Lecture 5: Multiple Descriptions

Jun Chen

1 Achievable Multiple Description Region:

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

Definition 1.1. *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:*

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

and decoding functions:

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

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

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

We shall prove the following consistency lemma before presenting the achievable MD rate-distortion region.

Lemma 1.1. *Let π be a permutation on $\mathcal{I}_k$ and let $(Z_1, \cdots, Z_k)$ be a set of k random variables. The sum*

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

is invariant under π.

Proof. Since any permutation can be decomposed into a sequence of operations that exchange two adjacent positions, we only need to show that under permutation $\pi = (1, \cdots, i, i+1, \cdots, k)$ and permutation $\pi' = (1, \cdots, i-1, i+1, i, i+2, \cdots, k)$, the values of (1) are the same. This further boils down to show

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

This is indeed true because

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

which completes the proof. □

Definition 1.2. *For any set of random variables $\{Z_1, Z_2, \cdots, Z_k\}$, let*

$$\psi(Z_1, \cdots, Z_k) = \sum_{i=1}^{k-1} I(Z_1, \cdots, Z_i; Z_{i+1})$$

if $k \geq 2$ and let $\psi(Z_1, \cdots, Z_k) = 0$ if $k = 1$. Remark: By Lemma 1.1, $\psi(Z_1, \cdots, Z_k)$ does not depend on the ordering in $\{Z_1, Z_2, \cdots, Z_k\}$ and thus is well-defined.

Theorem 1.1. *$\{R_i, i \in \mathcal{I}_L, D_{\mathcal{A}}, \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L\}$ is achievable if there exist functions $g_{\mathcal{A}}(\cdot)$, $\emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$, and random variables $(U_1, \cdots, U_L)$ jointly distributed with the generic source variable X such that*

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

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

Let

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

Now we proceed to study the geometric structure of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$, which sheds light on the associated coding schemes.

Theorem 1.2. *$\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is a contra-polymatroid.*

Proof. The "normalized" property and "nondecreasing" property are obviously satisfied. So we only need to verify the "supermodular" property. For any $\mathcal{S}, \mathcal{T} \subseteq \mathcal{I}_L$, suppose $\mathcal{S} = \{i_1, \cdots, i_j, i_{j+1}, \cdots, i_k\}$, $\mathcal{T} = \{i_{j+1}, \cdots, i_k, i_{k+1}, \cdots, i_l\}$, and $\mathcal{S} \cap \mathcal{T} = \{i_{j+1}, \cdots, i_k\}$. We have

$$\begin{aligned} & \psi(X, U_i, i \in \mathcal{S}) + \psi(X, U_i, i \in \mathcal{T}) \\ = \;& I(X; U_{i_1}) + \sum_{m=1}^{k-1} I(X, U_{i_1}, \cdots, U_{i_m}; U_{i_{m+1}}) + (X; U_{i_{j+1}}) + \sum_{m=j+1}^{l-1} I(X, U_{i_{j+1}}, \cdots, U_{i_m}; U_{i_{m+1}}) \\ = \;& I(X; U_{i_1}) + \sum_{m=1}^{k-1} I(X, U_{i_1}, \cdots, U_{i_m}; U_{i_{m+1}}) + \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T}) + \sum_{m=k}^{l-1} I(X, U_{i_{j+1}}, \cdots, U_{i_m}; U_{i_{m+1}}) \\ \leq \;& I(X; U_{i_1}) + \sum_{m=1}^{k-1} I(X, U_{i_1}, \cdots, U_{i_m}; U_{i_{m+1}}) + \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T}) + \sum_{m=k}^{l-1} I(X, U_{i_1}, \cdots, U_{i_m}; U_{i_{m+1}}) \\ = \;& \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T}) + \psi(X, U_i, i \in \mathcal{S} \cup \mathcal{T}), \end{aligned}$$

which completes the proof. □

Since $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is a contra-polymatroid, it must have $L!$ vertices. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $(R_1(\pi), \cdots, R_L(\pi))$ by

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

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

2 Successive Quantization and Quantization Splitting:

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

Successive Quantization for Vertex $(R_1(\pi), \cdots, R_L(\pi))$**:**

1. Codebook Generation: Encoder $\pi(1)$ independently generates $2^{n[I(X;U_{\pi(1)})+\epsilon_{\pi(1)}]}$ codewords $\{\mathbf{U}_{\pi(1)}(j_{\pi(1)})\}_{j_{\pi(1)}=1}^{2^{n[I(X;U_{\pi(1)})+\epsilon_1]}}$ according to the distribution $\prod p(u_{\pi(1)})$. Encoder $\pi(i)$ independently generates $2^{n[I(X,U_{\pi(1)},\cdots,U_{\pi(i-1)};U_{\pi(i)})+\epsilon_{\pi(i)}]}$ codewords $\{\mathbf{U}_{\pi(i)}(j_{\pi(i)})\}_{j_{\pi(i)}=1}^{2^{n[I(X,U_{\pi(1)},\cdots,U_{\pi(i-1)};U_{\pi(i)})+\epsilon_{\pi(i)}]}}$ according to the distribution $\prod p(u_{\pi(i)})$, $i = 2, \cdots, L$.

2. Encoding Procedure: Given $\mathbf{X}$, encoder $\pi(1)$ finds the codeword $\mathbf{U}_{\pi(1)}(j^*)$ 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 $(\mathbf{X}, \mathbf{U}_{\pi(1)}(j^*_{\pi(1)}), \cdots, \mathbf{U}_{\pi(1)}(j^*_{\pi(i-1)}))$, $i = 2, \cdots, L$. Index $j^*_{\pi(i)}$ is transmitted through channel i, $i = 1, \cdots, L$.

3. Reconstruction: Decoder $\mathcal{A}$ constructs $\hat{\mathbf{X}}_{\mathcal{A}}$ 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 t-th entries of $\mathbf{U}_i(j^*_i)$, $t = 1, \cdots, n$, $i = 1, \cdots, L$.

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 $(\mathbf{X}, \mathbf{U}_{\pi(1)}, \cdots, \mathbf{U}_{\pi(i-1)})$ its codebook needs to cover becomes larger. Intuitively, we can view each encoder as a quantizer. But it is not necessarily a conventional quantizers because its input (except that of quantizer $\pi(1)$) consists of multiple sequences. If there exists a sequence of functions $\varphi_1(\cdot), \cdots, \varphi_{L-1}(\cdot)$[1] such that

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

form a Markov chain, where $V_{\pi(i)} = \varphi_{\pi(i)}(X, U_{\pi(1)}, \cdots, U_{\pi(i)})$, $i = 1, \cdots, L-1$, then we can first map $(\mathbf{X}, \mathbf{U}_{\pi(1)}, \cdots, \mathbf{U}_{\pi(i)})$ to $\mathbf{V}_{\pi(i)}$ and let $\mathbf{V}_{\pi(i)}$ be the input of encoder $\pi(i+1)$, $i = 1, \cdots, L-1$. This can be justified by the Markov Lemma.

It is easy to verify that $\sum_{i=1}^{L} R_i(\pi) = \psi(X, U_i, i \in \mathcal{I}_L)$, i.e., the sum-rate $\psi(X, U_i, i \in \mathcal{I}_L)$ is achieved at every vertex. So the set

$$\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L) = \left\{ (R_1, \cdots, R_L) \in \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L) : \sum_{i=1}^{L} R_i = \psi(X, U_i, i \in \mathcal{I}_L) \right\}$$

is non-empty and clearly $dim\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L) \le L-1$. We call $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ the dominant face of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ since any rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ has the property that

$$(R'_1, \cdots, R'_L) \le (R_1, \cdots, R_L) \Rightarrow (R'_1, \cdots, R'_L) = (R_1, \cdots, R_L)$$

for all $(R'_1, \cdots, R'_L) \in \mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$. That is to say, any rate tuple $(R'_1, \cdots, R'_L)$ strictly inside $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is dominated by some rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ in terms of compression efficiency. Hence in searching for the optimal scheme, the attention can be restricted to rate pairs on the dominant face without loss of generality.

[1] These functions actually always exist. For example, we can let them be bijective functions.

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 coding scheme. We now extend this result to the rate tuples on the boundary of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$. Note: The boundary of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ is the union of all its faces (except $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ itself). Specifically, each vertex is a 0-dimensional face.

Again, characterizing the boundary of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ relies heavily on its contra-polymatroid structure. For $\emptyset \subset \mathcal{A} \subset \mathcal{I}_L$, define the hyperplane

$$\mathcal{H}(\mathcal{A}) = \left\{(R_1, \cdots, R_L) : \sum_{i \in \mathcal{A}} R_i = \psi(X, U_i, i \in \mathcal{A})\right\}$$

and let $\mathcal{F}_{\mathcal{A}} = \mathcal{H}(\mathcal{A}) \cap \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$. It is easy to verify that for any rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$, we have

$$\begin{aligned}
\psi(X, U_i, i \in \mathcal{A}) &\leq \sum_{i \in \mathcal{A}} R_i = \sum_{i=1}^{L} R_i - \sum_{i \in \mathcal{A}^c} R_i \\
&= \psi(X, U_i, i \in \mathcal{I}_L) - \sum_{i \in \mathcal{A}^c} R_i \\
&\leq \psi(X, U_i, i \in \mathcal{I}_L) - \psi(X, U_i, i \in \mathcal{A}^c).
\end{aligned}$$

So for any rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$, if $\sum_{i \in \mathcal{A}} R_i = \psi(X, U_i, i \in \mathcal{A})$, then $(R_1, \cdots, R_L) \in \mathcal{F}_{\mathcal{A}}$; if $\sum_{i \in \mathcal{A}} R_i = \psi(X, U_i, i \in \mathcal{I}_L) - \psi(X, U_i, i \in \mathcal{A}^c)$, then $(R_1, \cdots, R_L) \in \mathcal{F}_{\mathcal{A}^c}$. Therefore, every face of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ is of the form $\bigcap_j \mathcal{F}_{\mathcal{A}_j}$, where $\emptyset \subset \mathcal{A}_j \subset \mathcal{I}_L$ for all j. For $\emptyset \subset \mathcal{S}, \mathcal{T} \subset \mathcal{I}_L$ such that $\mathcal{S} \nsubseteq \mathcal{T}$ and $\mathcal{S} \nsupseteq \mathcal{T}$, by the "supermodular" property of $\psi(\cdot)$ we only need to consider the following two cases.

1. $\psi(X, U_i, i \in \mathcal{S}) + \psi(X, U_i, i \in \mathcal{T}) < \psi(X, U_i, i \in \mathcal{S} \cup \mathcal{T}) + \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T})$: Suppose there exists a rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ such that $\sum_{i \in \mathcal{S}} R_i = \psi(X, U_i, i \in \mathcal{S})$ and $\sum_{i \in \mathcal{T}} R_i = \psi(X, U_i, i \in \mathcal{T})$. We have

$$\begin{aligned}
& \sum_{i \in \mathcal{S}} R_i + \sum_{i \in \mathcal{T}} R_i \\
= \; & \psi(X, U_i, i \in \mathcal{S}) + \psi(X, U_i, i \in \mathcal{T}) \\
< \; & \psi(X, U_i, i \in \mathcal{S} \cup \mathcal{T}) + \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T}) \\
\leq \; & \sum_{i \in \mathcal{S} \cup \mathcal{T}} R_i + \sum_{i \in \mathcal{S} \cap \mathcal{T}} R_i \\
= \; & \sum_{i \in \mathcal{S}} R_i + \sum_{i \in \mathcal{T}} R_i,
\end{aligned}$$

   which results in a contradiction. Hence $\mathcal{F}_{\mathcal{S}} \cap \mathcal{F}_{\mathcal{T}} = \emptyset$.

2. $\psi(X, U_i, i \in \mathcal{S}) + \psi(X, U_i, i \in \mathcal{T}) = \psi(X, U_i, i \in \mathcal{S} \cup \mathcal{T}) + \psi(X, U_i, i \in \mathcal{S} \cap \mathcal{T})$: Suppose there exists a rate tuple $(R_1, \cdots, R_L) \in \mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ such that $\sum_{i \in \mathcal{S}} R_i = \psi(X, U_i, i \in \mathcal{S})$

and $\sum_{i\in\mathcal{T}} R_i = \psi(X, U_i, i \in \mathcal{T})$. Since

$$\begin{aligned}
& \sum_{i\in\mathcal{S}} R_i + \sum_{i\in\mathcal{T}} R_i \\
= \; & \psi(X, U_i, i \in \mathcal{S}) + \psi(X, U_i, i \in \mathcal{T}) \\
= \; & \psi(X, U_i, i \in \mathcal{S}\cup\mathcal{T}) + \psi(X, U_i, i \in \mathcal{S}\cap\mathcal{T}) \\
\leq \; & \sum_{i\in\mathcal{S}\cup\mathcal{T}} R_i + \sum_{i\in\mathcal{S}\cap\mathcal{T}} R_i \\
= \; & \sum_{i\in\mathcal{S}} R_i + \sum_{i\in\mathcal{T}} R_i,
\end{aligned}$$

it implies $\sum_{i\in\mathcal{S}\cup\mathcal{T}} R_i = \psi(X, U_i, i \in \mathcal{S}\cup\mathcal{T})$ and $\sum_{i\in\mathcal{S}\cap\mathcal{T}} R_i = \psi(X, U_i, i \in \mathcal{S}\cap\mathcal{T})$. So we have $\mathcal{F}_{\mathcal{S}} \cap \mathcal{F}_{\mathcal{T}} \subseteq \mathcal{F}_{\mathcal{S}\cup\mathcal{T}} \cap \mathcal{F}_{\mathcal{S}\cap\mathcal{T}}$. It can be shown using the same method that the converse is also true, i.e., $\mathcal{F}_{\mathcal{S}\cup\mathcal{T}} \cap \mathcal{F}_{\mathcal{S}\cap\mathcal{T}} \subseteq \mathcal{F}_{\mathcal{S}} \cap \mathcal{F}_{\mathcal{T}}$. Therefore, $\mathcal{F}_{\mathcal{S}} \cap \mathcal{F}_{\mathcal{T}} = \mathcal{F}_{\mathcal{S}\cup\mathcal{T}} \cap \mathcal{F}_{\mathcal{S}\cap\mathcal{T}}$.

From 1) and 2), we can see that for $\emptyset \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_k \subset \mathcal{I}_L$ and $\emptyset \subset \mathcal{T} \subset \mathcal{I}_L$, we have either

$$\left(\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{S}_j}\right) \cap \mathcal{F}_{\mathcal{T}} = \emptyset,$$

or

$$\begin{aligned}
& \left(\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{S}_j}\right) \cap \mathcal{F}_{\mathcal{T}} \\
= \; & \bigcap_{j=1}^{k} (\mathcal{F}_{\mathcal{S}_j} \cap \mathcal{F}_{\mathcal{T}}) \\
= \; & \bigcap_{j=1}^{k} \left(\mathcal{F}_{\mathcal{S}_j\cup\mathcal{T}} \cap \mathcal{F}_{\mathcal{S}_j\cap\mathcal{T}}\right) \\
= \; & \mathcal{F}_{\mathcal{S}_1\cap\mathcal{T}} \cap \cdots \cap \mathcal{F}_{\mathcal{S}_k\cap\mathcal{T}} \cap \mathcal{F}_{\mathcal{S}_1\cup\mathcal{T}} \cap \cdots \cap \mathcal{F}_{\mathcal{S}_k\cup\mathcal{T}}.
\end{aligned}$$

Therefore, by induction we can conclude that if $\emptyset \subset \mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_k \subset \mathcal{I}_L$, is a telescopic sequence of subsets, then $\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{A}_i}$ is a face of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$; conversely, every face of $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ can be written in this form. Let $\mathcal{B}_j = \mathcal{A}_j - \mathcal{A}_{j-1}$, $j = 1, 2, \cdots, k+1$, where we set $\mathcal{A}_0 = \emptyset$ and $\mathcal{A}_{k+1} = \mathcal{I}_L$. Let Π be the set of permutation π on $\mathcal{I}_L$ such that

$$\left\{\pi\left(\sum_{l=0}^{j-1} |\mathcal{B}_l| + 1\right), \cdots, \pi\left(\sum_{l=0}^{j} |\mathcal{B}_l|\right)\right\} = \mathcal{B}_j$$

for $j = 1, 2, \cdots, k+1$, where $\mathcal{B}_0 = \emptyset$ and $|\mathcal{B}_0| = 0$. Each permutation $\pi \in \Pi$ is associated with a vertex of $\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{A}_j}$ and vice versa. Hence $\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{A}_j}$ has totally $|\Pi| = \prod_{j=1}^{k+1}(|\mathcal{B}_j|!)$ vertices. Moreover, we have $dim(\bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{A}_j}) \leq L - k - 1$. Any rate tuple $(R_1, \cdots, R_L) \in \bigcap_{j=1}^{k} \mathcal{F}_{\mathcal{A}_j}$ can be achieved via a $k+1$-step successive group coding scheme. That is, at step j, encoders in group $\mathcal{B}_j$ jointly search for codewords $\mathbf{U}_i$ $(i \in \mathcal{B}_j)$ such that $(\mathbf{U}_i, i \in \mathcal{B}_j)$ are jointly typical with $(\mathbf{X}, \mathbf{U}_i, i \in \bigcup_{l=1}^{j-1} \mathcal{B}_l)$, $j = 1, \cdots, k+1$. So successive coding is implemented among different groups while joint coding is implemented within each group.

Now we propose a splitting method to achieve a general point on $\mathcal{D}_{MD}(U_i, i \in \mathcal{I}_L)$ via successive coding.

Let $\mathcal{U} = \{U_{1,1}, \cdots, U_{1,m_1}, U_{2,1}, \cdots, U_{2,m_2}, \cdots, U_{L,1}, \cdots, U_{L,m_L}\}$ jointly distributed with the generic source variables X such that

1. $\sum_{i=1}^{L} m_i \leq 2L - 1$ and $1 \leq m_i \leq 2$ for all $i \in \mathcal{I}_L$;
2. $U_{i,m_i} \rightarrow U_{i,1} \rightarrow (X, \{U_{j,1}, \cdots, U_{j,m_j}\}_{j \neq i})$ form a Markov chain for all $i \in \mathcal{I}_L$;
3. $(X, U_1, \cdots, U_L) = (X, U_{1,1}, \cdots, U_{L,1})$ 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 Y. 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 Quantization with Quantization 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 marginal 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 t-th 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, $i = 1, \cdots, L$.
3. Reconstruction: For all $\mathcal{A} \subseteq \mathcal{I}_L$, decoder $\mathcal{A}$ constructs $\hat{\mathbf{X}}_\mathcal{A}$ with $\mathbf{U}_{i,1}(j_{i,2}^*, j_{i,1}^*)$ (or $\mathbf{U}_{i,1}(j_{i,1}^*)$), $i \in \mathcal{A}$.

It is easy to verify that the sum-rate of this scheme is exactly $\psi(X, U_i, i \in \mathcal{I}_L)$ via induction over subsets of random variables in $\mathcal{U}_\sigma$. Furthermore, it can be shown that a general point on the dominant face of $\mathcal{R}_{MD}(U_i, i \in \mathcal{I}_L)$ is achievable via a successive quantization scheme of at most $2L - 1$ steps if the quantization splitting is used, and each U_i gets splits at most once.

It is noteworthy that $2L - 1$ is just an upper bound, for rate tuples on the boundary of $\mathcal{D}(U_i, i \in \mathcal{I}_L)$, the coding steps can be further reduced. Specifically, for any rate tuple $(R_1, \cdots, R_L) \in \bigcap_{i=1}^k \mathcal{F}_{\mathcal{A}_i}$, the splitting can be carried out in $\mathcal{B}_i$, $i = 1, 2, \cdots, k+1$ separately. So the number of successive coding steps is $L + dim(\bigcap_{i=1}^k \mathcal{F}_{\mathcal{A}_i}) \leq \sum_{i=1}^{k+1}(2|\mathcal{B}_i| - 1) = 2L - k - 1$.

References

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

[2] 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.

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

[4] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. on Inform. Theory*, vol. IT-33, pp. 502-521, July 1987.

[5] R. Venkataramani, G. Kramer and V. K. Goyal, "Multiple Description Coding With Many Channels," *IEEE Trans. Inform. Theory*, vol. IT-49, NO. 9, pp. 2106-2114, Sep. 2003.

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

[7] J. Edmonds, "Submodular functions, matroids and certain polyhedra, in *Combinatorial structures and their applications* (R. Guy, H. Hanani, N. Sauer, and J. Schonheim, eds.), pp. 69-87, Gordon and Breach, New York, 1970. (Proc. Calgary Int. Conf. 1969).

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

[9] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications* (G. Longo, ed.), vol. 229 of *CISM Courses and Lectures*, pp. 171-231, Springer-Verlag, Vienna/New York, 1978.

[10] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. on Inform. Theory*, submitted for publication.