Correspondence

Successive Coding in Multiuser Information Theory

Xin Zhang, Jun Chen, *Member, IEEE*,
Stephen B. Wicker, *Senior Member, IEEE*, and
Toby Berger, *Fellow, IEEE*

***Abstract*—In this correspondence, we show that solutions to the multiple description coding problem and the broadcast channel coding problem share a common encoding procedure: successive source encoding. We use this connection as the basis for establishing connections between the achievable multiple description rate region and Marton's region for broadcast channels. Specifically, we show that Marton's encoding scheme can be viewed as a multiple description coding procedure. We also explore the dual problem, namely, the relationship between successive channel decoding in multiple access communication and distributed source coding. By illuminating these connections to multiple description, we hope to motivate a solution to what remains a mostly unsolved problem.**

***Index Terms*—Broadcast channel, dirty paper coding, distributed source coding, Gram–Schmidt orthogonalization, multiple access, multiple descriptions, successive coding.**

I. Introduction

Shannon was the first to point out the intimate connection between source and channel coding [1]. The existence of this duality has since been relentlessly mined with significant success. This connection has recently been extended to the multiuser information theory scenario. Some noteworthy demonstrations of this connection include the random coding level duality [2]–[5] and the operational duality [6] between the Wyner–Ziv problem [7] and Gel'fand–Pinsker problem [8]; the convex duality between Gaussian broadcast channels and multiaccess channels [9]–[11]; and the sum-rate duality between distributed source and channel coding problems [12], [13]. The recognition of these connections has not only deepened our understanding of these subjects, but has also lead to solutions to some long-standing open problems.

There remain branches of multiuser information theory that are still more or less isolated from the above connections, one such branch being the multiple description (MD) problem. The multiple description problem can be stated as follows: given a total available bit rate R and a pair of channels, both of which are subject to failure, how can one allocate rate and coded representations between the two channels such that if one channel fails, an adequate reconstruction of the source is possible, but if both channels are available, an improved reconstruction over the single-channel reception results? Some important contributions to the MD problem can be found in [14]–[17]. Recently the MD problem has been generalized to the L-channel case [18], [19].

Although the MD problem remains a largely unsolved problem, it does exhibit some interesting connections with other problems in mulituser information theory. In this correspondence, we show that the multiple description coding and broadcast channel coding share a common encoding procedure: successive source encoding. We also explore the dual problem, namely, successive channel decoding in multiple access communication and distributed source coding.

The remainder of this correspondence is divided into four sections. In Section II, we characterize the geometric structure of polymatroids and contra-polymatroids. In Section III we propose an achievable MD rate region and prove it is a contra-polymatroid. This region is shown to be achievable via a low-complexity successive source encoding scheme. For the Gaussian case, this successive source encoding scheme has a simple implementation via Gram–Schmidt orthogonalization. We establish some connections between this achievable MD rate region and Marton's region for broadcast channels. Specifically, Marton's encoding scheme can be viewed as a multiple description coding procedure. Dual results in distributed source coding and multiple access communication are presented in Section IV. We conclude the correspondence in Section V.

We use boldfaced letters to indicate (n-dimensional) vectors, capital letters for random objects, and small letters for their realizations. For example, we let $\mathbf{X} = (X(1), \ldots, X(n))^T$ and $\mathbf{x} = (x(1), \ldots, x(n))^T$. Calligraphic letters are used to indicate a set (say, $\mathcal{A}$). For any positive integer K, we define $\mathcal{I}_K = \{1, 2, \ldots, K\}$.

II. Polymatroid and Contra-Polymatroid

The polymatroid and the contra-polymatroid are two important geometric objects arising in various multiuser information theoretic problems. In this section we give a simple characterization of their geometric and combinatorial structures. These structures, when interpreted in the information theoretic context, have intrinsic connections with successive coding schemes.

Definition 1 ([20], [21]): Let $f : 2^{\mathcal{I}_L} \to \mathcal{R}_+$ be a set function. The polyhedron

$$\mathcal{B}(f) \triangleq \left\{ (x_1, \ldots, x_L) \in \mathcal{R}_+^L : \sum_{i \in \mathcal{S}} x_i \le f(\mathcal{S}), \forall \mathcal{S} \subseteq \mathcal{I}_L \right\}$$

is a polymatroid if the set function f satisfies

1) $f(\emptyset) = 0$ (normalized).
2) $f(\mathcal{S}) \le f(\mathcal{T})$ if $\mathcal{S} \subset \mathcal{T}$ (nondecreasing).
3) $f(\mathcal{S}) + f(\mathcal{T}) \ge f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T})$ (submodular).

The polyhedron

$$\mathcal{G}(f) \triangleq \left\{ (x_1, \ldots, x_L) : \sum_{i \in \mathcal{S}} x_i \ge f(\mathcal{S}), \forall \mathcal{S} \subseteq \mathcal{I}_L \right\}$$

Manuscript received June 2, 2005; revised August 22, 2006. The work of X. Zhang and S. B. Wicker was supported in part by the NSF Nets NOSS and ITR programs and the NSF TRUST Science and Technology Center. The work of J. Chen and T. Berger was supported in part by the National Science Foundation under Grant CCR-033 0059 and under a grant from the National Academies Keck Futures Initiative (NAKFI).

X. Zhang was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. She is now with Qualcomm, San Diego, CA 92121 USA (e-mail: xz56@cornell.edu).

J. Chen was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: jc353@cornell.edu).

S. B. Wicker is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: wicker@ece.cornell.edu).

T. Berger was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA (e-mail: tb6n@virginia.edu).

Communicated by V. A. Vaishampayan, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2007.896857

is a contra-polymatroid if f satisfies

1) $f(\emptyset) = 0$ (normalized).
2) $f(\mathcal{S}) \leq f(\mathcal{T})$ if $\mathcal{S} \subset \mathcal{T}$ (nondecreasing).
3) $f(\mathcal{S}) + f(\mathcal{T}) \leq f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T})$ (supermodular).

If f satisfies the three properties, f is called a rank function in both cases.

One of the most important properties of polymatroid and contra-polymatroid is that their vertices can be easily characterized. If π is a permutation on the set $\mathcal{I}_L$, define the vector $\mathbf{v}(\pi) \in \mathcal{R}^L$ by $v_{\pi(1)}(\pi) = f(\{\pi(1)\})$ and

$$v_{\pi(i)}(\pi) = f(\{\pi(1), \ldots, \pi(i)\}) - f(\{\pi(1), \ldots, \pi(i-1)\})$$

for $i = 2, 3, \ldots, L$.

Lemma 1 ([20], [21]): Let $\mathcal{B}(f)$ (or $\mathcal{G}(f)$) be a polymatroid (or contra-polymatroid). Then the points $\mathbf{v}(\pi)$ where π is a permutation on $\mathcal{I}_L$ are precisely the vertices of $\mathcal{B}(f)$ (or $\mathcal{G}(f)$).

Now we proceed to characterize the boundary of polymatroid $\mathcal{B}(f)$ and contra-polymatroid $\mathcal{G}(f)$. The boundary of $\mathcal{B}(f)$ (or $\mathcal{G}(f)$) is the union of all its faces. Specifically, each vertex is a 0-dimensional face. But for $\mathcal{B}(f)$, those trivial faces $\mathcal{B}(f) \cap \{(x_1, \ldots, x_L) : x_i = 0\}$, $i = 1, 2, \ldots, L$ are precluded.

For $\emptyset \subset \mathcal{S} \subseteq \mathcal{I}_L$, define the hyperplane

$$\mathcal{H}(\mathcal{S}) = \left\{ (x_1, \ldots, x_L) : \sum_{i \in \mathcal{S}} x_i = f(\mathcal{S}) \right\}.$$

Theorem 1: Let $\mathcal{B}(f)$ (or $\mathcal{G}(f)$) be a polymatroid (or contra-polymatroid). $\mathcal{F}$ is a face of $\mathcal{B}(f)$ (or $\mathcal{G}(f)$) if and only if $\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$ (or $\mathcal{F} = \mathcal{G}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$) for some $\emptyset \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_k \subseteq \mathcal{I}_L$.

Proof: See Appendix I. □

Definition 2: Let $\mathcal{D}_{\mathcal{B}(f)} = \mathcal{B}(f) \cap \mathcal{H}(\mathcal{I}_L)$. We call $\mathcal{D}_{\mathcal{B}(f)}$ the dominant face of $\mathcal{B}(f)$. Similarly, $\mathcal{D}_{\mathcal{G}(f)} = \mathcal{G}(f) \cap \mathcal{H}(\mathcal{I}_L)$ is referred to as the dominant face of $\mathcal{G}(f)$.

Among all the faces of $\mathcal{B}(f)$ (or $\mathcal{G}(f)$), $\mathcal{D}_{\mathcal{B}(f)}$ (or $\mathcal{D}_{\mathcal{G}(f)}$) is of special importance. For any vector $(x'_1, \ldots, x'_L) \in \mathcal{B}(f)$, there exists a vector $(x_1, \ldots, x_L) \in \mathcal{D}_{\mathcal{B}(f)}$ such that $(x_1, \ldots, x_L) \geq (x'_1, \ldots, x'_L)$. That is to say, every vector in $\mathcal{B}(f)$ is dominated (componentwise) by a vector in $\mathcal{D}_{\mathcal{B}(f)}$. Similarly, for any vector $(x'_1, \ldots, x'_L) \in \mathcal{G}(f)$, there exists a vector $(x_1, \ldots, x_L) \in \mathcal{D}_{\mathcal{G}(f)}$ such that $(x_1, \ldots, x_L) \leq (x'_1, \ldots, x'_L)$.

Now we give a finer characterization of these two dominant faces. It is easy to verify that

$$\sum_{i=1}^{L} v_{\pi(i)}(\pi) = f(\mathcal{I}_L).$$

So all the vertices are contained in $\mathcal{D}_{\mathcal{B}(f)}$ (or $\mathcal{D}_{\mathcal{G}(f)}$). By Theorem 1, $\mathcal{F}$ is a face of $\mathcal{D}_{\mathcal{B}(f)}$ (or $\mathcal{D}_{\mathcal{G}(f)}$) if and only if $\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$ (or $\mathcal{F} = \mathcal{G}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$) for some $\emptyset = \mathcal{S}_0 \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_k = \mathcal{I}_L$. Let Π be the set of permutation π on $\mathcal{I}_L$ such that $\{\pi(1), \ldots, \pi(|\mathcal{S}_i|)\} = \mathcal{S}_i$ for $i = 1, 2, \ldots, k-1$. We can verify that $\mathbf{v}(\pi)$ is a vertex of $\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$ (or $\mathcal{F} = \mathcal{G}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$) for each permutation $\pi \in \Pi$, and vice versa. Hence $\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$ (or $\mathcal{F} = \mathcal{G}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)$) has totally $|\Pi| = \prod_{j=1}^{k} (|\mathcal{S}_j \setminus \mathcal{S}_{j-1}|!)$ vertices.[1] Moreover, we have $\dim\left(\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)\right)$ (or $\dim\left(\mathcal{F} = \mathcal{G}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}(\mathcal{S}_i)\right)\right)$) $\leq L - k$, where the equality holds when all the vertices of $\mathcal{F}$ are distinct.

[1]These vertices may not be distinct.

III. Successive Source Encoding

In multiuser information theory, rate regions with polymatroid or contra-polymatroid structure often have intimate connections with successive coding schemes. We focus on the successive source encoding in this section, while the successive channel decoding is left to next section.

A. Successive Coding Function

We first prove the following invariant lemma before introducing the successive coding function.

Lemma 2: Let π be a permutation on $\mathcal{I}_k$ and let $(X_1, \ldots, X_k)$ be a set of k random variables. The sum

$$\sum_{i=1}^{k-1} I(Z_{\pi(1)}, \ldots, Z_{\pi(i)}; Z_{\pi(i+1)})$$

is invariant under π.

Proof: See Appendix II. □

Definition 3: For any set of random variables $\{Z_1, Z_2, \ldots, Z_k\}$, let

$$\psi(Z_i, i \in \mathcal{I}_k) = \sum_{i=1}^{k-1} I(Z_1, \ldots, Z_i; Z_{i+1}) \tag{1}$$

if $k \geq 2$ and let $\psi(Z_i, i \in \mathcal{I}_k) = 0$ if $k = 1$. We refer to $\psi(\cdot)$ as the successive coding function.

Remark: By Lemma 2, $\psi(Z_i, i \in \mathcal{I}_k)$ does not depend on the ordering in $\{Z_1, Z_2, \ldots, Z_k\}$ and thus is well-defined.

Lemma 3: The successive coding function $\psi(\cdot)$ has the following two properties:

1) $\psi(Z_i, i \in \mathcal{S}) \leq \psi(Z_i, i \in \mathcal{T})$ if $\mathcal{S} \subset \mathcal{T}$ (nondecreasing).
2) $$\psi(Z_i, i \in \mathcal{S}) + \psi(Z_i, i \in \mathcal{T}) \leq \psi(Z_i, i \in \mathcal{S} \cup \mathcal{T}) + \psi(Z_i, i \in \mathcal{S} \cap \mathcal{T})$$ (supermodular).

Proof: See Appendix III. □

We will show that the successive coding function $\psi(\cdot)$ arises naturally in both successive source encoding and successive channel decoding. Since Lemma 3 says that $\psi(\cdot)$ is essentially a rank function, it is not surprising that successive coding schemes give rise to rate regions with either polymatroid or contra-polymatroid structure.

B. Multiple Description Coding

Let $\{X(t)\}_{t=1}^{\infty}$ be an i.i.d. random process with $X(t) \sim p(x)$ for all t. Let $d(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ be a distortion measure.

Definition 4: We say that the rates $\{R_i, i \in \mathcal{I}_L\}$ and distortions $\{D_{\mathcal{A}}, \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L\}$ are achievable if for all $\epsilon > 0$, there exist, for n sufficiently large, encoding functions

$$f_i^{(n)} : \mathcal{X}^n \to \mathcal{C}_i^{(n)} \quad \log|\mathcal{C}_i^{(n)}| \leq n(R_i + \epsilon) \quad i = 1, \ldots, L$$

and decoding functions

$$g_{\mathcal{A}}^{(n)} : \prod_{i \in \mathcal{A}} \mathcal{C}_i^{(n)} \to \mathcal{X}^n \quad \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$$

such that for $\hat{\mathbf{X}}_{\mathcal{A}} = g_{\mathcal{A}}^{(n)}(f_i^{(n)}(\mathbf{X}), i \in \mathcal{A})$

$$\frac{1}{n} \mathbb{E} \sum_{t=1}^{n} d(X(t), \hat{X}_{\mathcal{A}}(t)) < D_{\mathcal{A}} + \epsilon, \quad \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L.$$

Theorem 2: $\{R_i, i \in \mathcal{I}_L, D_\mathcal{A}, \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L\}$ is achievable if there exist functions $g_\mathcal{A}(\cdot)$, $\emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$, and random variables $(U_1, \ldots, U_L)$ jointly distributed with the generic source variable X such that

$$\sum_{i\in\mathcal{A}} R_i \geq \psi(X, U_i, i \in \mathcal{A})$$
$$D_\mathcal{A} \geq \mathbb{E}(X, g_\mathcal{A}(U_i, i \in \mathcal{A}))$$

for all $\emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$.

Proof: It can be verified that the resulting achievable region is contained in the region proved in [18] and thus must be achievable. □

Our proposed region, although not the largest known achievable MD region, is of special importance. One can easily recover the region in [18] from ours by appending some superimposed refinement; one can also recover the region in [19] by incorporating the random binning procedure. Therefore, to certain extent, our region is a fundamental building block of all existing achievable MD rate-distortion regions.

Let

$$\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) = \left\{(R_1, \ldots, R_L) : \sum_{i\in\mathcal{S}} R_i \geq \psi(X, U_i, i \in \mathcal{S}),\ \forall \mathcal{S} \subseteq \mathcal{I}_L\right\}.$$

By the properties of $\psi(\cdot)$ proved in Lemma 3, it is easy to see that $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is a contra-polymatroid. By Lemma 1, $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ has $L!$ vertices. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $(R_1(\pi), \ldots, R_L(\pi))$ by

$$\begin{aligned} R_{\pi(1)}(\pi) &= \psi(X, U_{\pi(1)}) = I(X; U_{\pi(1)}), \\ R_{\pi(i)}(\pi) &= \psi(X, U_{\pi(1)}, \ldots, U_{\pi(i)}) - \psi(X, U_{\pi(1)} \ldots, U_{\pi(i-1)}) \\ &= I(X, U_{\pi(1)}, \ldots, U_{\pi(i-1)}; U_{\pi(i)}), \quad i = 2, \ldots, L. \end{aligned}$$

Then $(R_1(\pi), \ldots, R_L(\pi))$ is a vertex of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ for every permutation π.

The expressions of these vertices directly lead to the following successive source encoding scheme.

Successive Source Encoding for Vertex

1) Codebook Generation: Encoder $\pi(1)$ independently generates $2^{n[I(X;U_{\pi(1)})+\epsilon_{\pi(1)}]}$ codewords $\{\mathbf{U}_{\pi(1)}(j_{\pi(1)})\}_{j_{\pi(1)}=1}^{2^{n[I(X;U_{\pi(1)})+\epsilon_{\pi(1)}]}}$ according to the distribution $\prod p(u_{\pi(1)})$. Encoder $\pi(i)$ independently generates $2^{n[I(X,U_{\pi(1)},\ldots,U_{\pi(i-1)};U_{\pi(i)})+\epsilon_{\pi(i)}]}$ codewords $\{\mathbf{U}_{\pi(i)}(j_{\pi(i)})\}_{j_{\pi(i)}=1}^{2^{n[I(X,U_{\pi(1)},\ldots,U_{\pi(i-1)};U_{\pi(i)})+\epsilon_{\pi(i)}]}}$ according to the distribution $\prod p(u_{\pi(i)})$, $i = 2, \ldots, L$.
2) Encoding Procedure: Given $\mathbf{X}$, encoder $\pi(1)$ finds the codeword $\mathbf{U}_{\pi(1)}(j^*_{\pi(1)})$ such that $\mathbf{U}_{\pi(1)}(j^*_{\pi(1)})$ is strongly typical with $\mathbf{X}$, then encoder $\pi(i)$ finds the codeword $\mathbf{U}_{\pi(i)}(j^*_{\pi(i)})$ such that $\mathbf{U}_{\pi(i)}(j^*_{\pi(i)})$ is strongly typical with $(\mathbf{X}, \mathbf{U}_{\pi(1)}(j^*_{\pi(1)}), \ldots, \mathbf{U}_{\pi(i-1)}(j^*_{\pi(i-1)}))$, $i = 2, \ldots, L$. Index $j^*_{\pi(i)}$ is transmitted through channel $\pi(i)$, $i = 1, \ldots, L$.
3) Reconstruction: Decoder $\mathcal{A}$ reconstructs $\hat{\mathbf{X}}_\mathcal{A}$ with $\hat{X}_\mathcal{A}(t) = g_\mathcal{A}(U_i(j^*_i, t), i \in \mathcal{A})$, $\emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$. Here $U_i(j^*_i, t)$ is the tth entries of $\mathbf{U}_i(j^*_i)$, $t = 1, \ldots, n$, $i = 1, \ldots, L$.

For this scheme, encoder $\pi(1)$ does the encoding first, then encoder $\pi(2)$, encoder $\pi(3)$, and so on. Generally the design of encoder $\pi(i)$ becomes more complicated as i gets larger since the sequence space $(\mathbf{X}, \mathbf{U}_{\pi(1)}, \ldots, \mathbf{U}_{\pi(i-1)})$ its codebook needs to cover becomes larger. To simplify the encoder design, we can replace its input by a sufficient statistic. Specifically, let $V_{\pi(i)}$ be a sufficient statistic for estimating $U_{\pi(i+1)}$ from $(X, U_{\pi(1)}, \ldots, U_{\pi(i)})$, i.e.

$$(X, U_{\pi(1)}, \ldots, U_{\pi(i)}) \to V_{\pi(i)} \to U_{\pi(i+1)}$$

form a Markov chain, $i = 1, \ldots, L-1$. We can first map $(\mathbf{X}, \mathbf{U}_{\pi(1)}, \ldots, \mathbf{U}_{\pi(i)})$ to $\mathbf{V}_{\pi(i)}$ and let $\mathbf{V}_{\pi(i)}$ be the input of encoder $\pi(i+1)$, $i = 1, \ldots, L-1$. This can be justified by the Markov Lemma [22].

For the quadratic Gaussian case, we can let

$$V_{\pi(i)} = \mathbb{E}(U_{\pi(i+1)} | X, U_{\pi(1)}, \ldots, U_{\pi(i)})$$

which is the MMSE estimate of $U_{\pi(i+1)}$ given $(X, U_{\pi(1)}, \ldots, U_{\pi(i)})$, $i = 1, \ldots, L-1$. A closer look reveals that the above successive construction of sufficient statistics is nothing but Gram–Schmidt orthogonalization on $\{X, U_{\pi(1)}, \ldots, U_{\pi(L)}\}$. In the standard form, we have

$$\begin{aligned} &I_0 = X, \\ &I_1 = U_{\pi(1)} - \mathbb{E}(U_{\pi(1)}|X), \\ &I_i = U_{\pi(i)} - \mathbb{E}(U_{\pi(i)}|X, U_{\pi(1)}, \ldots, U_{\pi(i-1)}), \quad i = 2, \ldots, L. \end{aligned}$$

$\{I_0, \ldots, I_L\}$ are independent Gaussian random variables and sometimes referred to as the *innovation process*. For vertex $(R_1(\pi), \ldots, R_L(\pi))$ of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$, we have

$$\begin{aligned} R_{\pi(1)}(\pi) &= \psi(X, U_{\pi(1)}) = I(X; X + I_1) \\ R_{\pi(i)}(\pi) &= I(X, U_{\pi(1)}, \ldots, U_{\pi(i-1)}; U_{\pi(i)}) \\ &= I(\mathbb{E}(U_{\pi(i)}|X, U_{\pi(1)}, \ldots, U_{\pi(i-1)}); \\ &\quad \mathbb{E}(U_{\pi(i)}|X, U_{\pi(1)}, \ldots, U_{\pi(i-1)}) + I_i), \quad i = 2, \ldots, L. \end{aligned}$$

Intuitively, I_i can be viewed as the quantization error of encoder $\pi(i)$. The independence between quantization errors of different encoders significantly simplifies the design of MD quantization system. See [23] for a practical implementation using ECDQ.

Now we proceed to discuss the operational results associated with the contra-polymatroid struture of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$. For $\emptyset \subset \mathcal{S} \subseteq \mathcal{I}_L$, define the hyperplane

$$\mathcal{H}_{MD}(\mathcal{S}) = \left\{(R_1, \ldots, R_L) : \sum_{i\in\mathcal{S}} R_i = \psi(X, U_i, i \in \mathcal{S})\right\}.$$

Let

$$\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L) = \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) \cap \mathcal{H}_{MD}(\mathcal{I}_L)$$

which is the dominant face of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$. Every rate tuple $(R'_1, \ldots, R'_L)$ strictly inside $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is dominated (componentwise) by a rate tuple $(R_1, \ldots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ in terms of compression efficiency. Hence in search of the optimal scheme, the attention can be restricted to rate pairs on the dominant face without loss of generality.

It has been shown that every vertex of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ is achievable via an L-step successive source encoding scheme. We now extend this result to the rate tuples on the boundary of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$. From the discussion in the preceding section, we know that $\mathcal{F}$ is a face of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ if and only if $\mathcal{F} = \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) \cap \left(\bigcap_{i=1}^k \mathcal{H}_{MD}(\mathcal{S}_i)\right)$ for some $\emptyset = \mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_k = \mathcal{I}_L$. For any rate tuple $(R_1, \ldots, R_L) \in \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) \cap \left(\bigcap_{i=1}^k \mathcal{H}_{MD}(\mathcal{S}_i)\right)$, we have

$$\begin{aligned} \sum_{m\in\mathcal{S}_j\setminus\mathcal{S}_{j-1}} R_m &= \psi(X, U_i, i \in \mathcal{S}_j) - \psi(X, U_i, i \in \mathcal{S}_{j-1}) \\ &= \psi(X_{j-1}, U_i, i \in \mathcal{S}_j \setminus \mathcal{S}_{j-1}) \end{aligned} \tag{2}$$

where $X_{j-1} = (X, U_i, i \in \mathcal{S}_{j-1})$, $j = 1, \ldots, k$. Note: X_{j-1} should be viewed as a single random variable when expand $\psi(X_{j-1}, U_i, i \in \mathcal{S}_j \setminus \mathcal{S}_{j-1})$ according to (1). From (2), it is clear that $(R_1, \ldots, R_L)$ can be achieved via a k-step successive group source encoding scheme.

That is, at step j, encoders in group $\mathcal{S}_j \setminus \mathcal{S}_{j-1}$ jointly search for codewords $\mathbf{U}_i$ ($i \in \mathcal{S}_j \setminus \mathcal{S}_{j-1}$) such that $(\mathbf{U}_i, i \in \mathcal{S}_j \setminus \mathcal{S}_{j-1})$ are jointly typical with $(\mathbf{X}, \mathbf{U}_i, i \in \mathcal{S}_{j-1})$, $j = 1, \ldots, k$. So the successive encoding is implemented between different groups while the joint encoding is implemented within each group.

Now we propose a splitting method to achieve a general point in $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ via successive source encoding. Let $\mathcal{U} = \{U_{1,1}, \ldots, U_{1,m_1}, U_{2,1}, \ldots, U_{2,m_2}, \ldots, U_{L,1}, U_{L,m_L}\}$ jointly distributed with the generic source variables X such that

1) $\sum_{i=1}^{L} m_i \le 2L - 1$ and $1 \le m_i \le 2$ for all $i \in \mathcal{I}_L$;
2) $U_{i,m_i} \to U_{i,1} \to (X, \{U_{j,1}, \ldots, U_{j,m_j}\}_{j \ne i})$ form a Markov chain for all $i \in \mathcal{I}_L$;
3) $(X, U_1, \ldots, U_L) = (X, U_{1,1}, \ldots, U_{L,1})$ in distribution.

If we view U_i as a description of X, then $U_{i,1}$ is an identical copy of U_i and $U_{i,2}$ (if exists) is a coarse description of X. We can say that $U_{i,2}$ is *split* from U_i. Let $\mathcal{U}_\sigma$ be a permutation on $\mathcal{U}$ such that for all $i \in \mathcal{I}_L$, $U_{i,2}$ (if exists) is placed before $U_{i,1}$ with at least one random variable between them (we refer this type of permutation as the well-ordered permutation). Let $\{U_{i,j}\}_\sigma^-$ denote all the random variables that appear before $U_{i,j}$ in $\mathcal{U}_\sigma$.

Successive Source Encoding with Splitting:

1) Codebook Generation: For all $i \in \mathcal{I}_L$, if $m_i = 1$, then encoder i independently generates $2^{n[I(X,\{U_{i,1}\}_\sigma^-; U_{i,1}) + \epsilon_{i,1}]}$ codewords $\{\mathbf{U}_{i,1}(j_{i,1})\}_{j_{i,1}=1}^{2^{n[I(X,\{U_{i,1}\}_\sigma^-; U_{i,1}) + \epsilon_{i,1}]}}$ according to the distribution $\prod p(u_{i,1})$; if $m_i = 2$, encoder i first independently generates $2^{n[I(X,\{U_{i,2}\}_\sigma^-; U_{i,2}) + \epsilon_{i,2}]}$ codewords $\{\mathbf{U}_{i,2}(j_{i,2})\}_{j_{i,2}=1}^{2^{n[I(X,\{U_{i,2}\}_\sigma^-; U_{i,2}) + \epsilon_{i,2}]}}$ according to the distribution $\prod p(u_{i,2})$, then for each codeword $\mathbf{U}_{i,2}(j_{i,2})$, encoder i independently generates $2^{n[I(X,\{U_{i,1}\}_\sigma^-; U_{i,1}|U_{i,2}) + \epsilon_{i,1}]}$ codewords $\{\mathbf{U}_2(j_{i,2}, j_{i,1})\}_{j_{i,1}=1}^{2^{n[I(X,\{U_{i,1}\}_\sigma^-; U_{i,1}|U_{i,2}) + \epsilon_{i,1}]}}$ according to the conditional distribution $\prod_t p(u_{i,1}|U_{i,2}(j_{i,2}, t))$. Here $U_{i,2}(j_{i,2}, t)$ is the tth entry of $\mathbf{U}_{i,2}(j_{i,2})$.
2) Encoding Procedure: Encoding is carried out according to the ordering in $\mathcal{U}_\sigma$. Given $\mathbf{X}$, if $m_i = 1$, then encoder i finds the codeword $\mathbf{U}_{i,1}(j_{i,1}^*)$ such that $\mathbf{U}_{i,1}(j_{i,1}^*)$ is strongly typical with $(\mathbf{X}, \{\mathbf{U}_{i,1}(j_{i,1}^*)\}_\sigma^-)$; if $m_i = 2$, encoder i first finds the codeword $\mathbf{U}_{i,2}(j_{i,2}^*)$ such that $\mathbf{U}_{i,2}(j_{i,2}^*)$ is strongly typical with $(\mathbf{X}, \{\mathbf{U}_{i,2}(j_{i,2}^*)\}_\sigma^-)$, then encoder i finds the codeword $\mathbf{U}_{i,1}(j_{i,2}^*, j_{i,1}^*)$ such that $\mathbf{U}_{i,1}(j_{i,2}^*, j_{i,1}^*)$ is strongly typical with $(\mathbf{X}, \{\mathbf{U}_{i,1}(j_{i,2}^*, j_{i,1}^*)\}_\sigma^-)$. Index $j_{i,1}^*$ and $j_{i,2}^*$ (if exists) are transmitted through channel i, $i = 1, \ldots, L$.
3) Reconstruction: For all $\mathcal{A} \subseteq \mathcal{I}_L$, decoder $\mathcal{A}$ reconstructs $\hat{\mathbf{X}}_\mathcal{A}$ with $\mathbf{U}_{i,1}(j_{i,2}^*, j_{i,1}^*)$ (or $\mathbf{U}_{i,1}(j_{i,1}^*)$ if $m_i = 1$), $i \in \mathcal{A}$.

It can be shown via an argument similar to that in [24] that a general point on the dominant face of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is achievable via a successive source encoding scheme of at most $2L - 1$ steps if the splitting is used, and each U_i gets split at most once. It is noteworthy that $2L - 1$ is just an upper bound. For rate tuples on the boundary of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$, the encoding steps can be further reduced. Intuitively, the splitting method can viewed as a way to create an order for $\{U_1, \ldots, U_L\}$. We have already known that rate tuples on the boundary of dominant face are achievable via successive group source encoding, i.e., the order exists between different groups but not within each group. So we just need to use the splitting method to create order within each group. Specifically, for any rate tuple $(R_1, \ldots, R_L)$ in

$$\mathcal{F} = \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) \cap \left(\bigcap_{i=1}^{k} \mathcal{H}_{MD}(\mathcal{S}_i) \right)$$

for some $\emptyset = \mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_k = \mathcal{I}_L$, the splitting can be carried out in $\mathcal{S}_i \setminus \mathcal{S}_{i-1}$, $i = 1, 2, \ldots, k$ separately. So the number of successive encoding steps is $L + \dim(\mathcal{F}) \le \sum_{i=1}^{k} (2|\mathcal{S}_i \setminus \mathcal{S}_{i-1}| - 1) = 2L - k$.

Similar to the vertex case, we can successively construct sufficient statistics along the order created by the splitting method to simplify the encoder design. This procedure becomes Gram–Schmidt orthogonalization in the quadratic Gaussian case. Let $\mathcal{U}_\sigma$ be a well-ordered permutation on $\mathcal{U}$. Applying Gram–Schmidt orthogonalization to $\mathcal{U}_\sigma$ yields

$$I_i' = U_\sigma(i) - \mathbb{E}(U_\sigma(i)|X, \{U_\sigma(i)\}_\sigma^-)$$

for $i = 1, \ldots, |\mathcal{U}|$, where $U_\sigma(i)$ is the ith element of $\mathcal{U}_\sigma$. For $i \in \mathcal{I}_L$, if $m_i = 1$ and $U_\sigma(j) = U_{i,1}$, then

$$\begin{aligned} R_i &= I(X, \{U_{i,1}\}_\sigma^-; U_{i,1}) \\ &= I(\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-); \mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) + I_j'); \end{aligned}$$

if $m_i = 2$, $U_\sigma(k) = U_{i,2}$ and $U_\sigma(l) = U_{i,1}$, then

$$\begin{aligned} R_i &= I(X, \{U_{i,2}\}_\sigma^-; U_{i,2}) + I(X, \{U_{i,1}\}_\sigma^-; U_{i,1}|U_{i,2}) \\ &= I(\mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-); \mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-) + I_k') \\ &\quad + I(\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-); \mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) + I_l'|U_{i,2}) \\ &= I(\mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-); \mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-) + I_k') \\ &\quad + I(\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) - \mathbb{E}[\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-)|U_{i,2}]; \\ &\quad \mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) - \mathbb{E}[\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-)|U_{i,2}] + I_l') \\ &= I(\mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-); \mathbb{E}(U_{i,2}|X, \{U_{i,2}\}_\sigma^-) + I_k') \\ &\quad + I(\mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) - \mathbb{E}(U_{i,1}|U_{i,2}); \\ &\quad \mathbb{E}(U_{i,1}|X, \{U_{i,1}\}_\sigma^-) - \mathbb{E}(U_{i,1}|U_{i,2}) + I_l'). \end{aligned}$$

C. Broadcast Channel Coding

It is well-known [6], [25] that the encoding part of the Gel'fand–Pinsker scheme [8] is essentially a source encoding. Now we extend this result to the broadcast channel. The largest known achievable rate region for the broadcast channel is by Marton [26]. Since the corner points of Marton's region are achievable via successive Gel'fand–Pinsker coding [27], it is not surprsing that the encoding scheme for the broadcast channel can be interpreted as successive source encoding.

Theorem 3 (Marton's Region [26]): The rate tuple $(R_1, \ldots, R_L)$ is achievable for discrete memoryless broadcast channel $p(y_1, \ldots, y_L|x)$ if there exist random variables $(W_1, \ldots, W_L)$ with $(W_1, \ldots, W_L) \to X \to (Y_1, \ldots, Y_L)$ forming a Markov chain such that

$$\sum_{i \in \mathcal{S}} R_i \le \sum_{i \in \mathcal{S}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{S}), \quad \forall \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L.$$

Let

$$\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L) = \left\{ (R_1, \ldots, R_L) \in \mathcal{R}_+^L : \sum_{i \in \mathcal{S}} R_i \le \sum_{i \in \mathcal{S}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{S}),\ \forall \emptyset \subset \mathcal{S} \subseteq \mathcal{I}_L \right\}.$$

It may seem natural to expect that $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ is a polymatroid. Although the "submodular" property of $\sum_{i \in \mathcal{S}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{S})$ follows directly from the "supermodular" property of $\psi(\cdot)$, it turns out that the "monotone" property, which amounts to requiring $I(W_i; Y_i) \ge I(W_i; W_j, j \ne i)$ for all i, is unguaranteed. Nevertheless, for the illustrative purpose, we assume that $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ is a polymatroid. Therefore, it follows from

Lemma 1 that $\mathcal{R}_{MD}(W_i, i \in \mathcal{I}_L)$ has $L!$ vertices. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $(R_1(\pi), \ldots, R_L(\pi))$ by

$$\begin{aligned} R_{\pi(1)}(\pi) &= I(W_{\pi(1)}; Y_{\pi(1)}), \\ R_{\pi(i)}(\pi) &= \sum_{j=1}^{i} I(W_{\pi(j)}; Y_{\pi(j)}) - \psi(W_{\pi(1)}, \ldots, W_{\pi(i)}) \\ &\quad - \sum_{j=1}^{i-1} I(W_{\pi(j)}; Y_{\pi(j)}) + \psi(W_{\pi(1)}, \ldots, W_{\pi(i-1)}) \\ &= I(W_{\pi(i)}; Y_{\pi(i)}) - I(W_{\pi(1)}, \ldots, W_{\pi(i-1)}; W_{\pi(i)}), \\ &\qquad i = 2, \ldots, L. \end{aligned}$$

Then $(R_1(\pi), \ldots, R_L(\pi))$ is a vertex of $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ for every permutation π. These vertices are achievable via L-step successive Gel'fand–Pinsker coding.

Successive Gel'fand–Pinsker Coding for Vertex $(R_1(\pi), \ldots, R_L(\pi))$

1) Codebook Generation: There are L bin arrays, with $2^{n(R_i(\pi)-\epsilon_{\pi(i)})}$ bins in the ith array. First generate $2^{n(R_{\pi(1)}(\pi)-\epsilon_{\pi(1)})} = 2^{n(I(W_{\pi(1)};Y_{\pi(1)})-\epsilon_{\pi(1)})}$ codewords $\{\mathbf{W}_{\pi(1)}(j_{\pi(1)})\}_{j_{\pi(1)}=1}^{2^{n[I(W_{\pi(1)};Y_{\pi(1)})-\epsilon_{\pi(1)}]}}$ independently according to the distribution $\prod p(w_{\pi(1)})$, and distribute them into the bins in the $\pi(1)$th array such that each bin contains exactly one codeword. Then independently generate $2^{n(I(W_{\pi(i)};Y_{\pi(i)})+\epsilon'_{\pi(i)})}$ codewords $\{\mathbf{W}_{\pi(i)}(j_{\pi(i)})\}_{j_{\pi(i)}=1}^{2^{n[I(W_{\pi(i)};Y_{\pi(i)})+\epsilon'_{\pi(i)}]}}$ according to the distribution $\prod p(w_{\pi(i)})$, and distribute them uniformly into the bins in the $\pi(i)$th array, $i = 2, \ldots, L$.
2) Encoding Procedure: Given L independent messages $M_1, M_2, \ldots, M_L$, where M_i is the message for receiver i, $i = 1, 2, \ldots, L$, first find, in the $\pi(1)$th array, the codeword $\mathbf{W}_{\pi(1)}(j^*_{\pi(1)})$ in the bin with index $M_{\pi(1)}$; then successively search in the $\pi(i)$th array for the codeword $\mathbf{W}_{\pi(i)}(j^*_{\pi(i)})$ in the bin with index $M_{\pi(i)}$ such that $\mathbf{W}_{\pi(i)}(j^*_{\pi(i)})$ is strongly typical with $(\mathbf{W}_{\pi(1)}(j^*_{\pi(1)}), \ldots, \mathbf{W}_{\pi(i-1)}(j^*_{\pi(i-1)}))$, $i = 2, \ldots, L$. Finally encoder converts $(\mathbf{W}_1(j^*_1), \ldots, \mathbf{W}_L(j^*_L))$ into $\mathbf{X}$ according to $\prod p(x|w_1, \ldots, w_L)$, and transmits $\mathbf{X}$ to the receivers.
3) Decoding Procedure: Given $\mathbf{Y}_i$, receiver i searches for the codeword $\mathbf{W}_i$ such that $\mathbf{Y}_i$ and $\mathbf{W}_i$ are jointly typical, and declares the index of the bin containing $\mathbf{W}_i$ as the transmitted message.

It is straightforward to see from the above coding scheme that each bin in the $\pi(i)$th array contains roughly $2^{[nI(W_{\pi(1)},\ldots,W_{\pi(i-1)};W_{\pi(i)})]}$ codewords, $i = 2, \ldots, L$, and the encoding procedure is essentially a successive source encoding in the order $\pi(1) \to \pi(2) \to \cdots \to \pi(L)$. In other words, pick an arbitrary bin from the $\pi(i)$th array, $i = 2, \ldots, L$, then these $L-1$ bin together form an $(L-1)$-MD code if we view $W_{\pi(1)}$ as the the source. Therefore, many results regarding to the multiple description problem are also applicable here. For example, we can successively construct sufficient statistics to reduce the encoding complexity, which becomes Gram–Schmidt orthogonalization in the Gaussian case. We use the renowned Costa's dirty paper coding [28] (which is a special case of Gel'fand–Pinsker coding) to illustrate this point.

Consider the Gaussian broadcast channel $Y_i = X + Z_i$ with the power constraint $\mathbb{E}X^2 \le P$, where $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, \ldots, L$. Without loss of generality, we assume $N_1 \ge N_2 \ge \cdots \ge N_L$. Following Costa's construction, we let

$$W_1 = X_1 \tag{3}$$

$$W_i = \alpha_{i-1} \sum_{j=1}^{i-1} X_j + X_i, \quad i = 2, \ldots, L \tag{4}$$

where $X_1, \ldots, X_L$ are zero-mean, independent Gaussian random variables with $\mathbb{E}X_i^2 = P_i$ and $\sum_{i=1}^{L} P_i = P$. Furthermore, let $\alpha_i = P_{i+1}/(N_{i+1} + \sum_{j=i+2}^{L} P_j)$, $i = 1, \ldots, L-1$, and $X = \sum_{i=1}^{L} X_i$. By this construction, the following rate region

$$\begin{aligned} R_i &\le I(W_i; Y_i) - I(W_1, \ldots, W_{i-1}; W_i) \\ &= \frac{1}{2} \log \left(1 + \frac{P_i}{N_i + \sum_{j=i+1}^{L} P_j} \right), \quad i = 1, \ldots, L \end{aligned} \tag{5}$$

is achievable. It was shown by Bergmans [29] that by varying P_i ($i = 1, \ldots, L$) under the constraint $\sum_{i=1}^{L} P_i = P$, the rate region given in (5) is exactly the capacity region of the Gaussian broadcast channel. Writing (3) and (4) in the following equivalent form:

$$\begin{aligned} X_1 &= W_1, \\ X_i &= W_i - \alpha_{i-1} \sum_{j=1}^{i-1} X_j \\ &= W_i - \mathbb{E}(W_i | W_1, \ldots, W_{i-1}), \quad i = 2, \ldots, L \end{aligned}$$

we can see that it is exactly Gram–Schmidt orthogonalization on $(W_1, W_2, \ldots, W_L)$.

Returning to the general case, one can easily derive the operational results associated with the polymatroid struture by imitating the approach used in the multiple description problem. Also, the splitting method can be used to achieve general points on the dominant face via successive coding. Of course, all these are based on the assumption that $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ is a polymatroid.

IV. Successive Channel Decoding

We have studied the properties of $\psi(\cdot)$ and its consequences in the successive source encoding scenario. We show in this section that $\psi(\cdot)$ also arises naturally in the successive channel decoding scenario.

A. Multiple-Access Channel Coding

For the memoryless multiaccess channel $p(y|x_1, \ldots, x_L)$ with the fixed input distribution $\prod_{i=1}^{L} p(x_i)$, the following rate region is achievable

$$\mathcal{R}_{MC}(X_1, \ldots, X_L) = \left\{ (R_1, \ldots, R_L) \in \mathcal{R}_+^L : \sum_{i \in \mathcal{S}} R_i \le I(X_i, i \in \mathcal{S}; Y | X_i, i \in \mathcal{S}^c), \forall \emptyset \subset \mathcal{S} \subseteq \mathcal{I}_L \right\}.$$

The following lemma follows from the fact that $X_1, \ldots, X_L$ are independent.

Lemma 4: $I(X_i, i \in \mathcal{S}; Y) = \psi(Y, X_i, i \in \mathcal{S})$ for all $\emptyset \subset \mathcal{S} \subseteq \mathcal{I}_L$.

Proof: Without loss of generality, we assume $\mathcal{S} = \{1, 2, \ldots, k\}$.

$$\begin{aligned} &I(X_i, i \in \mathcal{S}; Y) \\ &= I(X_1; Y) + \sum_{i=2}^{k} I(X_i; Y | X_1, \ldots, X_{i-1}) \\ &= I(X_1; Y) + \sum_{i=2}^{k} [I(X_i; Y, X_1, \cdots, X_{i-1}) - I(X_i; X_1, \cdots, X_{i-1})] \\ &= I(X_1; Y) + \sum_{i=2}^{k} I(X_i; Y, X_1, \ldots, X_{i-1}) \\ &= \psi(Y, X_i, i \in \mathcal{S}). \end{aligned}$$

□

By Lemma 4 we can write

$$I(X_i, i\in\mathcal{S}; Y|X_i, i\in\mathcal{S}^c)=I(X_i, i\in\mathcal{I}_L; Y)-I(X_i, i\in\mathcal{S}^c; Y)$$
$$=I(X_i, i\in\mathcal{I}_L; Y)-\psi(Y, X_i, i\in\mathcal{S}^c).$$

Then it follows directly from Lemma 3 that $\mathcal{R}_{MC}(X_1,\ldots,X_L)$ is a polymatroid. Therefore, the rate tuple $(R_1(\pi),\ldots,R_L(\pi))$ defined by

$$R_{\pi(i)}(\pi)$$
$$=\psi(Y, X_{\pi(j)}, j=i,\cdots,L)-\psi(Y, X_{\pi(j)}, j=i+1,\ldots,L)$$
$$= I(X_{\pi(i)}; Y, X_{\pi(i+1)},\ldots,X_{\pi(L)}),\quad i=1,\ldots,L-1,$$
$$R_{\pi(L)}(\pi)$$
$$= I(X_{\pi(L)}; Y)$$

is a vertex of $\mathcal{R}_{MC}(X_1,\ldots,X_L)$ for every permutation π on $\mathcal{I}_L$. It is well-known that these vertices are achievable via successive channel decoding. Specifically, we first decode $\mathbf{X}_{\pi(L)}$, then decode $\mathbf{X}_{\pi(L-1)}$ using $\mathbf{X}_{\pi(L)}$ as the side information, and so on. An equivalent but more illuminating interpretation is as follows.

1) At the first step, view $\mathbf{X}_{\pi(L)}$ as the input, which generates the output $\mathbf{Y}$ through a memoryless channel $p(y|x_{\pi(L)})$.
2) At the ith decoding step, view $\mathbf{X}_{\pi(L-i+1)}$ as the input, which generates the output $(\mathbf{Y}, \mathbf{X}_{\pi(L-i+2)},\ldots,\mathbf{X}_{\pi(L)})$ through a single-input–multiple-output (SIMO) channel with transition probability

$$p(y, x_{\pi(L-i+2)},\ldots,x_{\pi(L)}|x_{\pi(L-i+1)})$$
$$= p(y|x_{\pi(L-i+1)},\ldots,x_{\pi(L)})\prod_{j=L-i+2}^{L} p(x_{\pi(j)}),\ i=2,\ldots,L.$$

Here

$$p(y|x_{\pi(i)}, x_{\pi(i+1)},\ldots,x_{\pi(L)})$$
$$=\sum_{x_{\pi(1)},\ldots,x_{\pi(i-1)}}(p(y|x_1,\ldots,x_L)\prod_{j=1}^{i-1}p(x_{\pi(j)})),\ i=2,\ldots,L.$$

So each step is just a single-user channel decoding. Similar to the successive source encoding scenario, at each step we can replace the channel output by a sufficient statistic to reduce the decoding complexity. Let's consider the Gaussian multiaccess channel $Y=\sum_{i=1}^{L}X_i+N$, where $X_1,\ldots,X_L$, and N are zero-mean, independent Gaussian random variables. Applying Gram–Schmidt orthogonalization to $\{Y, X_{\pi(L)}, X_{\pi(L-1)},\ldots,X_{\pi(1)}\}$ yields

$$I_0=Y$$
$$I_1=X_{\pi(L)}-\mathbb{E}(X_{\pi(L)}|Y)=X_{\pi(L)}-\frac{\mathbb{E}X_{\pi(L)}^2}{EN^2+\sum_{j=1}^{L}\mathbb{E}X_j^2}Y$$
$$I_i=X_{\pi(L-i+1)}-\mathbb{E}(X_{\pi(L-i+1)}|Y, X_{\pi(L-i+2)},\ldots,X_{\pi(L)})$$
$$=X_{\pi(L-i+1)}-\frac{\mathbb{E}X_{\pi(L-i+1)}^2}{\mathbb{E}N^2+\sum_{j=1}^{L-i+1}\mathbb{E}X_{\pi(j)}^2}\left(Y-\sum_{j=L-i+2}^{L}X_{\pi(j)}\right)$$
$$i=2,\ldots,L.$$

This is exactly the classic successive cancellation procedure.

The polymatroid structure of $\mathcal{R}_{MC}(X_1,\ldots,X_L)$ as well as the associated operational meaning has been characterized in [21], [30]. See [24], [31]–[33] for the discussion of the splitting method in multiaccess communication.

B. Distributed Source Coding

The general achievable rate region for the distributed lossy source coding problem is sometimes referred to as the Berger-Tung region [22], [34]. Specifically, let $V_1,\ldots,V_L$ be the auxiliary random variables jointly distributed with the generic source variables $X_1,\ldots,X_L$ such that $V_i\to X_i\to(X_j, V_j, j\neq i)$ form a Markov chain for all $i=1,\ldots,L$. The Berger-Tung region is defined as

$$\mathcal{R}_{DS}(V_1,\ldots,V_L)=\left\{(R_1,\ldots,R_L)\in\mathcal{R}^L\right.$$
$$\left.:\sum_{i\in\mathcal{S}}R_i\geq I(X_i, i\in\mathcal{S}; V_i, i\in\mathcal{S}|V_i, i\in\mathcal{S}^c), \forall\emptyset\subset\mathcal{S}\subseteq\mathcal{I}_L\right\}.$$

Lemma 5: If $V_i\to X_i\to(X_j, V_j, j\neq i)$ form a Markov chain for all $i=1,\ldots,L$, then

$$I(X_i, i\in\mathcal{S}; V_i, i\in\mathcal{S}|V_i, i\in\mathcal{S}^c)$$
$$=\sum_{i\in\mathcal{S}}I(X_i; V_i)-\psi(V_i, i\in\mathcal{I}_L)+\psi(V_i, i\in\mathcal{S}^c).$$

Proof: Without loss of generality, we assume $\mathcal{S}=\{1,\ldots,k\}$. See the derivative at the bottom of the page, where (6) follows from the fact that $(X_i, V_i, i\neq k-j)\to X_{k-j}\to V_{k-j}$ form a Markov chain. □

Theorem 4: $\mathcal{R}_{DS}(V_1,\ldots,V_L)$ is a contra-polymatroid.

Proof: By Lemma 5, the "supermodular" property of $I(X_i, i\in\mathcal{S}; V_i, i\in\mathcal{S}|V_i, i\in\mathcal{S}^c)$ follows directly from the "supermodular" property of $\psi(\cdot)$. In order to satisfy the "monotone" property, by Lemma 5 we need $I(X_i, V_i)\geq I(V_i; V_j, j\neq i)$ for all i. This is true because $V_i\to X_i\to(V_j, j\neq i)$ form a Markov chain for all i. □

By Theorem 4 and Lemma 1, the rate tuple $(R_1(\pi),\ldots,R_L(\pi))$ defined by

$$R_{\pi(i)}(\pi)$$
$$=I(X_{\pi(i)}; V_{\pi(i)})+\psi(V_{\pi i+1},\ldots,V_{\pi(L)})-\psi(V_{\pi i},\ldots,V_{\pi(L)})$$
$$=I(X_{\pi(i)}; V_{\pi(i)})-I(V_{\pi(i)}; V_{\pi(i+1)},\cdots,V_{\pi(L)}),\ i=1,\cdots,L-1,$$
$$R_{\pi(L)}(\pi)$$
$$=I(X_{\pi(L)}; V_{\pi(L)})$$

$$\begin{aligned}I(X_i, i\in\mathcal{S}; V_i, i\in\mathcal{S}|V_i, i\in\mathcal{S}^c)&=\sum_{j=0}^{k-1}I(X_i, i\in\mathcal{S}; V_{k-j}|V_{k-j+1},\ldots,V_L)\\&=\sum_{j=0}^{k-1}[I(V_{k-j+1},\ldots,V_L, X_i, i\in\mathcal{S}; V_{k-j})-I(V_{k-j}; V_{k-j+1},\ldots,V_L)]\\&=\sum_{j=0}^{k-1}[I(X_{k-j}; V_{k-j})-I(V_{k-j}; V_{k-j+1},\ldots,V_L)]\\&=\sum_{i\in\mathcal{S}}I(X_i; V_i)-\psi(V_i, i\in\mathcal{I}_L)+\psi(V_i, i\in\mathcal{S}^c)\end{aligned}\tag{6}$$

is a vertex of $\mathcal{R}_{DS}(V_1,\ldots,V_L)$ for every permutation π on $\mathcal{I}_L$. These vertices are achievable via successive Wyner–Ziv coding.

Successive Wyner-Ziv Coding for Vertex $(R_1(\pi),\ldots,R_L(\pi))$

1) Codebook Generation: Encoder $\pi(L)$ independently generates $2^{n[I(X_{\pi(L)};V_{\pi(L)})+\epsilon_{\pi(L)}]}$ codewords $\{\mathbf{V}_{\pi(L)}(j_{\pi(L)})\}_{j_{\pi(L)}=1}^{2^{n[I(X_{\pi(L)};V_{\pi(L)})+\epsilon_{\pi(L)}]}}$ according to the distribution $\prod p(v_{\pi(L)})$, and distributes them into $2^{n[I(X_{\pi(L)};V_{\pi(L)})+\epsilon_{\pi(L)}]}$ bins such that each bin contains exactly one codeword. Encoder $\pi(i)$ independently generates $2^{n[I(X_{\pi(i)};V_{\pi(i)})+\epsilon_{\pi(i)}]}$ codewords $\{\mathbf{V}_{\pi(i)}(j_{\pi(i)})\}_{j_{\pi(i)}=1}^{2^{n[I(X_{\pi(i)};V_{\pi(i)})+\epsilon_{\pi(i)}]}}$ according to the distribution $\prod p(v_{\pi(i)})$, and uniformly distributes them into $2^{n[R_{\pi(i)}(\pi)+\epsilon_{\pi(i)}]}$ bins, $i=1,\ldots,L-1$.
2) Encoding Procedure: Given $\mathbf{X}_i$, encoder i searches for the codeword $\mathbf{V}_i(j_i^*)$ such that $\mathbf{V}_i(j_i^*)$ is strongly typical with $\mathbf{X}_i$. The index of the bin containing $\mathbf{V}_i(j_i^*)$, say b_i^*, is sent to the decoder.
3) Decoding Procedure: Decoder first decodes the codeword $\mathbf{V}_{\pi(L)}$ in the bin $b_{\pi(L)}^*$ of encoder $\pi(L)$. Clearly, $\mathbf{V}_{\pi(L)}=\mathbf{V}_{\pi(L)}(j_{\pi(L)}^*)$. Then the decoder successively searches in the bin $b_{\pi(L-i)}^*$ of encoder $\pi(L-i)$ for the codeword $\mathbf{V}_{\pi(L-i)}$ such that $\mathbf{V}_{\pi(L-i)}$ is strongly typical with $(\mathbf{V}_{\pi(L-i+1)},\ldots,\mathbf{V}_{\pi(L)})$, $i=1,\ldots,L-1$.

We can view the above decoding procedure as successive channel decoding. There are exactly one codeword in each bin of encoder $\pi(L)$, and roughly $2^{nI(V_{\pi(i)};V_{\pi(i+1)},\ldots,V_{\pi(L)})}$ codewords in each bin of encoder $\pi(i), i=1,\ldots,L-1$. We first decode $\mathbf{V}_{\pi(L)}(j_{\pi(L)}^*)$ unambiguously; then regard the codewords in bin $b_{\pi(L-1)}^*$ of encoder $\pi(L-1)$ as a channel codebook, and $\mathbf{V}_{\pi(L)}(j_{\pi(L)}^*)$ as the channel output generated by $\mathbf{V}_{\pi(L-1)}(j_{\pi(L-1)}^*)$ through the channel $p(v_{\pi(L)}|v_{\pi(L-1)})$. Therefore, recovering $\mathbf{V}_{\pi(L-1)}(j_{\pi(L-1)}^*)$ based on $\mathbf{V}_{\pi(L)}(j_{\pi(L)}^*)$ is a channel decoding operation. Similarly, for $i=L-2,L-3,\ldots,1$, viewing the codewords in bin $b_{\pi(i)}^*$ of encoder $\pi(i)$ as a channel codebook, and $(\mathbf{V}_{\pi(i+1)}(j_{\pi(i+1)}^*),\ldots,\mathbf{V}_{\pi(L)}(j_{\pi(L)}^*))$ as the channel output generated by $\mathbf{V}_{\pi(i)}(j_{\pi(i)}^*)$ through the SIMO channel $p(v_{\pi(i+1)},\ldots,v_{\pi(L)}|v_{\pi(i)})$, we can successively decode $\mathbf{V}_{\pi(L-2)}(j_{\pi(L-2)}^*),\mathbf{V}_{\pi(L-3)}(j_{\pi(L-3)}^*),\ldots,\mathbf{V}_{\pi(1)}(j_{\pi(1)}^*)$ via channel decoding. Similar to the multiaccess channel case, at each decoding step, we can replace the channel output by a sufficient statistic to simplify the decoding complexity, i.e., replace $(V_{\pi(i+1)},\ldots,V_{\pi(L)})$ by $\varphi(V_{\pi(i+1)},\ldots,V_{\pi(L)})$ with the property that $(V_{\pi(i+1)},\ldots,V_{\pi(L)})\rightarrow\varphi(V_{\pi(i+1)},\ldots,V_{\pi(L)})\rightarrow V_{\pi(i)}$ form a Markov chain, $i=1,\ldots,L-1$. In the quadratic Gaussian case, we can let $\varphi(V_{\pi(i+1)},\ldots,V_{\pi(L)})=\mathbb{E}(V_{\pi(i)}|V_{\pi(i+1)},\ldots,V_{\pi(L)})$, which is essentially Gram–Schmidt orthogonalization on $\{V_{\pi(L)},V_{\pi(L-1)},\ldots,V_{\pi(1)}\}$. The contra-polymatroid structure of $\mathcal{R}_{DS}(V_1,\ldots,V_L)$ was first observed in [12], [35]. See [38] for the characterization of the dominant face of $\mathcal{R}_{DS}(V_1,\ldots,V_L)$, and [25], [36]–[38] for the application of the splitting method in distributed source coding.

V. Conclusion

We synthesized many classic results in multiuser information theory with an emphasis on the underlying successive coding schemes. Successive coding is practically important because it reduces a complicated multiuser coding problem into sequences of low-complexity single-user source encoding or channel decoding problems. Although a direct successive coding order may not exist for general points on the dominant face, we can use the splitting method to create a coding order as long as the rate region possesses a polymatroid or contra-polymatroid structure. Once such an order is given, sufficient statistics can be successively constructed along this order to reduce the source encoding complexity or the channel decoding complexity. In the Gaussian case, this successive construction of sufficient statistics becomes Gram–Schmidt orthogonalization.

Appendix I
Proof of *Theorem* 1

We only prove the results regarding polymatroid. The proof for contra-polymatroid is completely analogous.

The "if" part is straightforward. We only need to check that for every $\emptyset\subset\mathcal{S}_1\subset\mathcal{S}_2\subset\cdots\mathcal{S}_k\subseteq\mathcal{I}_L$, $\mathcal{F}=\mathcal{B}(f)\cap\left(\bigcap_{i=1}^k\mathcal{H}(\mathcal{S}_i)\right)$ is nonempty. This can be verified by noticing that for every permutation π on $\mathcal{I}_L$ satisfying $\{\pi(1),\pi(2),\ldots,\pi(|\mathcal{S}_i|)\}=\mathcal{S}_i$, $i=1,2,\ldots,k$, the vertex $\mathbf{v}(\pi)$ is contained in $\mathcal{F}$.

Now we proceed to prove the "only if" part. For $\emptyset\subset\mathcal{S},\mathcal{T}\subseteq\mathcal{I}_L$, by the "submodular" property of f we only need to consider the following two cases.

1) $f(\mathcal{S})+f(\mathcal{T})>f(\mathcal{S}\cup\mathcal{T})+f(\mathcal{S}\cap\mathcal{T})$: Suppose there exists a vector $(x_1,\ldots,x_L)\in\mathcal{B}(f)$ such that $\sum_{i\in\mathcal{S}}x_i=f(\mathcal{S})$ and $\sum_{i\in\mathcal{T}}x_i=f(\mathcal{T})$. We have
$$\begin{aligned}\sum_{i\in\mathcal{S}}x_i+\sum_{i\in\mathcal{T}}x_i&=f(\mathcal{S})+f(\mathcal{T})\\&>f(\mathcal{S}\cup\mathcal{T})+f(\mathcal{S}\cap\mathcal{T})\\&\geq\sum_{i\in\mathcal{S}\cup\mathcal{T}}x_i+\sum_{i\in\mathcal{S}\cap\mathcal{T}}x_i\\&=\sum_{i\in\mathcal{S}}x_i+\sum_{i\in\mathcal{T}}x_i\end{aligned}$$
which results in a contradiction. Hence $\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S})\cap\mathcal{H}(\mathcal{T})=\emptyset$.
2) $f(\mathcal{S})+f(\mathcal{T})=f(\mathcal{S}\cup\mathcal{T})+f(\mathcal{S}\cap\mathcal{T})$: Suppose there exists a vector $(x_1,\ldots,x_L)\in\mathcal{B}(f)$ such that $\sum_{i\in\mathcal{S}}x_i=f(\mathcal{S})$ and $\sum_{i\in\mathcal{T}}x_i=f(\mathcal{T})$. Since
$$\begin{aligned}\sum_{i\in\mathcal{S}}x_i+\sum_{i\in\mathcal{T}}x_i&=f(\mathcal{S})+f(\mathcal{T})\\&=f(\mathcal{S}\cup\mathcal{T})+f(\mathcal{S}\cap\mathcal{T})\\&\geq\sum_{i\in\mathcal{S}\cup\mathcal{T}}x_i+\sum_{i\in\mathcal{S}\cap\mathcal{T}}x_i\\&=\sum_{i\in\mathcal{S}}x_i+\sum_{i\in\mathcal{T}}x_i\end{aligned}$$
it implies $\sum_{i\in\mathcal{S}\cup\mathcal{T}}x_i=f(\mathcal{S}\cup\mathcal{T})$ and $\sum_{i\in\mathcal{S}\cap\mathcal{T}}x_i=f(\mathcal{S}\cap\mathcal{T})$. So we have $\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S})\cap\mathcal{H}(\mathcal{T})\subseteq\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S}\cup\mathcal{T})\cap\mathcal{H}(\mathcal{S}\cap\mathcal{T})$. It can be shown using the same method that the converse is also true, i.e., $\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S}\cup\mathcal{T})\cap\mathcal{H}(\mathcal{S}\cap\mathcal{T})\subseteq\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S})\cap\mathcal{H}(\mathcal{T})$. Therefore, $\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S})\cap\mathcal{H}(\mathcal{T})=\mathcal{B}(f)\cap\mathcal{H}(\mathcal{S}\cup\mathcal{T})\cap\mathcal{H}(\mathcal{S}\cap\mathcal{T})$.

From 1) and 2), we can see that for $\emptyset\subset\mathcal{A}_1\subset\mathcal{A}_2\subset\cdots\subset\mathcal{A}_i\subseteq\mathcal{I}_L$ and $\emptyset\subset\mathcal{T}\subseteq\mathcal{I}_L$, we have either
$$\mathcal{B}(f)\cap\left(\bigcap_{j=1}^i\mathcal{H}(\mathcal{A}_j)\right)\cap\mathcal{H}(\mathcal{T})=\emptyset$$
or
$$\begin{aligned}&\mathcal{B}(f)\cap\left(\bigcap_{j=1}^i\mathcal{H}(\mathcal{A}_j)\right)\cap\mathcal{H}(\mathcal{T})\\&=\mathcal{B}(f)\cap\left(\bigcap_{j=1}^i(\mathcal{H}(\mathcal{A}_j)\cap\mathcal{H}(\mathcal{T}))\right)\\&=\mathcal{B}(f)\cap\left(\bigcap_{j=1}^i(\mathcal{H}(\mathcal{A}_j\cup\mathcal{T})\cap\mathcal{H}(\mathcal{A}_j\cap\mathcal{T}))\right)\\&=\mathcal{B}(f)\cap\mathcal{H}(\mathcal{A}_1\cap\mathcal{T})\cap\cdots\cap\mathcal{H}(\mathcal{A}_i\cap\mathcal{T})\cap\mathcal{H}(\mathcal{A}_1\cup\mathcal{T})\\&\quad\cap\cdots\cap\mathcal{H}(\mathcal{A}_i\cup\mathcal{T}).\end{aligned}\tag{7}$$
Notice in (7) we have $(\mathcal{A}_1\cap\mathcal{T})\subseteq\cdots\subseteq(\mathcal{A}_i\cap\mathcal{T})\subseteq(\mathcal{A}_1\cup\mathcal{T})\subseteq\cdots\subseteq(\mathcal{A}_i\cup\mathcal{T})$. Therefore, for any face $\mathcal{F}=\mathcal{B}(f)\cap\left(\bigcap_{j=1}^l\mathcal{T}_j\right)$,

by successively applying (7) we can write $\mathcal{F}$ in the form $\mathcal{F} = \mathcal{B}(f) \cap \left(\bigcap_{i=1}^{k} \mathcal{S}_i\right)$ for some $\emptyset \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_k \subseteq \mathcal{I}_L$.

Appendix II
Proof of Lemma 2

Since any permutation can be decomposed into a sequence of operations that exchange two adjacent positions, we only need to show that the sum is invariant under permutation $\pi = (1, \ldots, i, i+1, \ldots, k)$ and permutation $\pi' = (1, \ldots, i-1, i+1, i, i+2, \ldots, k)$, which further boils down to show

$$I(Z_1, \ldots, Z_{i-1}; Z_i) + I(Z_1, \ldots, Z_i; Z_{i+1}) = I(Z_1, \ldots, Z_{i-1}; Z_{i+1}) + I(Z_1, \ldots, Z_{i-1}, Z_{i+1}; Z_i).$$

This is true because

$$\begin{aligned} &I(Z_1, \ldots, Z_{i-1}; Z_{i+1}) + I(Z_1, \ldots, Z_{i-1}, Z_{i+1}; Z_i) \\ &= I(Z_1, \ldots, Z_{i-1}; Z_{i+1}) + I(Z_1, \ldots, Z_{i-1}; Z_i) \\ &\quad + I(Z_{i+1}; Z_i | Z_1, \ldots, Z_{i-1}) \\ &= I(Z_1, \ldots, Z_{i-1}; Z_i) + I(Z_1, \ldots, Z_i; Z_{i+1}) \end{aligned}$$

which completes the proof.

When the entropy (or differential entropy) exists for all the random variables and random vectors of interest, the following proof is more straightforward. Since

$$\begin{aligned} &I(Z_{\pi(1)}, \ldots, Z_{\pi(i)}; Z_{\pi(i+1)}) \\ &\quad = H(Z_{\pi(1)}, \ldots, Z_{\pi(i)}) + H(Z_{\pi(i+1)}) - H(Z_{\pi(1)}, \ldots, Z_{\pi(i+1)}), \end{aligned}$$

it follows that

$$\sum_{i=1}^{k-1} I(Z_{\pi(1)}, \ldots, Z_{\pi(i)}; Z_{\pi(i+1)}) = \sum_{i=1}^{k} H(Z_i) - H(Z_i, i \in \mathcal{I}_k),$$

which clearly does not depend on π.

Appendix III
Proof of Lemma 3

The "nondecreasing" property is obvious. So we only need to verify the "supermodular" property. For any $\mathcal{S}, \mathcal{T} \subseteq \mathcal{I}_k$, suppose $\mathcal{S} = \{i_1, \ldots, i_l, i_{l+1}, \ldots, i_m\}$, $\mathcal{T} = \{i_{l+1}, \ldots, i_m, i_{m+1}, \ldots, i_n\}$, and $\mathcal{S} \cap \mathcal{T} = \{i_{l+1}, \ldots, i_m\}$. We have

$$\begin{aligned} &\psi(Z_i, i \in \mathcal{S}) + \psi(Z_i, i \in \mathcal{T}) \\ &= \sum_{j=1}^{m-1} I(Z_{i_1}, \ldots, Z_{i_j}; Z_{i_{j+1}}) + \sum_{j=l+1}^{n-1} I(Z_{i_{l+1}}, \ldots, Z_{i_j}; Z_{i_{j+1}}) \\ &= \sum_{j=1}^{m-1} I(Z_{i_1}, \ldots, Z_{i_j}; Z_{i_{j+1}}) + \psi(Z_i, i \in \mathcal{S} \cap \mathcal{T}) \\ &\quad + \sum_{j=m}^{n-1} I(Z_{i_{l+1}}, \ldots, Z_{i_j}; Z_{i_{j+1}}) \\ &\leq \sum_{j=1}^{m-1} I(Z_{i_1}, \ldots, Z_{i_j}; Z_{i_{j+1}}) + \psi(Z_i, i \in \mathcal{S} \cap \mathcal{T}) \\ &\quad + \sum_{j=m}^{n-1} I(Z_{i_1}, \ldots, Z_{i_j}; Z_{i_{j+1}}) \\ &= \psi(Z_i, i \in \mathcal{S} \cap \mathcal{T}) + \psi(Z_i, i \in \mathcal{S} \cup \mathcal{T}) \end{aligned}$$

which completes the proof.

Remark: Suppose the entropy (or differential entropy) exists for all the random variables and random vectors of interest. We can write $\psi(Z_i, i \in \mathcal{S}) = \sum_{i \in \mathcal{S}} H(Z_i) - H(Z_i, i \in \mathcal{S})$. Then the "supermodular" property of $\psi(\cdot)$ is equivalent to the "submodular" property of $H(\cdot)$.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Proc. IRE Nat. Conv. Rec.*, Mar. 1959, pp. 142–163.

[2] J. K. Su, J. J. Eggers, and B. Girod, "Illustration of the duality between channel coding and rate distortion with side information," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2000.

[3] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1629–1638, Jun. 2002.

[4] B. Chen and G. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1159–1180, May 2003.

[5] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1181–1203, May 2003.

[6] H. Wang and P. Viswanath, "Fixed binning schemes for channel and source coding problems: An operational duality," *IEEE Trans. Inf. Theory*, submitted for publication.

[7] A. D. Wyner and J. Ziv, "The rate distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[8] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[9] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

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

[11] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1875–1892, Sep. 2004.

[12] P. Viswanath, "Sum rate of a class of Gaussian multiterminal source coding problems," in *Advances in Network Information Theory*, P. Gupta, G. Kramer, and A. Wijngaarden, Eds. Providence, RI: DIMACS, American Mathematical Society, 2004, pp. 43–60.

[13] W. Yu, "Duality and the value of cooperation in distributive source and channel coding problems," in *Proc. 41st Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2003.

[14] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[15] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[16] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. IT-31, pp. 721–726, Nov. 1985.

[17] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-33, pp. 502–521, Jul. 1987.

[18] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 2106–2114, Sep. 2003.

[19] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions-part I: (n, k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, pp. 47–61, Jan. 2004.

[20] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Combinatorial Structures and their Applications*, R. Guy, H. Hanani, N. Sauer, and J. Schonheim, Eds. New York: Gordon and Breach, 1970, pp. 69–87.

[21] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels-part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.

[22] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978, vol. 229, CISM Courses and Lectures, pp. 171–231.

[23] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5197–5217, Dec. 2006.
[24] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, pp. 873–890, Mar. 2001.
[25] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1250–1276, Jun. 2002.
[26] K. Marton, "A coding theorem for the discrete memoryless broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
[27] G. Caire and S. Shamai, "On the achievable throughput in multantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, Jul. 2003.
[28] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 439–441, May 1983.
[29] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, pp. 279–280, Mar. 1974.
[30] B. Rimoldi and R. Urbanke, "On the structure of the dominant face of multiple access channels," in *Proc. Inf. Theory Commun. Workshop*, South Africa, Jun. 20–25, 1999, pp. 12–14.
[31] A. B. Carleial, "On the Capacity of Multiple-Terminal Communication Networks," Ph.D., Stanford Univ., Stanford, CA, 1975.
[32] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 364–375, Mar. 1996.
[33] B. Rimoldi, "Generalized time sharing: A low-complexity capacity-achieving multiple-access technique," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2432–2442, Sep. 2001.
[34] S. Y. Tung, "Multiterminal Source Coding," Ph.D., Cornell Univ., School of Electrical Engineering, Ithaca, NY, 1978.
[35] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 977–987, Aug. 2004.
[36] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting," in *Proc. IEEE Int. Symp. Inf. Theory*, Ulm, Germany, Jun.–Jul. 29–4, 1997, p. 271.
[37] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3546–3561, Aug. 2006.
[38] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its application to the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, submitted for publication.

Information Rates Subject to State Masking

Neri Merhav, *Fellow, IEEE*, and
Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*—We consider the problem of rate–R channel coding with causal/noncausal side information at the transmitter, under an additional requirement of minimizing the amount of information that can be learned from the channel output about the state sequence, which is defined in terms of the mutual information between the state sequence and the channel output sequence. A single-letter characterization is provided for the achievable region of pairs $\{(R, E)\}$. Explicit results for the Gaussian case (Costa's dirty-paper channel) are derived in full detail.**

***Index Terms*—Binning, causal side information, equivocation, dirty-paper channel, Gel'fand–Pinsker channel, noncausal side information, secrecy.**

I. Introduction

The problem of information transfer via state–dependent channels is classical (see [14] for a partial review). One of the most interesting models is the case where the channel states are available at the transmitter either causally or noncausally. This framework has been fully characterized for independent and identically distributed (i.i.d.) states in famous studies by Shannon [19] and by Gel'fand and Pinsker (G-P) [9], repectively. These models, and in particular the G-P setting, have gained much interest in the last few years, mainly due to the wide scope application areas, such as watermarking [3], [15], [17], [20], [16], multiple–input multiple–output (MIMO) broadcast channels [1], [2], [13], and cooperative networks [11], just to name a few applications.

One of the most interesting and well-known examples is the G-P channel is the Gaussian setting where the states impact the channel additively. The surprising result by Costa [4] demonstrates that no loss in capacity is suffered no matter how strong that independent interfering state sequence is. Evidently, the many applications and the challenge here motivated much work in terms of actual coding strategies that come close to the optimum. These coding strategies (see, e.g., [26] and references therein), build on the insight of random binning which is the central mechanism in showing achievability in this problem [9], and can, in fact, be interpreted as practical binning strategies. In the Gaussian channel, nicknamed "dirty paper" [4], efficient techniques based on modern codes were recently reported as well (see [8], [21], and references therein). Source—channel coding aspects in the framework of state–dependent channel of this type are also considered [18], and the source–channel separation principle has been shown valid in various scenarios, in which the model itself is intimately related to the Wyner–Ziv (W-Z) source coding problem with side information at the decoder [25], and the G-P channel [9].

While in models addressed in [18] the source and channel states are assumed independent, this is not always the case. In some applications, the channel–state process is not inherently channel–related (like in fading), but may rather be an information–bearing signal on its own.

Manuscript received March 26, 2006; revised December 20, 2006. This research was supported by the Israel Science Foundation. The material in this correspondence was presented at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

The authors are with the Electrical Engineering Department, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il; sshlomo@ee.technion.ac.il).

Communicated by G. Kramer, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2007.896860