal, "Pseudocodeword weights and stopping sets," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 67.

[11] K. M. Krishnan and P. Shankar, "Computing the stopping distance of a Tanner graph is NP-hard," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2278–2280, Jun. 2007.

[12] M. G. Luby, M. Mitzenbacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[13] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[14] A. McGregor and O. Milenkovic, "On the hardness of approximating stopping and trapping sets in LDPC codes," in *Proc. IEEE Information Theory Workshop*, Lake Tahoe, CA, Sep. 2007, pp. 248–253.

[15] A. Orlitsky, R. Urbanke, K. Viswanathan, and J. Zhang, "Stopping sets and the girth of Tanner graphs," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun. /Jul. 2002, p. 2.

[16] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 929–953, Mar. 2005.

[17] H. Pishro-Nik and F. Fekri, "On decoding of low-density parity-check codes over the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 439–454, Mar. 2004.

[18] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1611–1635, Jul. 2003.

[19] M. Schwartz and A. Vardy, "On the stopping distance and the stopping redundancy of codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 922–932, Mar. 2006.

[20] J. H. Weber and K. A. S. Abdel-Ghaffar, "Stopping set analysis for Hamming codes," in *Proc. Information Theory Workshop on Coding and Complexity*, Rotorua, New Zealand, Aug./Sep. 2005, pp. 244–247.

[21] J. H. Weber and K. A. S. Abdel-Ghaffar, "On decoding failure probabilities for linear block codes on the binary erasure channel," in *Proc. IEEE Int. Information Theory Workshop*, Chengdu, China, Oct. 2006, pp. 24–28.

[22] S.-T. Xia and F. -W. Fu, "On the stopping distance of finite geometry LDPC codes," *IEEE Commun. Lett.*, vol. 10, no. 5, pp. 381–383, May 2006.

Signaling Over Arbitrarily Permuted Parallel Channels

Frans M. J. Willems, *Fellow, IEEE*, and Alexei Gorokhov, *Member, IEEE*

***Abstract*—The problem of transmission of information over arbitrarily permuted parallel channels is studied here. The transmitter does not know over which channel a certain code-sequence will actually be transmitted, however the receiver knows how the sequences are permuted. The permutation is arbitrary but constant during the (simultaneous) transmission of the code-sequences via the parallel channels. It is shown first that the sum of the capacities of each channel is achievable for such a communication system in the special case where the capacity achieving input distributions of all channels are identical. More important is that this sum-capacity can also be achieved using a single channel code for all channels combined with a sequential decoding method. The construction of a rate-matching code based on Maximum Distance Separable (MDS) codes turns out to be crucial. Finally, the case where the parallel channels have different capacity-achieving input distributions is investigated. Also for this case the capacity is determined. Again, this capacity is achievable with a sequential decoding procedure.**

***Index Terms*—Capacity, maximum distance separable (MDS) codes, parallel channels, permuted channels, rate-matching code, sequential decoding method.**

I. INTRODUCTION

In a communication system, it is often the case that the channel is not constant. The transmitter may not be aware of the state of the channel, but for the decoder it is in general not very difficult to find out what the actual state is (was). Certain fading channels correspond to such a situation. Note that the channel can be changing (fading) over time and/or over frequency. Despite the fact that the transmitter is not aware of the state of the channel it would be desirable if the largest possible rate could be achieved at any time and/or for all frequencies. Here we want to investigate how this can be realized. We therefore model the varying channel as a collection of parallel channels and study coding techniques for this situation. We first focus on the case where the same input distribution achieves capacity for all channels in the collection but later we will also consider the case where this assumption does not hold.

The outline of the correspondence is as follows. In Section II, we describe the model that we use in our investigations. Essential is the concept of an input-permuter that connects code-sequences in an arbitrary way to each of the parallel channels. The permutation remains constant as long as it takes to transmit the code-sequences via the channels. The permuter embodies the fact that the transmitter does not know the actual state of the channel. We first consider the case where all channels have the same capacity-achieving input distribution. Section III shows

Manuscript received March 25, 2005; revised December 29, 2006. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, July 2003.

F. M. J. Willems is with Eindhoven University of Technology, Electrical Engineering Department, 5600 MB Eindhoven, The Netherlands, and also with Philips Research, 5656 AA Eindhoven, The Netherlands (e-mail: f.m.j.willems@tue.nl).

A. Gorokhov was with Philips Research, 5656 AA Eindhoven, The Netherlands. He is now with Qualcomm Inc., San Diego, CA 92121-1714 USA (e-mail: gorokhov@qualcomm.com).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.

Color version of Figure 7 in this correspondence is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2007.915912

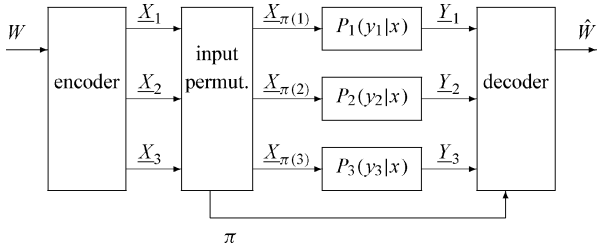

Fig. 1. A transmitter that communicates via an input-permuter followed by three parallel channels to a receiver.

that the capacity of the system is the sum of the capacities of the parallel channels. The proof of this statement is based on the asymptotic equipartition property (AEP), see Cover and Thomas [1]. In Section IV, we propose a sequential decoding procedure combined with a single code for all channels. A crucial ingredient of this method is a so-called rate-matching code that creates the required dependency between the codewords for the parallel channels. We give the definitions of a rate-matching code and discuss some of its properties. Then in Section V we use the AEP to demonstrate that sequential procedures achieve the sum-capacity. In Section VI we propose several methods that can be used to construct rate-matching codes. Section VII discusses an application of the results that we have obtained here. In particular we show that for two flat-fading additive white Gaussian noise (AWGN) channels in parallel, we can decrease the outage probability. The case where the parallel channels have different capacity-achieving input distributions is considered in Section VIII. We determine the capacity also for this case. We conclude by making some final remarks in Section IX.

II. Problem Description

Suppose that we have a communication system in which the transmitter (encoder) is connected to the receiver (decoder) by S parallel channels (see Fig. 1 where $S = 3$). The encoder produces S code-sequences, all having length T. These S code-sequences are transmitted over the S channels, each sequence via one of them. Within T transmissions a message-index W is conveyed by the encoder to the decoder. This index assumes values in $\{1, 2, \ldots, M\}$. The distribution of W is uniform, i.e., $\Pr\{W = w\} = 1/M$ for all $w \in \{1, 2, \ldots, M\}$.

The input alphabets of all S parallel channels are assumed to be discrete and identical; thus

$$\mathcal{X}_1 = \mathcal{X}_2 = \cdots = \mathcal{X}_S = \mathcal{X}. \tag{1}$$

For each message-index W the encoder generates S code-sequences $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_S$, hence

$$(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_S) = e(W) \tag{2}$$

where $\underline{X}_s \triangleq (x_{s1}, x_{s2}, \ldots, x_{sT})$ for $s = 1, \ldots, S$, with $X_{st} \in \mathcal{X}$ for $t = 1, 2, \ldots, T$.

An input-permuter then permutes the S code-sequences over the S parallel channels. More specifically it "connects" channel s to code-sequence $\underline{X}_r$ where $r = \pi(s)$ for $s = 1, \ldots, S$, hence code-sequence $\underline{X}_r$ is transmitted via channel $s = \pi^{-1}(r)$. The mapping $\pi(\cdot)$ from $\{1, \ldots, S\}$ to $\{1, \ldots, S\}$ is one-to-one (a permutation). Since there are $S!$ different permutations we label these permutations $1, 2, \ldots, S!$ here. At a certain moment prior to transmission, one of these permutations is chosen. This permutation then remains constant as long as it takes to transmit the S code-sequences simultaneously via the parallel channels. The encoder is not aware of the chosen permutation π, the decoder is supposed to be informed about π however (or is able to determine it somehow).

All S parallel channels are memoryless. Channel s for $s = 1, \ldots, S$ has transition probability matrix $\{P_s(y_s|x), x \in \mathcal{X}, y_s \in \mathcal{Y}_s\}$. Note that the discrete channel output alphabets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_S$ need not be identical.

The decoder, knowing the actual permutation π, forms an estimate $\hat{W}(\pi) \in \{1, 2, \ldots, M\}$ of the transmitted message-index based on the received S channel output sequences $\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_S$, i.e.

$$\hat{W}(\pi) = d(\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_S, \pi). \tag{3}$$

Here channel output sequence $\underline{Y}_s$, for $s = 1, \ldots, S$, is the output of channel s, hence

$$\underline{Y}_s \triangleq (Y_{s1}, Y_{s2}, \ldots, Y_{sT}) \text{ with } Y_{st} \in \mathcal{Y}_s \text{ for } t = 1, 2, \ldots, T. \tag{4}$$

We now say that rate R is achievable if, for all $\delta > 0$ and all T large enough, there exist an encoder $e(\cdot)$ and decoder $d(\cdot)$ such that both

$$\begin{aligned} &\frac{1}{T}\log_2 M \geq R - \delta \\ &\Pr\{\hat{W}(\pi) \neq W\} \leq \delta, \text{ for all } S! \text{ permutations } \pi. \end{aligned} \tag{5}$$

The largest achievable rate is called the capacity C_{Π}. In the next section we will determine this capacity for the case that one single input distribution achieves capacity for all S parallel channels.

III. Basic Result

The capacity of channel s for $s = 1, \ldots, S$ in bits per transmission is

$$C_s = \max_{Q(\cdot)} I(X; Y_s) \tag{6}$$

for

$$I(X; Y_s) \triangleq \sum_{x, y_s} Q(x) P_s(y_s|x) \log_2 \frac{P_s(y_s|x)}{\sum_{x'} Q(x') P_s(y_s|x')} \tag{7}$$

where in the summations x and x' run over $\mathcal{X}$ and y_s over $\mathcal{Y}_s$. Moreover $\{Q(x), x \in \mathcal{X}\}$ is the channel input distribution.

Theorem 1: The capacity for transmission over S arbitrarily permuted parallel channels equals

$$C_{\Pi} = \sum_{s=1}^{S} C_s, \tag{8}$$

when all channels have an identical input alphabet $\mathcal{X}$ and the same capacity-achieving input distribution $\{Q^*(x), x \in \mathcal{X}\}$.

Note that a certain channel may not have a unique capacity-achieving input distribution. What matters in the theorem however is that there is an input distribution $Q^*(\cdot)$ that achieves capacity for all parallel channels.

In what follows we will prove this result. First note that $\sum_{s=1}^{S} C_s$ is the capacity if both the encoder and the decoder know the actual permutation π. Since the encoder is not aware of the actual permutation we obtain the trivial upper bound

$$C_{\Pi} \leq \sum_{s=1}^{S} C_s. \tag{9}$$

This upper bound turns out to be achievable as we will see next. Observe that for each of the $S!$ permutations π the channel between encoder and decoder is a memoryless product-channel with input alphabet

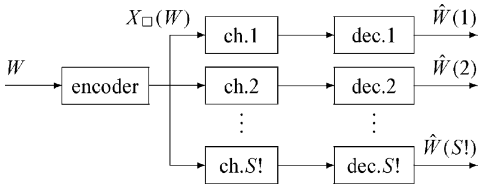

Fig. 2. Simultaneous transmission of a code-block over $S!$ product channels.

$\mathcal{X}^S$ and output alphabet $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_S$. Assume that the super-input distribution of all $S!$ product-channels is

$$\Pr\{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_S \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_S \end{pmatrix}\} = \Pi_{s=1}^{S} Q^*(x_s) \tag{10}$$

where $Q^*(\cdot)$ is the distribution achieving capacity for all S channels. Then for all permutations π

$$\begin{aligned} I(X_1, X_2, \ldots, X_S; Y_1, Y_2, \ldots, Y_S) &= \sum_{s=1}^{S} I(X_{\pi(s)}; Y_s) \\ &= \sum_{s=1}^{S} C_s. \end{aligned} \tag{11}$$

We now design a code whose code-blocks $X_\square \triangleq \{X_{st}, s = 1, \ldots, S, t = 1, \ldots, T\}$ are transmitted over all these $S!$ product channels simultaneously, see Fig. 2, and for which the error probabilities realized by all $S!$ decoders connected to these channels, can be made arbitrarily small. Note that $\Pr\{\hat{W}(\pi) \neq W\}$ denotes the error probability for the decoder that corresponds to permutation π.

Consider a random coding argument along the lines of Cover and Thomas [1, Ch. 7]. Generate for each message index $w \in \{1, 2, \ldots, M = 2^{TR}\}$ an $S \times T$ code-block $X_\square(w) \triangleq \{X_{st}(w), s = 1, \ldots, S, t = 1, \ldots, T\}$. Each component of this block gets value $x \in \mathcal{X}$ with probability $Q^*(x)$ independently of all other components. Note that this code is a random code for all $S!$ product channels where the input distribution of a certain column (super-input) is given by (10).

Every decoder π can now perform decoding by joint typicality, it has to search for a message index w such that the corresponding code-block $X_\square(w)$ is jointly typical with the received block $Y_\square \triangleq \{Y_{st}, s = 1, \ldots, S, t = 1, \ldots, T\}$, relative to the super-channel determined by π. By (11), for rate $R = \sum_{s=1}^{S} C_s - 4\epsilon$, the error probability averaged over the ensemble of codes $\overline{\Pr}\{\hat{W}(\pi) \neq W\}$ is not larger than 2ϵ for each decoder π, for all large enough block lengths T. Here $\epsilon > 0$ is the parameter that specifies the typical set $\mathcal{A}_\epsilon^T$. The probability (averaged over the ensemble of codes) $\overline{\Pr}\{\bigcup_{\pi=1}^{S!} \hat{W}(\pi) \neq W\}$ that any of these $S!$ decoders produces the wrong estimate is therefore not larger than $2S!\epsilon$. Consequently there exists at least one code with total error probability $\Pr\{\bigcup_{\pi=1}^{S!} \hat{W}(\pi) \neq W\}$ not larger than $2S!\epsilon$. Therefore, for this code, for all $S!$ decoders $\pi \in \{1, 2, \ldots, S!\}$ the error probability

$$\Pr\{\hat{W}(\pi) \neq W\} \leq 2S!\epsilon. \tag{12}$$

If we let $\epsilon \to 0$, we can make this error probability and the difference between R and $\sum_{s=1}^{S} C_s$ arbitrarily small. Therefore we may conclude that $\sum_{s=1}^{S} C_s$ is achievable. Together with the upper bound (9) this proves Theorem 1.

Note that our achievability proof resembles the proof given by Cover, McEliece, and Posner [2] for the asynchronous multiple-access channel. In [2] a single code has to be good for all possible delays, here one code must be reliable for all possible permutations.

IV. A Sequential Decoding Procedure

The product-channel approach that was considered in the previous section leads to a large decoding complexity. The decoder has to check all code-blocks $X_\square(w)$ for $w = 1, \ldots, 2^{TR}$ against the received block $Y_\square = \{Y_{st}, s = 1, \ldots, S, t = 1, \ldots, T\}$. In this section we show that $\sum_{s=1}^{S} C_s$ can also be achieved using a single code for all channels and performing a sequential decoding procedure. To make things simple we focus on the case $S = 3$ here. The method that we describe generalizes to larger S however. Also the case where $S = 2$ will be discussed. First, we will describe an important ingredient of a sequential decoding procedure, i.e., the so-called rate-matching code. In Section V we will then prove that it is possible to achieve capacity with a sequential procedure.

To make the decoding effort simpler we use a single code for each channel. Assume that this channel code consist of $M_1 = 2^{TR_1}$ codewords $\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(M_1)$. Codeword $\underline{x}(w_1) = (x_1(w_1), x_2(w_1), \ldots, x_T(w_1))$ is a sequence of T symbols from $\mathcal{X}$, for $w_1 = 1, \ldots, M_1$.

It is obvious that we need some dependency between the codewords that are transmitted over the three channels. This dependence is created by the rate-matching code. This code maps the message-index w onto a triple (w_1, w_2, w_3) of indices. These three indices specify the codewords $\underline{x}(w_1), \underline{x}(w_2)$, and $\underline{x}(w_3)$ that are then sent over three arbitrarily permuted channels. A rate-matching code for $S = 3$ is therefore defined by three mappings

$$\begin{aligned} f_1 &: \{1, 2, \ldots, M\} \to \{1, 2, \ldots, M_1\} \\ f_2 &: \{1, 2, \ldots, M\} \to \{1, 2, \ldots, M_1\} \\ f_3 &: \{1, 2, \ldots, M\} \to \{1, 2, \ldots, M_1\}. \end{aligned} \tag{13}$$

These mappings result in three kinds of subsets of $\{1, 2, \ldots, M_1\}$. For $i, j, k \in \{1, 2, 3\}$ and $i \neq j$, $j \neq k$, and $k \neq i$, we define the subsets

$$\begin{aligned} \mathcal{B}_i &\triangleq \{f_i(w) : \text{for } w = 1, 2, \ldots, M\} \\ \mathcal{B}_{i|j}(w_j) &\triangleq \{f_i(w) : \text{for } w = 1, 2, \ldots, M \text{ such that } f_j(w) = w_j\} \\ \mathcal{B}_{i|jk}(w_j, w_k) &\triangleq \{f_i(w) : \text{for } w = 1, 2, \ldots, M \text{ such that } f_j(w) = w_j \text{ and } f_k(w) = w_k\}. \end{aligned} \tag{14}$$

We are now ready to state what we mean by a rate-matching code.

Definition 1: An (M, M_1, M_2, M_3)-rate-matching code must be one-to-one, hence for all $w = 1, 2, \ldots, M$, $w' = 1, 2, \ldots, M$, and $w' \neq w$

$$(f_1(w), f_2(w), f_3(w)) \neq (f_1(w'), f_2(w'), f_3(w')). \tag{15}$$

Moreover an (M, M_1, M_2, M_3)-rate-matching code must satisfy for all $i, j, k \in \{1, 2, 3\}$ and $i \neq j$, $j \neq k$, and $k \neq i$ the three equalities

$$\begin{aligned} |\mathcal{B}_i| &= M_1 \\ |\mathcal{B}_{i|j}(w_j)| &= M_2 \\ |\mathcal{B}_{i|jk}(w_j, w_k)| &= M_3 \end{aligned} \tag{16}$$

for all $w_j \in \{1, 2, \ldots, M_1\}$ and $w_k \in \mathcal{B}_{k|j}(w_j)$.

A first consequence of (14) and Definition 1 is that for an (M, M_1, M_2, M_3)-rate-matching code

$$M_1 \geq M_2 \geq M_3. \tag{17}$$

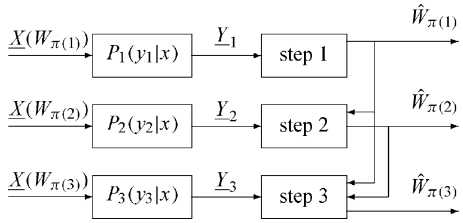

Fig. 3. Information flow during the decoding procedure.

Note also that

$$M = |\{(f_1(w), f_2(w), f_3(w)) \text{ for } w \in \{1, 2, \ldots, M\}\}| = M_1 M_2 M_3. \quad (18)$$

The first equality holds since the rate-matching code is one-to-one, the second equality follows from (16). Rate-matching codes with parameters M, M_1, M_2, and M_3 violating (17) or (18) therefore do not exist. For the moment we will just assume that it is possible to construct rate-matching codes for enough M, M_1, M_2, and M_3 that satisfy (17) and (18). We will however further investigate this issue in Section VI.

Now that we have introduced the concept of a rate-matching code we can give an outline of the sequential decoding procedure. Note that for actual permutation π code-sequences $\underline{X}(W_{\pi(1)})$, $\underline{X}(W_{\pi(2)})$, and $\underline{X}(W_{\pi(3)})$, are the inputs to channel 1, 2, and 3, respectively, see Fig. 3. First we assume, without losing generality, that

$$C_1 \geq C_2 \geq C_3. \quad (19)$$

The decoder now starts the decoding procedure with the channel having the largest capacity, i.e., channel 1. If $\frac{1}{T}\log_2 M_1$ is smaller than C_1 reliable reconstruction of $w_{\pi(1)}$ based on the channel 1 output $\underline{Y}_1$ is possible. Then the decoder proceeds with the output sequence $\underline{Y}_2$ of the second-best channel, i.e., channel 2. Note that $w_{\pi(1)}$ is known to the decoder, and that there are only M_2 possible indices $w_{\pi(2)} \in \mathcal{B}_{2|1}(w_{\pi(1)})$ for the second codeword that need to be considered. When $\frac{1}{T}\log_2 M_2$ is smaller than C_2 reliable reconstruction of $w_{\pi(2)}$ is feasible. Finally, the output $\underline{Y}_3$ of the worst channel, i.e., channel 3, is processed. There are M_3 indices $w_{\pi(3)} \in \mathcal{B}_{3|12}(w_{\pi(1)}, w_{\pi(2)})$ that need to be checked now. For $\frac{1}{T}\log_2 M_3$ smaller than C_3 the correct index $w_{\pi(3)}$ can be found with probability arbitrarily close to one. Finally, from the actual permutation π and $(w_{\pi(1)}, w_{\pi(2)}, w_{\pi(3)})$ the index-triple (w_1, w_2, w_3) and the message-index w can be determined.

V. Achievability Proof for a Sequential Procedure

In this section, we will demonstrate the achievability of $C_1+C_2+C_3$ using a single channel code, a rate-matching code, and a sequential decoding procedure.

A. Code Generation

Consider the input distribution $\{Q^*(x) : x \in \mathcal{X}\}$ that is capacity-achieving for all three channels. We use this distribution to generate a code with $M_1 = 2^{TR_1}$ codewords $\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(M_1)$ all having length T at random. More precisely

$$\Pr\{\underline{X}(w_1) = \underline{x}\} = \Pi_{t=1}^{T} Q^*(x_t) \text{ for all } w_1 = 1, \ldots, M_1. \quad (20)$$

Here $\underline{x} = (x_1, x_2, \ldots, x_T)$. The probability of producing a particular code $\mathcal{C}$ with codewords $\underline{x}(1), \underline{x}(2), \ldots, \underline{x}(M_1)$ is therefore $\Pr\{\mathcal{C}\} = \Pi_{w_1=1}^{M_1} \Pr\{\underline{X}(w_1) = \underline{x}(w_1)\}$.

B. Encoding and Transmission

We will use a $(2^{TR}, 2^{TR_1}, 2^{TR_2}, 2^{TR_3})$-rate-matching code with

$$R = R_1 + R_2 + R_3 \quad (21)$$

for $R_1 \geq R_2 \geq R_3$ that will be specified later. This rate-matching code transforms message index w into an index-triple (w_1, w_2, w_3). The resulting codewords $\underline{X}(w_1)$, $\underline{X}(w_2)$, and $\underline{X}(w_3)$ are now ready for transmission over the three channels, i.e.

$$\underline{X}_1 = \underline{X}(w_1), \qquad \underline{X}_2 = \underline{X}(w_2), \qquad \underline{X}_3 = \underline{X}(w_3). \quad (22)$$

The input-permuter permutes these codewords over the three channels. For $s = 1, 2, 3$ the input of channel s is $\underline{X}_{\pi(s)}$. The decoder receives the three sequences $\underline{Y}_1$, $\underline{Y}_2$, and $\underline{Y}_3$. Note that

$$\Pr\{\underline{y}_1, \underline{y}_2, \underline{y}_3 | \underline{x}_{\pi(1)}, \underline{x}_{\pi(2)}, \underline{x}_{\pi(3)}\} = \Pi_{s=1}^{3} \Pi_{t=1}^{T} P_s(y_{st}|x_t(w_{\pi(s)})). \quad (23)$$

Here $x_t(w_r)$ is the t-th component of codeword $\underline{X}(w_r)$, for $w_r = 1, M_1$ and $r = 1, 2, 3$.

C. Decoding Procedure

The decoder uses a sequential procedure based on decoding by joint typicality. For exact definitions of the sets of jointly typical sequences $\mathcal{A}_\epsilon^T(X, Y_1)$, $\mathcal{A}_\epsilon^T(X, Y_2)$, and $\mathcal{A}_\epsilon^T(X, Y_3)$, we refer to Cover and Thomas [1, Sec. 7.6]. The distributions that determine these sets are $P(x, y_s) = Q^*(x)P_s(y_s|x)$ for $s = 1, 2, 3$. Note that by (19) we have

$$\begin{aligned} I(X_{\pi(1)}; Y_1) = C_1 &\geq I(X_{\pi(2)}; Y_2) = C_2 \\ &\geq I(X_{\pi(3)}; Y_3) = C_3. \end{aligned} \quad (24)$$

The decoder first decodes the message-index transmitted over the channel with capacity C_1. It declares that the index $\hat{w}_{\pi(1)}$ was sent if there is a unique index $\hat{w}_{\pi(1)}$ such that the pair $(\underline{X}(\hat{w}_{\pi(1)}), \underline{Y}_1) \in \mathcal{A}_\epsilon^T(X, Y_1)$, i.e., if this pair is jointly typical. If no such $\hat{w}_{\pi(1)}$ exists or there are more than one such, then an error is declared and decoding stops. If not, the decoder proceeds with the channel having capacity C_2. It declares that $\hat{w}_{\pi(2)}$ was transmitted if there is a unique index $\hat{w}_{\pi(2)}$ such that the pair $(\underline{X}(\hat{w}_{\pi(2)}), \underline{Y}_2) \in \mathcal{A}_\epsilon^T(X, Y_2)$ and $\hat{w}_{\pi(2)} \in \mathcal{B}_{2|1}(\hat{w}_{\pi(1)})$. If no such $\hat{w}_{\pi(2)}$ exists or there are more than one such, an error is declared and decoding stops. If not, the decoder processes the output of channel 3, the channel with capacity C_3. It declares that index $\hat{w}_{\pi(3)}$ was sent if there is a unique index $\hat{w}_{\pi(3)}$ such that the pair $(\underline{X}(\hat{w}_{\pi(3)}), \underline{Y}_3) \in \mathcal{A}_\epsilon^T(X, Y_3)$ and $\hat{w}_{\pi(3)} \in \mathcal{B}_{3|12}(\hat{w}_{\pi(1)}, \hat{w}_{\pi(2)})$. If there is no such $\hat{w}_{\pi(3)}$ or there are more than one such, an error is declared.

From the actual permutation π and $(\hat{w}_{\pi(1)}, \hat{w}_{\pi(2)}, \hat{w}_{\pi(3)})$ the index-triple $(\hat{w}_1, \hat{w}_2, \hat{w}_3)$ and the message-index $\hat{w}$ can be determined.

D. Analysis of Probability of Error

First we fix a certain permutation π and investigate what happens when π is the actual permutation. We define the following events:

$$\begin{aligned} E_{w_1}^1 &= \{(\underline{X}(w_1), \underline{Y}_1) \in \mathcal{A}_\epsilon^T(X, Y_1)\}, \\ E_{w_2}^2 &= \{(\underline{X}(w_2), \underline{Y}_2) \in \mathcal{A}_\epsilon^T(X, Y_2)\}, \\ E_{w_3}^3 &= \{(\underline{X}(w_3), \underline{Y}_3) \in \mathcal{A}_\epsilon^T(X, Y_3)\} \end{aligned} \quad (25)$$

for $w_1, w_2, w_3 \in \{1, 2, \ldots, 2^{TR_1}\}$. Recall that $\underline{Y}_1$, $\underline{Y}_2$, and $\underline{Y}_3$ are the results of sending W. Moreover W_1, W_2, and W_3 are the random

variables induced by the random variable W. The error probability averaged over the ensemble of codes for actual permutation π is

$$\begin{aligned}
&\overline{\Pr}\{\hat{W}(\pi) \neq W\} \\
&= \overline{\Pr}\left\{E^{1c}_{W_{\pi(1)}} \cup \left(\cup_{w_1 \neq W_{\pi(1)}} E^1_{w_1}\right)\right. \\
&\quad \cup E^{2c}_{W_{\pi(2)}} \cup \left(\cup_{w_2 \neq W_{\pi(2)}, w_2 \in \mathcal{B}_{2|1}(W_{\pi(1)})} E^2_{w_2}\right) \\
&\quad \left. \cup E^{3c}_{W_{\pi(3)}} \cup \left(\cup_{w_3 \neq W_{\pi(3)}, w_3 \in \mathcal{B}_{3|12}(W_{\pi(1)}, W_{\pi(2)})} E^3_{w_3}\right)\right\} \\
&\leq \overline{\Pr}\{E^{1c}_{W_{\pi(1)}}\} + \sum_{w_1 \neq W_{\pi(1)}} \overline{\Pr}\{E^1_{w_1}\} + \overline{\Pr}\{E^{2c}_{W_{\pi(2)}}\} \\
&\quad + \sum_{w_2 \neq W_{\pi(2)}, w_2 \in \mathcal{B}_{2|1}(W_{\pi(1)})} \overline{\Pr}\{E^2_{w_2}\} + \overline{\Pr}\{E^{3c}_{W_{\pi(3)}}\} \\
&\quad + \sum_{w_3 \neq W_{\pi(3)}, w_3 \in \mathcal{B}_{3|12}(W_{\pi(1)}, W_{\pi(2)})} \overline{\Pr}\{E^3_{w_3}\} \qquad (26)
\end{aligned}$$

where it is understood that probability $\overline{\Pr}\{E^{1c}_{W_{\pi(1)}}\}$ denotes $\sum_{w=1}^{2^{TR}} \Pr\{W = w\}\ \overline{\Pr}\{E^{1c}_{f_{\pi(1)}(w)}\}$, and the sum of probabilities $\sum_{w' \neq W_{\pi(1)}} \overline{\Pr}\{E^1_{w'}\}$ denotes $\sum_{w=1}^{2^{TR}} \Pr\{W = w\} \sum_{w' \neq f_{\pi(1)}(w)} \overline{\Pr}\{E^1_{w'}\}$, etc. It follows from Cover and Thomas [1, Th. 7.6.1, part 1] that

$$\begin{aligned}
\overline{\Pr}\{E^{1c}_{W_{\pi(1)}}\} &\leq \epsilon \\
\overline{\Pr}\{E^{2c}_{W_{\pi(2)}}\} &\leq \epsilon \\
\overline{\Pr}\{E^{3c}_{W_{\pi(3)}}\} &\leq \epsilon \qquad (27)
\end{aligned}$$

for all T large enough. Moreover, part 3 of this theorem implies that

$$\begin{aligned}
\overline{\Pr}\{E^1_{w_1}\} &\leq 2^{-T(I(X;Y_1)-3\epsilon)} = 2^{-T(C_1-3\epsilon)} \\
\overline{\Pr}\{E^2_{w_2}\} &\leq 2^{-T(I(X;Y_2)-3\epsilon)} = 2^{-T(C_2-3\epsilon)} \\
\overline{\Pr}\{E^3_{w_3}\} &\leq 2^{-T(I(X;Y_3)-3\epsilon)} = 2^{-T(C_3-3\epsilon)} \qquad (28)
\end{aligned}$$

for $w_1 \neq W_{\pi(1)}$, $w_2 \neq W_{\pi(2)}$, and $w_3 \neq W_{\pi(3)}$. If we now use a rate-matching code with

$$R_1 = C_1 - 4\epsilon, \qquad R_2 = C_2 - 4\epsilon, \qquad R_3 = C_3 - 4\epsilon \qquad (29)$$

then using (16) we obtain

$$\begin{aligned}
&\overline{\Pr}\{\hat{W}(\pi) \neq W\} \\
&\leq \epsilon + 2^{TR_1} 2^{-T(C_1-3\epsilon)} + \epsilon + 2^{TR_2} 2^{-T(C_2-3\epsilon)} \\
&\quad + \epsilon + 2^{TR_3} 2^{-T(C_3-3\epsilon)} \\
&\leq 6\epsilon \qquad (30)
\end{aligned}$$

for all T large enough. There are $S! = 3! = 6$ possible permutations however. Therefore

$$\overline{\Pr}\{\cup_{\pi=1}^{6} \hat{W}(\pi) \neq W\} \leq 36\epsilon. \qquad (31)$$

This implies that for all T large enough there exist at least one code for which the error probability $\Pr\{\hat{W}(\pi) \neq W\}$ is smaller than 36ϵ for all 6 permutations π when $\frac{1}{T}\log_2 M = C_1 + C_2 + C_3 - 12\epsilon$. Consequently $C_1 + C_2 + C_3$ is achievable with a single code and a sequential decoding procedure.

Comment: Observe that this proof generalizes to arbitrary $S > 3$ but also to the case where $S = 2$. For $S = 2$ we can use a $(2^{T(R_1+R_2)}, 2^{TR_1}, 2^{TR_2})$-rate-matching code based on two mappings f_1 and f_2, with $R_1 = C_1 - 4\epsilon$ and $R_2 = C_2 - 4\epsilon$ which is very easy to construct as we shall see in the next section.

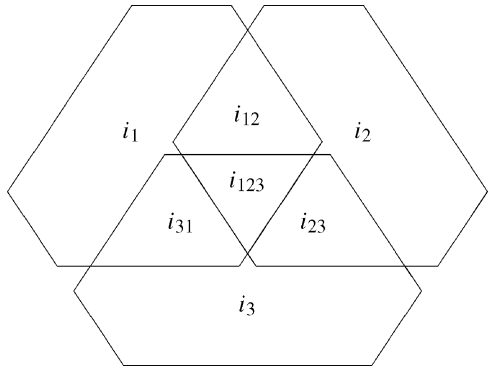

Fig. 4. Indices in a Venn-diagram.

VI. Construction of Rate-Matching Codes

In Section IV we have introduced rate-matching codes. We assumed there that rate-matching codes owning the properties stated in Definition 1 exist. Here we will show how such rate-matching codes can be constructed. We start with codes for $S = 3$.

A. Construction Based on a Venn-Diagram for $S = 3$

Although so far we have considered indices i assuming values in $\{1, \ldots, M\}$ for some positive integer M, from now on we assume that the set of possible indices is $\{0, 1, \ldots, 2^L - 1\}$ where L is a nonnegative integer. This implies that index i is equivalent to some binary sequence $(b_1, b_2, \ldots, b_L)$ in the sense that $i = \sum_{l=1}^{L} b_l 2^{l-1}$. Observe that digits $B_1, B_2, \ldots, B_L$ in the binary sequence should be independent and uniformly distributed if the index I must be uniform, and vice versa. Next we define the combined index (i', i'') of two indices i' and i'' as the index that is equivalent to the concatenation $(b'_1, b'_2, \ldots, b'_{L'}, b''_1, b''_2, \ldots, b''_{L''})$ of the binary sequences $(b'_1, b'_2, \ldots, b'_{L'})$ and $(b''_1, b''_2, \ldots, b''_{L''})$ that correspond to i' and i'' respectively. Note that the combined index (i', i'') assumes values in $\{0, 1, \ldots, 2^{L'+L''} - 1\}$. Three or more indices can been combined in a similar way.

In order to construct a rate-matching code, we consider 7 indices i_1, i_2, i_3, i_{12}, i_{23}, i_{31}, and i_{123}. Assume that indices i_1, i_2, and i_3 each correspond to sequences consisting of L_1 binary digits. Indices i_{12}, i_{23}, and i_{31} correspond to sequences of L_2 binary digits, and index i_{123} to a sequence of L_3 digits. The three message indices w_1, w_2, and w_3 are now obtained by combining the indices $i_1, i_2, \ldots$, and i_{123} in a "Venn-diagram"-manner, see Fig. 4, i.e.

$$\begin{aligned}
w_1 &= (i_1, i_{12}, i_{31}, i_{123}) \\
w_2 &= (i_2, i_{23}, i_{12}, i_{123}) \\
w_3 &= (i_3, i_{31}, i_{23}, i_{123}). \qquad (32)
\end{aligned}$$

This construction results in a $(2^{TR}, 2^{TR_1}, 2^{TR_2}, 2^{TR_3})$-rate-matching code with $R = R_1 + R_2 + R_3$ and

$$\begin{aligned}
TR_1 &= L_1 + 2L_2 + L_3 \\
TR_2 &= L_1 + L_2 \\
TR_3 &= L_1 \qquad (33)
\end{aligned}$$

for all nonnegative integers L_1, L_2, and L_3. The nonnegativity of the L_1, L_2, and L_3 now implies that the inequalities

$$\begin{aligned}
R_1 &\geq R_2 \geq R_3 \geq 0 \\
R_1 - R_2 &\geq R_2 - R_3 \qquad (34)
\end{aligned}$$

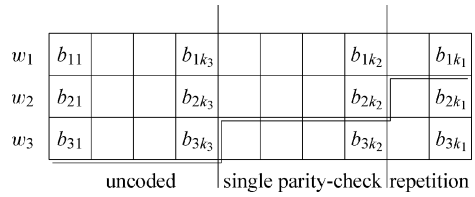

Fig. 5. A rate-matching code for three parallel channels.

must hold. The last inequality in (34) now causes a problem if we want to achieve $C_1+C_2+C_3$ for capacities for which $C_1-C_2 < C_2-C_3$. A construction that is effective also for such capacities will be presented next however.

The Venn-diagram construction was proposed and investigated by the second author in [3]. Note that for $S = 2$ (i.e., the case with two parallel channels) the Venn-diagram method does not impose a constraint on R_1 and R_2 other than $R_1 \geq R_2$. Therefore the Venn-diagram construction is optimal for $S = 2$.

B. Construction Based on Binary Codes for $S = 3$

Consider a matrix of binary digits $\{b_{sk} : s = 1, \ldots, 3, k = 1, \ldots, k_1\}$. Furthermore let $k_1 \geq k_2 \geq k_3 \geq 0$ for integers k_1, k_2, and k_3, see Fig. 5. The binary digits in the matrix relate to each other in the following way:

- The digits b_{sk}, for $s = 1$ and $k = 1, \ldots, k_1$, for $s = 2$ and $k = 1, \ldots, k_2$, and for $s = 3$ and $k = 1, \ldots, k_3$ are *information symbols*. They can be chosen independently and together these digits represent the message index w.
- Note that in columns $k = 1, \ldots, k_3$ the digits b_{1k}, b_{2k}, and b_{3k} are information symbols.
- In columns $k = k_3 + 1, \ldots, k_2$ the digits b_{1k} and b_{2k} are information symbols. The digits b_{3k} are *parity symbols*, i.e.

$$b_{3k} = b_{1k} \oplus b_{2k} \tag{35}$$

  where $\oplus$ is modulo-2 addition.
- In columns $k = k_2 + 1, \ldots, k_3$ the digits b_{1k} are information symbols. The digits b_{2k} and b_{3k} are *parities* again, but now

$$\begin{aligned} b_{2k} &= b_{1k} \\ b_{3k} &= b_{1k}. \end{aligned} \tag{36}$$

Note that in columns $k = k_2 + 1, k_1$ we are using a length-three binary repetition code and each of the binary digits in such a column determines the other two digits in that column. In columns $k = k_3+1$, k_2 we apply a binary single-parity-check code and each pair of binary digits in such a column determines the remaining digit in that column. Columns $k = 1, k_3$ are uncoded, all eight digit-combinations are possible in such a column.

If we now define the indices as follows:

$$\begin{aligned} w_1 &\triangleq \sum_{k=1}^{k_1} b_{1k} 2^{k-1} \\ w_2 &\triangleq \sum_{k=1}^{k_1} b_{2k} 2^{k-1} \\ w_3 &\triangleq \sum_{k=1}^{k_1} b_{3k} 2^{k-1} \end{aligned} \tag{37}$$

we have constructed a $(2^{TR}, 2^{TR_1}, 2^{TR_2}, 2^{TR_3})$-rate-matching code with $R = R_1 + R_2 + R_3$ and

$$\begin{aligned} TR_1 &= k_1 \\ TR_2 &= k_2 \\ TR_3 &= k_3 \end{aligned} \tag{38}$$

for integer k_1, k_2, and k_3. Although $k_1 \geq k_2 \geq k_3 \geq 0$ implies that condition

$$R_1 \geq R_2 \geq R_3 \geq 0 \tag{39}$$

should hold, this causes no problem. The consequence of this is that $C_1 + C_2 + C_3$ is also achievable with a single channel code together with a rate-matching code based on simple binary codes and sequential decoding.

C. Construction of a Rate-Matching Code for $S > 3$

If we consider the case where we have $S > 3$ parallel channels, then, to create a rate-matching code, for each $k = 1, \ldots, S$ we need codes of length S with the property that any k symbols of a codeword fully determine the remaining symbols. Codes that have this property are called maximum distance separable (MDS), see MacWilliams and Sloane [4, Ch. 11]. The symbols in these codes are not always binary, as in the previous subsection. Our construction is based on Reed–Solomon codes over GF(2^m), where we take m such that $2^m - 1 \geq S$. There exist $[2^m - 1, k, 2^m - k]$ Reed-Solomon codes for all $k = 1, \ldots, 2^m - 1$. Here $[n, k, d]$ refers to a linear code with length n, dimension k, and minimum Hamming distance d. These Reed-Solomon codes all have the MDS property and therefore any k symbols may be taken as information symbols.

Take $k = 1, \ldots, S$ in what follows. It is a consequence of [4, Corollary 3, Ch. 11] that since MDS-code $[2^m - 1, k, 2^m - k]$ exists over $GF(2^m)$ there exists another MDS-code $[2^m - 1 - j, k, 2^m - k - j]$ as long as $j \leq 2^m - 1 - k$. Just delete j columns from the original generator matrix, then there are at least k columns left. If we take $j = 2^m - 1 - S$ then we get an $[S, k, S - k + 1]$ MDS code over $GF(2^m)$. Since $S \geq k$ this is allowed.

Using a similar construction as in the previous subsection, based on the MDS property of the constituent codes, we obtain a $(2^{TR}, 2^{TR_1}, 2^{TR_2}, \ldots, 2^{TR_S})$-rate-matching code with $R = R_1 + R_2 + \cdots + R_S$ and

$$\begin{aligned} TR_1 &= mk_1 \\ TR_2 &= mk_2 \\ &\cdots \\ TR_S &= mk_S \end{aligned} \tag{40}$$

for negative integers $k_1, k_2, \ldots, k_S$ satisfying $k_1 \geq k_2 \geq \cdots \geq k_S$; see Fig. 6. Therefore, again the only restriction for the rates is that

$$R_1 \geq R_2 \geq \cdots \geq R_S \geq 0. \tag{41}$$

Therefore $\sum_{s=1}^{S} C_s$ is also achievable for $S > 3$ with a single channel code combined with an MDS-code based rate-matching code and sequential decoding.

D. A Pseudo Rate-Matching Code Based on Joint Typicality

In Section IV we have defined what we mean by a rate-matching code. Moreover in this section we have described how rate-matching codes can be constructed. Here we will consider special sets of jointly-typical sequences that have properties which are not as strict as those of rate-matching codes, but nevertheless these special sets can be used as

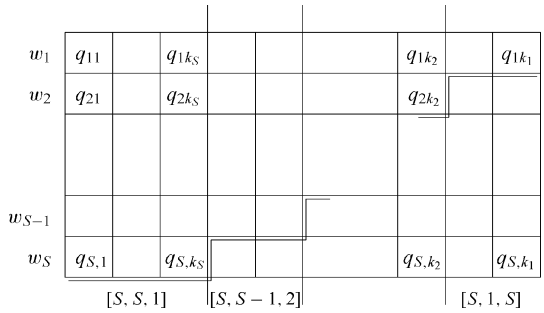

Fig. 6. A rate-matching code for S parallel channels.

(pseudo) rate-matching codes in our achievability proof. Consider three discrete random variables U_1, U_2, U_3 that have entropies $H(U_1) = H(U_2) = H(U_3) = h_1$, $H(U_1, U_2) = H(U_2, U_3) = H(U_3, U_1) = h_1 + h_2$, and $H(U_1, U_2, U_3) = h_1 + h_2 + h_3$. Fix an integer sequence length n, a $\delta > 0$, and observe now that for all $i, j, k \in \{1, 2, 3\}$ and $i \neq j$, $j \neq k$, and $k \neq i$

$$\begin{aligned} |\mathcal{A}_\delta^n(U_i)| &\leq 2^{n(h_1+\delta)} \\ |\mathcal{A}_\delta^n(U_i|\underline{u}_j)| &\leq 2^{n(h_2+2\delta)} \\ |\mathcal{A}_\delta^n(U_i|\underline{u}_j, \underline{u}_k)| &\leq 2^{n(h_3+2\delta)} \end{aligned} \tag{42}$$

where $\mathcal{A}_\delta^n(U_i)$, $\mathcal{A}_\delta^n(U_i|\underline{u}_j)$, and $\mathcal{A}_\delta^n(U_i|\underline{u}_j, \underline{u}_k)$ denote sets of (conditionally) typical sequences $\underline{u}_i$, of length n, see Cover and Thomas [1, Sec. 15.2]. Note that from comparing (16) and (42) we may conclude that the sets containing sequences $\underline{u}_1$, $\underline{u}_2$, and $\underline{u}_3$ satisfy inequalities that are similar to the equalities for the sets containing indices w_1, w_2 and w_3. Since the inequalities in (42) are upper bounds for the set cardinalities we could use the sequences $\underline{u}_1$, $\underline{u}_2$, and $\underline{u}_3$ instead of w_1, w_2 and w_3 as indices to $2^{n(h_1+\delta)}$ randomly chosen codewords in our achievability proof. If moreover we use te sequence-triple $(\underline{u}_1, \underline{u}_2, \underline{u}_3)$ as a replacement for message-index w it follows immediately that our pseudo rate-matching code is one-to-one. A crucial observation is that

$$|\mathcal{A}_\delta^n(U_1, U_2, U_3)| \geq 2^{n(h_1+h_2+h_3-2\delta)} \tag{43}$$

for all n large enough. Therefore we can use this pseudo rate-matching code in our achievability proof to convey at least $TR = n(h_1 + h_2 + h_2 - 2\delta)$ bits to the receiver if we set $n(h_1 + \delta) = TR_1$, $n(h_2 + 2\delta) = TR_2$, and $n(h_3 + 2\delta) = TR_3$. A problem that still remains to be solved is to find probability distributions $\{P(u_1, u_2, u_3), u_1 \in \mathcal{U}_1, u_2 \in \mathcal{U}_2, u_3 \in \mathcal{U}_3\}$ with desired (conditional) entropies h_1, h_2, and h_3.

As a final remark we mention that the above method can also be used in a more direct way to transmit the output sequences $\underline{u}_1$, $\underline{u}_2$, and $\underline{u}_3$ generated by a correlated i.i.d. source with generic random variables U_1, U_2, and U_3 as described before, to the receiver, over arbitrarily permuted channels if $T(C_1 - 4\epsilon) \geq n(h_1 + \delta)$, $T(C_2 - 4\epsilon) > n(h_2 + 2\delta)$, and $T(C_3 - 4\epsilon) > n(h_3 + 2\delta)$. This technique is a kind of Slepian-Wolf [5] coding, since the encoders for the (dependent) separate sequences can operate independently of each other and yet reliable transmission is possible for total source entropy $h_1 + h_2 + h_3$ not larger than but arbitrarily close to the total channel capacity $C_1 + C_2 + C_3$, forgetting about the factor T/n for a moment.

VII. An Application

We will study an application in this subsection based on AWGN channels. Note that so far we have only considered discrete channels here. It is not hard to see that our results carry over to the AWGN case however.

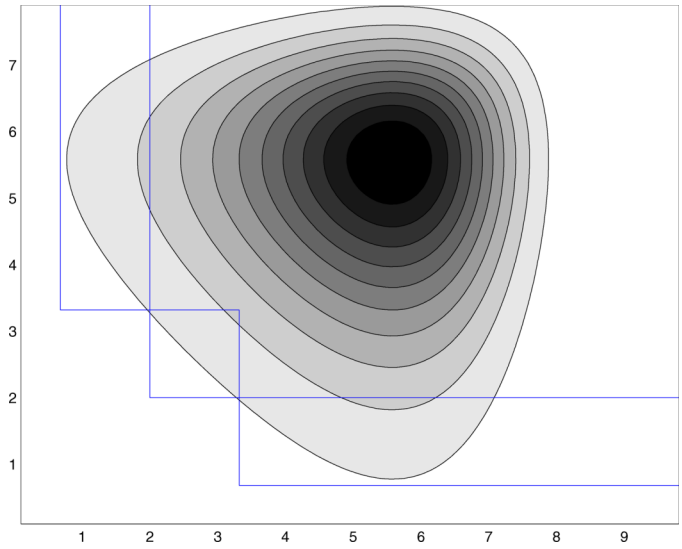

Fig. 7. Contour plot of the probability density function of the capacity pair (C_1, C_2).

Consider a communication system with two parallel AWGN channels that are described by the equations

$$\begin{aligned} Y_1 &= A_1 X_1 + N_1 \\ Y_2 &= A_2 X_2 + N_2 \end{aligned} \tag{44}$$

where X_1 and X_2 are complex channel input variables, N_1 and N_2 are circularly symmetric complex Gaussian noise variables, and A_1 and A_2 are complex channel coefficients. We assume that A_1 and A_2 are circularly symmetric complex randomly chosen Gaussians with variance 1. The moduli of the channel coefficients therefore have a Rayleigh density, i.e.,

$$p_{|A_1|}(a) = p_{|A_2|}(a) = 2a \exp(-a^2), \text{ for } a \geq 0. \tag{45}$$

Moreover we suppose that the coefficients are generated independently of each other and of the channel inputs, and that they are constant over the duration of the code-sequences. The channel inputs are power-constrained, i.e., $E[|X_1|^2] \leq P$ and $E[|X_2|^2] \leq P$. The noise variables have variances $E[|N_1|^2] = E[|N_2|^2] = 1$.

Capacity is now achieved if both X_1 and X_2 are circularly symmetric complex Gaussians with total variance P. The capacity of a channel however depends on the channel coefficient, hence

$$\begin{aligned} C_1 &= \log_2(1 + |A_1|^2 P) \\ C_2 &= \log_2(1 + |A_2|^2 P). \end{aligned} \tag{46}$$

Observe that therefore the capacities C_1 and C_2 are random variables. For $P = 48$ we have computed the probability density function of the pair (C_1, C_2). A contour plot of this density can be found in Fig. 7. Note that the density is largest when both capacities are between 5 and 6 bits.

A) Now suppose first that we communicate independently over both channels and apply a code with rate $R = 2$ bits for each channel. Then communication can only be reliable if both capacities $C_1 \geq 2$ bits and $C_2 \geq 2$ bits. If this is not the case we speak about outage. The outage probability is defined as

$$P_{\text{out}}^A \triangleq 1 - \Pr\{(C_1 \geq 2) \wedge (C_2 \geq 2)\} \tag{47}$$

see Fig. 7. Note that the channels have a capacity larger than 2 bits if $|A_1|^2 \geq 1/16$ and $|A_2|^2 \geq 1/16$. Therefore the outage probability

$$P_{\text{out}}^{A} = 1 - (\exp(-1/16))^2 = 0.1175. \tag{48}$$

B) Next note that a total rate of 4 bits can also be achieved if we can transmit with rate $R_1 = 3.3$ bits over the strong channel and with rate $R_2 = 0.7$ bits over the weak channel. Note that since the transmitter does not know which channel is strong and which is weak we could use the signaling method that we have developed here. We may assume that the receiver (e.g., by applying pilots) knows the state of the channels however. Assuming that channel i is the strong channel and channel j is the weak channel, both inequalities $|A_i|^2 \geq 0.1844$ and $|A_j|^2 \geq 0.0130$ should be satisfied. Now for the outage probability we can write

$$\begin{aligned} P_{\text{out}}^{B} &\triangleq 1 - \Pr\{((C_1 \geq 0.7) \wedge (C_2 \geq 3.3)) \\ &\quad \vee ((C_1 \geq 3.3) \wedge (C_2 \geq 0.7))\} \\ &= 1 - (e^{-0.0130})^2 + (e^{-0.0130} - e^{-0.1844})^2 \\ &= 0.0498 \end{aligned} \tag{49}$$

see again Fig. 7.

We may conclude that using our coding method for transmission over arbitrarily permuted parallel channels results in a smaller outage probability. The reason for getting an improvement is simply that we are more flexible. Condition $(C_1 \geq 2) \wedge (C_2 \geq 2)$ is just a special case of condition $((C_1 \geq 2-\delta) \wedge (C_2 \geq 2+\delta)) \vee ((C_1 \geq q2+\delta) \wedge (C_2 \geq 2-\delta))$. Better results can be obtained for $\delta = 1.3$.

VIII. Channels With Different Capacity-Achieving Distributions

So far we have only focussed on parallel channels that all achieve capacity for the same input distribution. In the present section we will consider the case where these channels have different capacity-achieving distributions however. The result that we obtain here is stated in the next theorem.

Theorem 2: The capacity for transmission over S arbitrarily permuted parallel channels equals

$$C_{\Pi} = \max_{Q(\cdot)} \sum_{s=1}^{S} I(X;Y_s) \tag{50}$$

where $I(X;Y_s)$ is as defined in (7). Note that all channels have an identical input alphabet $\mathcal{X}$.

Note that now C_{Π} is in general smaller than the sum of the capacities of each of the parallel channels. When there is a single distribution that achieves capacity for all parallel channels we get equality however.

The proof of this theorem actually consists only of a converse part. Achievability, both for the basic and the sequential case, follows immediately from the achievability proofs in Sections III and V if we replace $\{Q^*(x), x \in \mathcal{X}\}$ by the (or a) distribution that achieves the maximum in (50).

To prove the converse we first fix an $\epsilon > 0$ and a block-length T. Now consider a permutation π and a code that achieves $\Pr\{\hat{W}(\pi) \neq W\} \leq \epsilon$. Then

$$\begin{aligned} \log_2(M) &\leq H(W) \\ &= I(W;\underline{Y}_1,\underline{Y}_2,\ldots,\underline{Y}_S) + H(W|\underline{Y}_1,\underline{Y}_2,\ldots,\underline{Y}_S) \\ &\leq I(\underline{X}_1,\underline{X}_2,\ldots,\underline{X}_S;\underline{Y}_1,\underline{Y}_2,\ldots,\underline{Y}_S) + H(W|\hat{W}(\pi)) \\ &\leq \sum_{s=1}^{S} I(\underline{X}_{\pi(s)};\underline{Y}_s) + H(W|\hat{W}(\pi)) \\ &\leq \sum_{s=1}^{S}\sum_{t=1}^{T} I(X_{\pi(s),t};Y_{s,t}) + 1 + \epsilon \log_2(M). \end{aligned} \tag{51}$$

Note that in the last step we used Fano's inequality. Rewriting (51) leads, for all permutations π, to

$$\log_2(M) \leq \frac{1}{1-\epsilon}\left[\sum_{s=1}^{S}\sum_{t=1}^{T} I(X_{\pi(s),t};Y_{s,t}) + 1\right]. \tag{52}$$

Since this code has to be good for all permutations, we can combine all these inequalities. We then obtain

$$\begin{aligned} \frac{\log_2(M)}{T} &\leq \frac{1}{S!}\sum_{\pi=1}^{S!}\frac{1}{1-\epsilon}\left[\sum_{s=1}^{S}\frac{1}{T}\sum_{t=1}^{T} I(X_{\pi(s),t};Y_{s,t}) + \frac{1}{T}\right] \\ &= \frac{1}{1-\epsilon}\left[\sum_{s=1}^{S}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{S!}\sum_{\pi=1}^{S!} I(X_{\pi(s),t};Y_{s,t}) + \frac{1}{T}\right] \\ &\leq \frac{1}{1-\epsilon}\left[\sum_{s=1}^{S}\frac{1}{T}\sum_{t=1}^{T} I(X_t;Y_{s,t}) + \frac{1}{T}\right] \\ &\leq \frac{1}{1-\epsilon}\left[\sum_{s=1}^{S} I(X;Y_s) + \frac{1}{T}\right]. \end{aligned} \tag{53}$$

The second inequality follows from the convexity of mutual information over the channel's input distribution. We assume that for all $t = 1,\ldots,T$ the variables $\overline{X}_{1t},\overline{X}_{2t},\ldots,\overline{X}_{St}$ are obtained from permuting $X_{1t},X_{2t},\ldots,X_{St}$ in all $S!$ ways, uniformly. This results in

$$\begin{aligned} &\Pr\{\overline{X}_{1t} = x_1,\ldots,\overline{X}_{St} = x_S\} \\ &\qquad = \frac{1}{S!}\sum_{\pi=1}^{S!}\Pr\{X_{\pi(1)t} = x_1,\ldots,X_{\pi(S)t} = x_S\} \end{aligned} \tag{54}$$

for all $x_1, x_2, \ldots, x_S \in \mathcal{X}$. This implies that

$$\begin{aligned} \Pr\{\overline{X}_{1t} = x\} &= \Pr\{\overline{X}_{2t} = x\} = \cdots = \Pr\{\overline{X}_{St} = x\} \\ &= \frac{1}{S}\sum_{s=1}^{S}\Pr\{X_{st} = x\} \end{aligned} \tag{55}$$

for all $x \in \mathcal{X}$. Therefore $\overline{X}_{1t},\overline{X}_{2t},\ldots\overline{X}_{St}$ are all random variables with the same probability distribution and we denote them all by X_t.

Also the third inequality follows from the convexity of mutual information over the channel's input distribution. If we take

$$\Pr\{X = x\} = \frac{1}{T}\sum_{t=1}^{T}\Pr\{X_t = x\} \tag{56}$$

for all $x \in \mathcal{X}$, we get the third inequality.

The converse now ends in the standard way, i.e., by letting $T \to \infty$ and $\epsilon \downarrow 0$.

IX. Final Remarks

Our rate-matching code followed by the randomly chosen code resembles a Blokh–Zyablov generalized concatenated code [6]. In the Blokh–Zyablov construction outer codes are used with different dimensions but all having the same length, and a single inner code, see [7]. The column codes in our rate-matching construction in Sections VI-B and VI-C are equivalent to the outer codes. Our randomly generated code is the equivalent of the inner code in the Blokh–Zyablov construction. What makes our code construction special is the fact that our code is used to transmit efficiently over a number of arbitrary permuted parallel channels while the Blokh–Zyablov motivation was to find codes with good distance properties.

The methods that were proposed here have been used to improve the performance of V-BLAST systems proposed by Foschini [8], see

[9]. In a V-BLAST systems several input streams are transmitted over layers whose capacities can differ from block to block.

ACKNOWLEDGMENT

The authors wish to thank their colleagues Ludo Tolhuizen and Andries Hekstra, but especially the two anonymous reviewers, for valuable comments and insightful remarks. The remarks of one of the reviewers motivated us to investigate the case where the capacity-achieving input distributions for the parallel channels differ from each other.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second ed. New York: Wiley-Interscience, 2006.

[2] T. M. Cover, R. J. McEliece, and E. C. Posner, "Asynchronous multiple-access channel capacity," *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 409–414, Jul. 1981.

[3] A. Gorokhov, "Layered MIMO transceivers with enhanced capacities: A new coding strategy," in *Proc. Conference Record of 35th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 4–7, 2001, vol. 2, pp. 1004–1008.

[4] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1977.

[5] D. S. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[6] E. L. Blokh and V. V. Zyablov, "Coding of generalized concatenated codes," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 218–222, Jul. – Sep. 1974.

[7] V. Zyablov, S. Shavgulidze, and M. Bossert, "An introduction to generalized concatenated codes," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 609–622, Nov.-Dec. 1999.

[8] G. Foschini, "Layered space-time architecture for wireless communication in fading environment when using multi-element antennes," *Bell Labs Tech. J.*, vol. 1, no. , 2, pp. 41–59, 1996.

[9] A. Gorokhov and F. M. J. Willems, "Enhancing the capacity of layered MIMO transceivers via rate-matched encoding," in *Proc. 40th Allerton Conf. Commun., Contr. Comput.*, Oct. 1–3, 2002.