SYMMETRICAL MULTILEVEL DIVERSITY CODING AND SUBSET ENTROPY INEQUALITIES

A Dissertation

by

JINJING JIANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tie Liu
Committee Members,	Serap Savari
	Anxiao Jiang
	Srinivas Shakkottai
Head of Department,	Chanan Singh

August 2013

Major Subject: Department of Electrical and Computer Engineering

ABSTRACT

Symmetrical multilevel diversity coding (SMDC) is a classical model for coding over distributed storage. In this setting, a simple separate encoding strategy known as superposition coding was shown to be optimal in terms of achieving the minimum sum rate and the entire admissible rate region of the problem in the literature. The proofs utilized carefully constructed induction arguments, for which the classical subset entropy inequality of Han played a key role.

This thesis includes two parts. In the first part the existing optimality proofs for classical SMDC are revisited, with a focus on their connections to subset entropy inequalities. First, a new sliding-window subset entropy inequality is introduced and then used to establish the optimality of superposition coding for achieving the minimum sum rate under a weaker source-reconstruction requirement. Second, a subset entropy inequality recently proved by Madiman and Tetali is used to develop a new structural understanding to the proof of Yeung and Zhang on the optimality of superposition coding for achieving the entire admissible rate region. Building on the connections between classical SMDC and the subset entropy inequalities developed in the first part, in the second part the optimality of superposition coding is further extended to the cases where there is an additional all-access encoder, an additional secrecy constraint or an encoder hierarchy.

To my parents

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. Tie Liu, who believed in me, led me into the exciting area of information theoretic research and worked with me, side by side, on various problems. I can never forget those hours-long discussion in his office and even intensive phone calls at early mornings and midnights when we were at dawn of truly cracking our problem. I can never forget he always did some pioneer work even before I knew the problem, told me where to start and made everything so easy to begin with. I can never forget he would never be satisfied with something we had achieved and always tried to refine our understanding in a bigger picture, which always gave me deeper insight and new inspiration. I can never forget he worked harder on my talk and my paper than myself. I do not believe I can expect a better advisor and a better role model. I truly learn a lot from him not only as a researcher but also as a human being.

I would like to thank Dr. Savari, Dr. Shakkottai and Dr. Jiang for serving on my thesis committee and for providing their thoughtful comments on my research. I would like to thank Hung Ly, Jae Won Yoo and Neeharika Marukala for making our office an exceptionally wonderful place! Many thanks also go to my long-time friends Cong Shen and Yan Zhu for their help these many years!

Finally, I would like to extend my greatest gratitude towards my parents and my wife Ling. Their endless love is the only origin of all my courage.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

In recent years, the boom of cloud computing applications has mounted great challenges on the design of the distributed storage systems, where users would like to store the information on several distributed servers. The motivation of this thesis is in terms of the robustness issue in distributed storage systems as shown in Fig. 1.1.

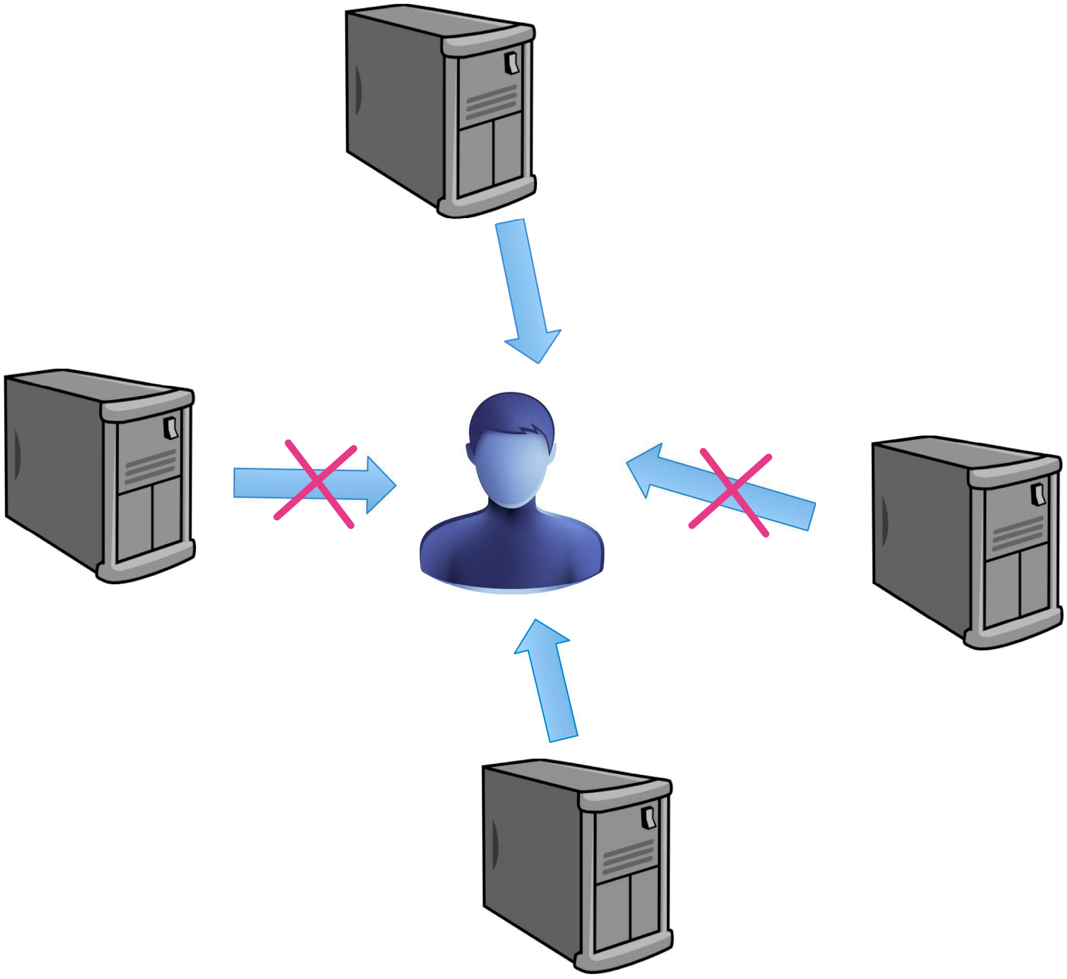


Figure 1.1: An example of distributed storage system with failures.

Particularly, when a network erasure happens temporarily or permanently due to link failure or disk malfunction, it is still desirable to recover the information through the erasure-resilient coding schemes. The schemes such as repetition codes or maximum distance separable (MDS) codes can decode information if at least α out of the total L servers remains available. General speaking, these erasure coding schemes are well understood in the literature. However, from the system design point of view, the problem left is how to choose the decoding threshold α, which controls the tradeoff between robustness and efficiency as shown in Fig. 1.2. One extreme is to set $\alpha = 1$, in which the system is the most robust as long as there is one node available. However, apparently the information should be repeated at every node, which may need huge amount of storage in total. On the contrary, when $\alpha = L$, the system is the most efficient but the least robust since the information is not decodable if any single node fails. Meanwhile, not all information are created equal, i.e., some information are more important than others. Take binary representations as an example: MSBs are more important than LSBs. Due to the information hierarchy, the design that adapts the decoding threshold to the importance of the source is preferred. From the architectural level, the fundamental information theoretic question about this adaptive design is which coding scheme is optimal, separate encoding or joint encoding. Separate encoding is easy to implement and manage while joint encoding is potentially more efficient in terms of minimizing the storage space due to the principle of network coding.

Symmetrical multilevel diversity coding (SMDC) is a classical model arising from the coding over distributed storage, which was first introduced by Roche [1] and Yeung [2]. In this setting, there are a total of L *independent* discrete memoryless sources $S_1, \ldots, S_L$, where the importance of the source S_l is assumed to decrease with the subscript l. The sources are to be encoded by a total of L *randomly accessible*

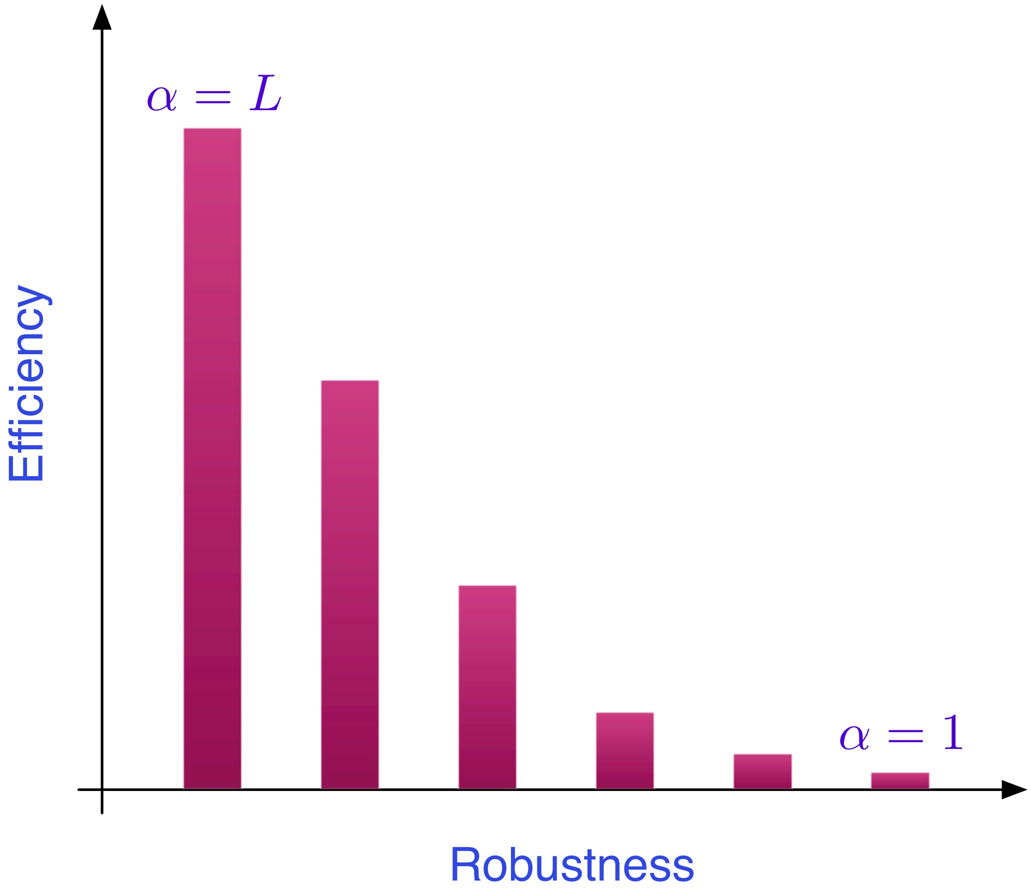


Figure 1.2: Decoding threshold determine the tradeoff between system robustness and storage efficiency.

encoders. The goal of encoding is to ensure that the number of sources that can be nearly perfectly reconstructed grows with the number of available encoder outputs at the decoder. More specifically, denote by $U \subseteq \Omega_L := \{1, \ldots, L\}$ the set of accessible encoders. The realization of U is *unknown* a priori at the encoders. However, the sources $S_1, \ldots, S_\alpha$ need to be nearly perfectly reconstructed whenever $|U| \geq \alpha$ at the decoder. The word "symmetrical" here refers to the fact that the sources that need to be nearly perfectly reconstructed depend on the set of accessible encoders only via its cardinality. The rate allocations at different encoders, however, can be different and are not necessarily symmetrical.

A natural strategy for SMDC is to encode the sources separately at each of the encoders (no coding across different sources) known as *superposition coding* [2]. To show that the natural superposition coding strategy is also optimal, however, turned out to be rather nontrivial. The optimality of superposition coding in terms of achieving the *minimum sum rate* was established by Roche, Yeung, and Hau [3]. The proof used a carefully constructed induction argument, for which the classical subset entropy inequality of Han [4] played a key role. Later, the optimality of superposition coding in terms of achieving the *entire admission rate region* was established by Yeung and Zhang [5]. Their proof was based on a *new* subset entropy inequality, which was established by carefully combining Han's subset inequality with several highly technical results on the analysis of a sequence of linear programs (which are used to characterize the performance of superposition coding).

This thesis includes two parts. In the first part (Section 2), the optimality proofs of [3] and [5] are revisited in light of two new subset entropy inequalities:

- First, a new *sliding-window* subset entropy inequality is introduced, which not only implies the classical subset entropy inequality of Han [4] in a trivial way, but also leads to a new proof of the optimality of superposition encoding for achieving the minimum sum rate under a *weaker* source-reconstruction requirement.

- Second, a subset entropy inequality recently proved by Madiman and Tetali [6] is leveraged to provide a new *structural* understanding to the subset entropy inequality of Yeung and Zhang [5]. Based on this new understanding, a *conditional* version of the subset entropy inequality of Yeung and Zhang [5] is further established, which plays a key role in extending the optimality of superposition coding to the case where there is an additional secrecy constraint.

In the second part of the thesis (Section 3 to Section 5), three extensions of classical SMDC are considered:

- The first extension, which we shall refer to as Symmetrical Multilevel Diversity Coding with an All-Access Encoder (SMDC-A), features an *all-access encoder*, in addition to the L randomly accessible encoders in the classical setting, whose output is available at the decoder *at all time*. This model is mainly motivated by the proliferation of mobile computing devices (laptop computers, tablets, smart phones etc.), which can access both remote storage nodes via unreliable wireless links and local hard disks which are always available but are of limited capacity. It is shown that in this setting, superposition coding remains optimal in terms of achieving the entire admissible rate region. Key to our proof is to identify the supporting hyperplanes that define the superposition coding rate region and then apply the subset entropy inequality of Yeung and Zhang [5].

- The second extension, which we shall refer to as Secure Multilevel Diversity Coding (S-SMDC), extends the problem of SMDC to the *secure* communication setting. The problem was first introduced in [7], where the optimality of superposition coding for achieving the minimum sum rate was established via the classical subset entropy inequality of Han [4]. Through the conditional version of the subset entropy inequality of Yeung and Zhang [5] established in the first part, here we show that superposition coding can, in fact, achieve the entire admissible rate region of the problem, resolving the conjecture of [7] by positive.

- The third extension, which we shall refer as *Hierarchical* Multilevel Diversity Coding (HMDC), further extends the problem of SMDC to a special *asymmet-*

rical setting. This model is a natural generalization of symmetrical multilevel diversity coding to heterogeneous distributed storage systems. Recall that in symmetrical multilevel diversity coding, the number of messages that needs to be recovered by the decoder depends on the available subset of the encoder outputs only via its cardinality. Therefore, the underlying assumption for symmetrical multilevel diversity coding is that all distributed storage nodes have the same reliability. For heterogeneous distributed storage systems, one may associate each storage node with a *reliability score* and design a coding scheme such that the number of messages that needs to be recovered by the decoder depends on the available subset of the encoder outputs via its accumulative reliability score. Then the encoders are classified into different ranks due to their different reliability scores. In our setting, the encoders show two reliability scores. The encoders with higher reliability score are referred as the strong encoders, while the others are weak encoders. With this encoder hierarchy, it is shown that superposition coding remains optimal in achieving the minimum sum rate but the optimality of achieving the entire admissible rate region is still unknown. We conjecture that superposition coding is optimal in achieving the entire admissible rate region.

2. SYMMETRICAL MULTILEVEL DIVERSITY CODING REVISITED

2.1 Problem Statement And Optimality Of Superposition Coding

2.1.1 Problem Statement

As illustrated in Figure 2.1, the problem of SMDC consists of:

- a total of L *independent* discrete memoryless sources $\{S_\alpha[t]\}_{t=1}^{\infty}$, where $\alpha = 1, \ldots, L$ and t is the time index;
- a set of L encoders (encoder 1 to L);
- a decoder which can access a nonempty subset $U \subseteq \Omega_L$ of the encoder outputs.

The realization of U is *unknown* a priori at the encoders. However, no matter which U actually materializes, the decoder needs to nearly perfectly reconstruct the sources $S_1, \ldots, S_\alpha$ whenever $|U| \geq \alpha$.

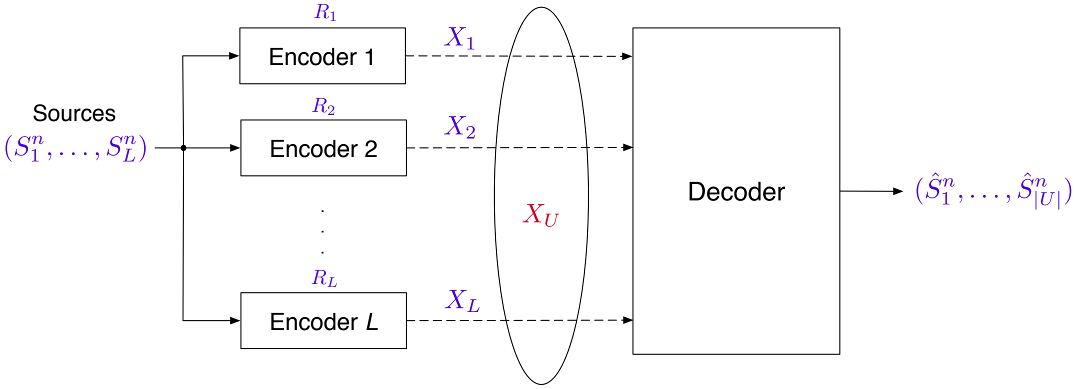


Figure 2.1: The classical SMDC problem where a total of L independent discrete memoryless sources $S_1, \ldots, S_L$ are to be encoded by a total of L encoders. The decoder, which has access to a subset U of the encoder outputs, needs to nearly perfectly reconstruct the sources $S_1, \ldots, S_{|U|}$ no matter what the realization of U is.

Formally, an $(n,(M_1,\ldots,M_L))$ code is defined by a collection of L encoding functions:

$$e_l : \prod_{\alpha=1}^{L} \mathcal{S}_\alpha^n \to \{1,\ldots,M_l\}, \quad \forall l = 1,\ldots,L \tag{2.1}$$

and $2^L - 1$ decoding functions:

$$d_U : \prod_{l\in U}\{1,\ldots,M_l\} \to \prod_{\alpha=1}^{|U|} \mathcal{S}_\alpha^n, \quad \forall U \subseteq \Omega_L \text{ s.t. } U \neq \emptyset. \tag{2.2}$$

A nonnegative rate tuple $(R_1,\ldots,R_L)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block-length n, an $(n,(M_1,\ldots,M_L))$ code such that:

- (Rate constraints at the encoders)

$$\frac{1}{n}\log M_l \le R_l + \epsilon, \qquad \forall l = 1,\ldots,L; \tag{2.3}$$

- (Asymptotically perfect reconstructions at the decoder)

$$\Pr\left\{d_U(X_U) \neq (S_1^n,\ldots,S_{|U|}^n)\right\} \le \epsilon, \qquad \forall U \subseteq \Omega_L \text{ s.t. } U \neq \emptyset \tag{2.4}$$

  where $S_\alpha^n := \{S_\alpha[t]\}_{t=1}^n$, $X_l := e_l(S_1^n,\ldots,S_L^n)$ is the output of encoder l, and $X_U := \{X_l : l \in U\}$.

The *admissible rate region* $\mathcal{R}$ is the collection of *all* admissible rate tuples $(R_1,\ldots,R_L)$. The *minimum sum rate* R_{ms} is defined as

$$R_{ms} := \min_{(R_1,\ldots,R_L)\in\mathcal{R}} \sum_{l=1}^{L} R_l. \tag{2.5}$$

2.1.2 Superposition Coding Rate Region

As mentioned previously, a natural strategy for SMDC is *superposition coding*, i.e., to encode the sources separately at the encoders and there is *no* coding across different sources. Formally, the problem of encoding a single source S_α can be viewed as a special case of the general SMDC problem, where the sources S_m are deterministic for all $m \neq \alpha$. In this case, the source S_α needs to be nearly perfectly reconstructed whenever the decoder can access at least α encoder outputs. Thus, the problem is essentially to transmit S_α over an *erasure* channel, and the following simple source-channel separation scheme is known to be optimal [1, 2]:

- First compress the source sequence S_α^n into a source message W_α using a *lossless* source code. It is well known [8, Ch. 5] that the rate of the source message W_α can be made arbitrarily close to the entropy rate $H(S_\alpha)$ for sufficiently large block-length n.

- Next, the source message W_α is encoded at encoders 1 to L using a *maximum distance separable* code [9]. It is well known [1, 2] that the source message W_α can be perfectly recovered at the decoder whenever

$$\sum_{l \in U} R_l \geq \frac{1}{n} H(W_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \tag{2.6}$$

  for sufficiently large block length n, where $\Omega_L^{(\alpha)}$ denotes the collection of all subsets of Ω_L of size α.

Combining the above two steps, we conclude that the admissible rate region for encoding a single source S_α is given by the collection of all nonnegative rate tuples

$(R_1, \ldots, R_L)$ satisfying

$$\sum_{l \in U} R_l \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}. \tag{2.7}$$

By definition, the superposition coding rate region $\mathcal{R}_{sup}$ for encoding the sources $S_1, \ldots, S_L$ is given by the collection of all nonnegative rate tuples $(R_1, \ldots, R_L)$ such that

$$R_l := \sum_{\alpha=1}^{L} r_l^{(\alpha)} \tag{2.8}$$

for some nonnegative $r_l^{(\alpha)}$, $\alpha = 1, \ldots, L$ and $l = 1, \ldots, L$, satisfying

$$\sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}. \tag{2.9}$$

In principle, an explicit characterization of the superposition coding rate region $\mathcal{R}_{sup}$ can be obtained by eliminating $r_l^{(\alpha)}$, $\alpha = 1, \ldots, L$ and $l = 1, \ldots, L$, via a Fourier-Motzkin elimination from (2.8) and (2.9). However, the elimination process is *unmanageable* even for moderate L, as there are simply too many equations involved. On the other hand, note that the superposition coding rate region $\mathcal{R}_{sup}$ is a convex polyhedron with polyhedral cone being $(\mathbb{R}^+)^L$, so an equivalent characterization is to characterize the supporting hyperplanes:

$$\sum_{l=1}^{L} \lambda_l R_l \geq f(\boldsymbol{\lambda}), \quad \forall \boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_L) \in (\mathbb{R}^+)^L \tag{2.10}$$

where

$$f(\boldsymbol{\lambda}) = \min_{(R_1, \ldots, R_L) \in \mathcal{R}_{sup}} \sum_{l=1}^{L} \lambda_l R_l. \tag{2.11}$$

To solving for $f(\boldsymbol{\lambda})$, (2.11) can be explicitly written as the following linear program,

$$\begin{aligned} \min \quad & \textstyle\sum_{l=1}^{L} \lambda_l \left(\sum_{\alpha=1}^{L} r_l^{(\alpha)} \right) \\ \text{subject to} \quad & \textstyle\sum_{l\in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \text{ and } \alpha = 1, \ldots, L \\ & r_l^{(\alpha)} \geq 0, \quad \forall \alpha = 1, \ldots, L \text{ and } l = 1, \ldots, L. \end{aligned} \tag{2.12}$$

The linear program can be further written as

$$\begin{aligned} \min \quad & \textstyle\sum_{l=1}^{L} \left(\sum_{\alpha=1}^{L} \lambda_l r_l^{(\alpha)} \right) \\ \text{subject to} \quad & \textstyle\sum_{l\in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \text{ and } \alpha = 1, \ldots, L \\ & r_l^{(\alpha)} \geq 0, \quad \forall \alpha = 1, \ldots, L \text{ and } l = 1, \ldots, L. \end{aligned} \tag{2.13}$$

Clearly, the above optimization problem can be separated into the following L sub-optimization problems:

$$f(\boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f'_\alpha(\boldsymbol{\lambda}) \tag{2.14}$$

where

$$f'_\alpha(\boldsymbol{\lambda}) = \min \sum_{l=1}^{L} \lambda_l r_l^{(\alpha)}, \tag{2.15}$$

subject to

$$\sum_{l\in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}, \tag{2.16}$$

$$r_l^{(\alpha)} \geq 0, \quad \forall l = 1, \ldots, L. \tag{2.17}$$

The minimization of $f'_\alpha(\boldsymbol{\lambda})$ can be explicitly written as the following linear program

$$\begin{array}{rl} \max & \left(\sum_{U\in\Omega_L^{(\alpha)}} c_{\boldsymbol{\lambda}}(U)\right) H(S_\alpha) \\ \text{subject to} & \sum_{\{U\in\Omega_L^{(\alpha)}:U\ni l\}} c_{\boldsymbol{\lambda}}(U) \le \lambda_l, \quad \forall l=1,\ldots,L \\ & c_{\boldsymbol{\lambda}}(U)\ge 0, \quad \forall U\in\Omega_L^{(\alpha)}. \end{array} \tag{2.18}$$

and (2.18) follows from the strong *duality* for linear programs. For any $\boldsymbol{\lambda}\in(\mathbb{R}^+)^L$ and any $\alpha=1,\ldots,L$, let

$$f_\alpha(\boldsymbol{\lambda}) \quad := \quad \begin{array}{rl} \max & \sum_{U\in\Omega_L^{(\alpha)}} c_{\boldsymbol{\lambda}}(U) \\ \text{subject to} & \sum_{\{U\in\Omega_L^{(\alpha)}:U\ni l\}} c_{\boldsymbol{\lambda}}(U) \le \lambda_l, \quad \forall l=1,\ldots,L \\ & c_{\boldsymbol{\lambda}}(U)\ge 0, \quad \forall U\in\Omega_L^{(\alpha)}. \end{array} \tag{2.19}$$

Then, we have $f'_\alpha(\boldsymbol{\lambda}) = f_\alpha(\boldsymbol{\lambda})H(S_\alpha)$ and hence

$$f(\boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha) \tag{2.20}$$

for any $\boldsymbol{\lambda}\in(\mathbb{R}^+)^L$. Substituting (2.20) into (2.10), we conclude that the superposition coding rate region $\mathcal{R}_{sup}$ is given by the collection of nonnegative rate tuples $(R_1,\ldots,R_L)$ satisfying

$$\sum_{l=1}^{L}\lambda_l R_l \ge \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha), \quad \forall\boldsymbol{\lambda}\in(\mathbb{R}^+)^L. \tag{2.21}$$

For a general $\boldsymbol{\lambda}$, the linear program (2.19) does not admit a *closed-form* solution. However, for $\boldsymbol{\lambda}=\mathbf{1}:=(1,\ldots,1)$ it can be easily verified that $c_{\mathbf{1}}^{(\alpha)} = \{c_{\mathbf{1}}(U) : U\in\Omega_L^{(\alpha)}\}$ where

$$c_{\mathbf{1}}(U) := \frac{1}{\binom{L-1}{\alpha-1}} \tag{2.22}$$

is an *optimal* solution to the linear program (2.19), and we thus have

$$f_\alpha(\mathbf{1}) = \sum_{U \in \Omega_L^{(\alpha)}} c_{\mathbf{1}}(U) = \frac{\binom{L}{\alpha}}{\binom{L-1}{\alpha-1}} = \frac{L}{\alpha} \tag{2.23}$$

for any $\alpha = 1, \ldots, L$. Hence, the minimum sum rate that can be achieved by superposition coding is given by

$$\min_{(R_1,\ldots,R_L) \in \mathcal{R}_{sup}} \sum_{l=1}^{L} R_l = f(\mathbf{1}) = \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) H(S_\alpha) = \sum_{\alpha=1}^{L} (L/\alpha) H(S_\alpha). \tag{2.24}$$

2.1.3 Optimality Of Superposition Coding: Known Proofs

To show that superposition coding is optimal in terms of achieving the entire admissible rate region, we need to show that for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ we have

$$\sum_{l=1}^{L} \lambda_l R_l \geq \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha), \quad \forall (R_1, \ldots, R_L) \in \mathcal{R}. \tag{2.25}$$

In particular, to show that superposition coding is optimal in terms of achieving the minimum sum rate, we need to show that

$$\sum_{l=1}^{L} R_l \geq \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) H(S_\alpha) = \sum_{\alpha=1}^{L} (L/\alpha) H(S_\alpha), \quad \forall (R_1, \ldots, R_L) \in \mathcal{R}. \tag{2.26}$$

Note that for any admissible rate tuple $(R_1, \ldots, R_L) \in \mathcal{R}$ and $\epsilon > 0$, by the rate constraints (2.3) we have

$$n(R_l + \epsilon) \geq H(X_l), \quad \forall l = 1, \ldots, L \tag{2.27}$$

for sufficiently large block-length n. Furthermore, by the asymptotically perfect reconstruction requirement (2.4) and the well-known Fano's inequality we have

$$H(S_1^n, \ldots, S_\alpha^n | X_U) \leq n\delta_\alpha^{(n)} \tag{2.28}$$

for any $U \in \Omega_L^{(\alpha)}$ and $\alpha = 1, \ldots, L$, where $\delta_\alpha^{(n)} \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$. Thus, for any $V \in \Omega_L^{(\alpha-1)}$ we have

$$\begin{aligned}
H(X_V | S_1^n, \ldots, S_{\alpha-2}^n) &= H(X_V | S_1^n, \ldots, S_{\alpha-1}^n) + I(X_V ; S_{\alpha-1}^n | S_1^n, \ldots, S_{\alpha-2}^n) && (2.29)\\
&= H(X_V | S_1^n, \ldots, S_{\alpha-1}^n) + H(S_{\alpha-1}^n | S_1^n, \ldots, S_{\alpha-2}^n) - \\
&\quad H(S_{\alpha-1}^n | S_1^n, \ldots, S_{\alpha-2}^n, X_V) && (2.30)\\
&\geq H(X_V | S_1^n, \ldots, S_{\alpha-1}^n) + H(S_{\alpha-1}^n) - H(S_1^n, \ldots, S_{\alpha-1}^n | X_V) \\
& && (2.31)\\
&\geq H(X_V | S_1^n, \ldots, S_{\alpha-1}^n) + nH(S_{\alpha-1}) - n\delta_{\alpha-1}^{(n)} && (2.32)
\end{aligned}$$

where (2.31) follows from the facts that all sources are independent so

$$H(S_{\alpha-1}^n | S_1^n, \ldots, S_{\alpha-2}^n) = H(S_{\alpha-1}^n)$$

and that

$$\begin{aligned}
H(S_{\alpha-1}^n | S_1^n, \ldots, S_{\alpha-2}^n, X_V) &= H(S_1^n, \ldots, S_{\alpha-1}^n | X_V) - H(S_1^n, \ldots, S_{\alpha-2}^n | X_V) \\
&\leq H(S_1^n, \ldots, S_{\alpha-1}^n | X_V). && (2.33)
\end{aligned}$$

Therefore, starting with (2.27) and applying (2.32) *iteratively* may lead us towards a proof of (2.25) and (2.26). Note, however, that to apply (2.32) iteratively we shall need to bound from below $H(X_V | S_1^n, \ldots, S_{\alpha-1}^n)$ in terms of $H(X_U | S_1^n, \ldots, S_{\alpha-1}^n)$ for some $U \in \Omega_L^{(\alpha)}$. The key observation of [3] and [5] is that such bounds exist, not for an arbitrary individual pair of U and V, but rather at the level of an appropriate *averaging* among $V \in \Omega_L^{(\alpha-1)}$ and $U \in \Omega_L^{(\alpha)}$.

More specifically, [3] considered the classical subset entropy inequality of Han [4], which can be written as follows.

Theorem 1 (A subset entropy inequality of Han [4]). *For any collection of L jointly distributed random variables $(X_1, \ldots, X_L)$, we have*

$$\frac{1}{\binom{L}{\alpha-1}} \sum_{V \in \Omega_L^{(\alpha-1)}} \frac{H(X_V)}{\alpha - 1} \geq \frac{1}{\binom{L}{\alpha}} \sum_{U \in \Omega_L^{(\alpha)}} \frac{H(X_U)}{\alpha} \tag{2.34}$$

for any $\alpha = 2, \ldots, L$.

Essentially, considering the average joint entropy of all subsets of fixed size, (2.34) says that the average joint entropy per element decreases with the size of the subsets. The proof of Theorem 1 can be found in Appendix A.

Iteratively applying (2.32) and (2.34), we may obtain

$$\begin{aligned} \frac{1}{L}\sum_{l=1}^{L} H(X_l) &= \frac{1}{\binom{L}{1}} \sum_{V \in \Omega_L^{(1)}} H(X_V) && (2.35) \\ &\geq \frac{1}{\binom{L}{m}} \sum_{U \in \Omega_L^{(m)}} \frac{H(X_U | S_1^n, \ldots, S_m^n)}{m} + n\sum_{\alpha=1}^{m} \frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{m} \frac{\delta_\alpha^{(n)}}{\alpha} && (2.36) \end{aligned}$$

for any $m = 1, \ldots, L$. In particular, let $m = L$, and we have

$$\begin{aligned} \frac{1}{L}\sum_{l=1}^{L} H(X_l) &\geq \frac{1}{\binom{L}{L}} \sum_{U \in \Omega_L^{(L)}} \frac{H(X_U | S_1^n, \ldots, S_L^n)}{L} + n\sum_{\alpha=1}^{L} \frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{L} \frac{\delta_\alpha^{(n)}}{\alpha} \\ &\geq n\sum_{\alpha=1}^{L} \frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{L} \frac{\delta_\alpha^{(n)}}{\alpha}. && (2.37) \end{aligned}$$

Substituting (2.27) into (2.37) and dividing both sides of the inequality by n, we have

$$\frac{1}{L}\sum_{l=1}^{L}(R_l+\epsilon) \geq \sum_{\alpha=1}^{L}\frac{H(S_\alpha)}{\alpha} - \sum_{\alpha=1}^{L}\frac{\delta_\alpha^{(n)}}{\alpha}. \tag{2.38}$$

Finally, letting $n \to \infty$ and $\epsilon \to 0$ completes the proof of (2.26), i.e., superposition coding can achieve the minimum sum rate for the general SMDC problem.

To prove that superposition coding can in fact achieve the entire admissible rate region, Yeung and Zhang [5] proved the following key subset entropy inequality.

Theorem 2 (A subset entropy inequality of Yeung and Zhang [5])**.** *For any* $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$*, there exists a function* $c_{\boldsymbol{\lambda}} : 2^{\Omega_L} \setminus \emptyset \to \mathbb{R}^+$ *such that:*

1) for each $\alpha = 1, \ldots, L$, $c_{\boldsymbol{\lambda}}^{(\alpha)} := \{c_{\boldsymbol{\lambda}}(U) : U \in \Omega_L^{(\alpha)}\}$ *is an* optimal *solution to the linear program* (2.19)*; and*

2) for each $\alpha = 2, \ldots, L$,

$$\sum_{V\in\Omega_L^{(\alpha-1)}} c_{\boldsymbol{\lambda}}(V)H(X_V) \geq \sum_{U\in\Omega_L^{(\alpha)}} c_{\boldsymbol{\lambda}}(U)H(X_U) \tag{2.39}$$

for any collection of L *jointly distributed random variables* $(X_1, \ldots, X_L)$.

Iteratively applying (2.32) and (2.39), we may obtain

$$\begin{aligned}\sum_{V\in\Omega_L^{(1)}} c_{\boldsymbol{\lambda}}(V)H(V) \geq \sum_{U\in\Omega_L^{(m)}} & c_{\boldsymbol{\lambda}}(U)H(X_U|S_1^n, \ldots, S_m^n)+ \\ & n\sum_{\alpha=1}^{m} f_\alpha(\boldsymbol{\lambda})H(S_\alpha) - n\sum_{\alpha=1}^{m} f_\alpha(\boldsymbol{\lambda})\delta_\alpha^{(n)}\end{aligned} \tag{2.40}$$

for any $m = 1, \ldots, L$. In particular, let $m = L$, and note that for $\alpha = 1$ the optimal solution to the linear program (2.19) is unique and is given by

$$c_{\boldsymbol{\lambda}}(\{l\}) = \lambda_l, \quad \forall l \in \Omega_L. \tag{2.41}$$

We have

$$\begin{aligned}\sum_{l=1}^{L} \lambda_l H(X_l) \geq & \sum_{U \in \Omega_L^{(L)}} c_{\boldsymbol{\lambda}}(U) H(X_U | S_1^n, \ldots, S_L^n) + \\ & n \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - n \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)} && (2.42) \\ & \geq n \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - n \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)}. && (2.43)\end{aligned}$$

Substituting (2.27) into (2.43) and dividing both sides of the inequality by n, we have

$$\sum_{l=1}^{L} \lambda_l (R_l + \epsilon) \geq \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)}. \tag{2.44}$$

Finally, letting $n \to \infty$ and $\epsilon \to 0$ completes the proof of (2.25), i.e., superposition coding can achieve the entire admissible rate region for the general SMDC problem.

2.2 Minimum Sum Rate Via A Sliding-window Subset Entropy Inequality

In this section, we prove a new *sliding-window* subset entropy inequality and then use it to provide an alternative proof of the optimality of superposition coding for achieving the minimum sum rate.

2.2.1 A Sliding-window Subset Entropy Inequality

For any integer l let

$$\langle l \rangle := \begin{cases} l \bmod L, & \text{if } l \bmod L \neq 0 \\ L, & \text{if } l \bmod L = 0 \end{cases} \tag{2.45}$$

and for any $l = 1, \ldots, L$ and $\alpha = 1, \ldots, L$ let

$$W_l^{(\alpha)} := \{l, \langle l+1 \rangle, \ldots, \langle l+\alpha-1 \rangle\}. \tag{2.46}$$

As illustrated in Figure 2.2, $W_l^{(\alpha)}$ represents a *sliding window* of length α starting with l when the integers $1, \ldots, L$ are circularly placed (clockwise or counter clockwise) based on their natural order. We have the following *sliding-window* subset entropy inequality.

Theorem 3 (A sliding-window subset entropy inequality)**.** *For any collection of L jointly distributed random variables* $(X_1, \ldots, X_L)$*, we have*

$$\sum_{l=1}^{L} \frac{H(X_{W_l^{(\alpha-1)}})}{\alpha-1} \geq \sum_{l=1}^{L} \frac{H(X_{W_l^{(\alpha)}})}{\alpha} \tag{2.47}$$

for any $\alpha = 2, \ldots, L$*. The equalities hold when* $X_1, \ldots, X_L$ *are mutually independent of each other.*

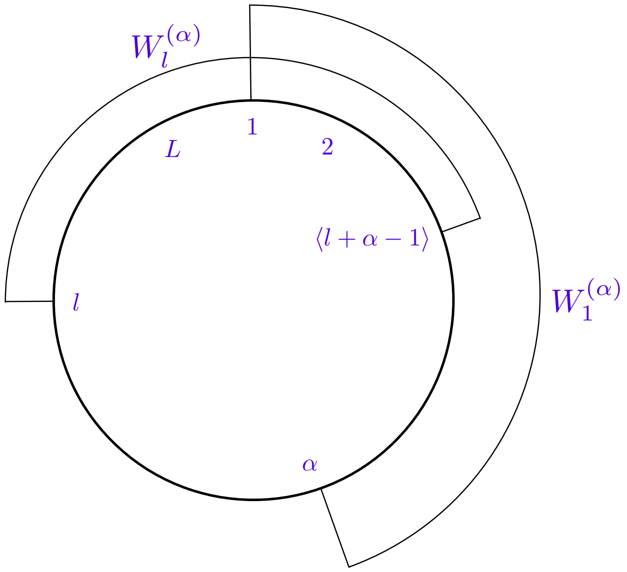


Figure 2.2: An illustration of the sliding windows of length α when the integers $1, \ldots, L$ are circularly placed (clockwise) based on their natural order.

Proof. Consider a proof via an induction on α. First, for $\alpha = 2$ we have

$$\begin{aligned}
\sum_{l=1}^{L} H(X_{W_l^{(1)}}) &= \sum_{l=1}^{L} H(X_l) && (2.48)\\
&= \sum_{l=1}^{L} \frac{H(X_l) + H(X_{\langle l+1\rangle})}{2} && (2.49)\\
&\geq \sum_{l=1}^{L} \frac{H(X_l, X_{\langle l+1\rangle})}{2} && (2.50)\\
&= \sum_{l=1}^{L} \frac{H(X_{W_l^{(2)}})}{2} && (2.51)
\end{aligned}$$

where (2.50) follows from the independence bound on entropy.

Next, assume that the inequality (2.47) holds for $\alpha = r$ for some $r \in \{2, \ldots, L -$

1}, i.e.,

$$\sum_{l=1}^{L} \frac{H(X_{W_l^{(r-1)}})}{r-1} \geq \sum_{l=1}^{L} \frac{H(X_{W_l^{(r)}})}{r}. \tag{2.52}$$

We have

$$\begin{aligned}
\sum_{l=1}^{L} H(X_{W_l^{(r)}}) &= \frac{1}{2}\sum_{l=1}^{L}\left[H(X_{W_l^{(r)}}) + H(X_{W_{\langle l+1\rangle}^{(r)}})\right] && (2.53)\\
&\geq \frac{1}{2}\sum_{l=1}^{L}\left[H(X_{W_l^{(r+1)}}) + H(X_{W_{\langle l+1\rangle}^{(r-1)}})\right] && (2.54)\\
&= \frac{1}{2}\sum_{l=1}^{L} H(X_{W_l^{(r+1)}}) + \frac{1}{2}\sum_{l=1}^{L} H(X_{W_{\langle l+1\rangle}^{(r-1)}}) && (2.55)\\
&= \frac{1}{2}\sum_{l=1}^{L} H(X_{W_l^{(r+1)}}) + \frac{1}{2}\sum_{l=1}^{L} H(X_{W_l^{(r-1)}}) && (2.56)\\
&\geq \frac{1}{2}\sum_{l=1}^{L} H(X_{W_l^{(r+1)}}) + \frac{1}{2}\cdot\frac{r-1}{r}\sum_{l=1}^{L} H(X_{W_l^{(r)}}) && (2.57)
\end{aligned}$$

where (2.54) follows from the submodularity of entropy [14, Ch. 14.A]

$$H(X_U) + H(X_V) \geq H(X_{U\cup V}) + H(X_{U\cap V}) \tag{2.58}$$

for $U = W_l^{(r)}$ and $V = W_{\langle l+1\rangle}^{(r)}$ so $U \cup V = W_l^{(r+1)}$ and $U \cap V = W_{\langle l+1\rangle}^{(r-1)}$, and (2.57) follows from the induction assumption (2.52). Moving the second term on the right-hand side of (2.57) to the left and multiplying both sides by $\frac{2}{r+1}$, we have

$$\frac{1}{r}\sum_{l=1}^{L} H(X_{W_l^{(r)}}) \geq \frac{1}{r+1}\sum_{l=1}^{L} H(X_{W_l^{(r+1)}}). \tag{2.59}$$

We have thus proved that the inequality (2.47) also holds for $\alpha = r+1$.

Finally, note that when $X_1, \dots, X_L$ are mutually independent, we have

$$\sum_{l=1}^{L} \frac{H(X_{W_l^{(\alpha)}})}{\alpha} = \sum_{l=1}^{L} H(X_l), \quad \forall \alpha = 1, \dots, L. \tag{2.60}$$

This completes the proof of Theorem 3. □

Note that for $\alpha = L$, the classical subset entropy inequality of Han (2.34) and the sliding-window subset entropy inequality (2.47) are equivalent, and both can be equivalently written as

$$\frac{1}{L-1} \sum_{l=1}^{L} H(X_{\Omega_L \setminus \{l\}}) \geq H(X_{\Omega_L}). \tag{2.61}$$

For a *general* α, the classical subset entropy inequality of Han (2.34) can be derived from the sliding-window subset entropy inequality (2.47) via a simple *permutation* argument as follows. Let π be a permutation on Ω_L. For any $l = 1, \dots, L$ and $\alpha = 1, \dots, L$, let

$$W_{\pi,l}^{(\alpha)} := \{\pi^{-1}(l), \pi^{-1}(\langle l+1 \rangle), \dots, \pi^{-1}(\langle l+\alpha-1 \rangle)\}. \tag{2.62}$$

By Theorem 3, we have

$$\frac{1}{\alpha-1} \sum_{l=1}^{L} H(X_{W_{\pi,l}^{(\alpha-1)}}) \geq \frac{1}{\alpha} \sum_{l=1}^{L} H(X_{W_{\pi,l}^{(\alpha)}}) \tag{2.63}$$

for any $\alpha = 2, \dots, L$. Averaging (2.63) over all possible permutations π, we have

$$\frac{1}{L!} \sum_{\pi} \left[\frac{1}{\alpha-1} \sum_{l=1}^{L} H(X_{W_{\pi,l}^{(\alpha-1)}}) \right] \geq \frac{1}{L!} \sum_{\pi} \left[\frac{1}{\alpha} \sum_{l=1}^{L} H(X_{W_{\pi,l}^{(\alpha)}}) \right]. \tag{2.64}$$

Note that for any $\alpha = 1, \ldots, L$,

$$\sum_{\pi} \sum_{l=1}^{L} H(X_{W_{\pi,l}^{(\alpha)}}) = L \cdot \alpha!(L-\alpha)! \sum_{U \in \Omega_L^{(\alpha)}} H(X_U). \tag{2.65}$$

Substituting (2.65) into (2.64) and dividing both sides of the inequality by L establish the classical subset entropy inequality of Han (2.34).

2.2.2 The Minimum Sum Rate

The sliding-window subset entropy inequality (2.47) can be used to provide an alternative proof of the optimality of superposition coding for achieving the minimum sum rate as follows. Let us first show that

$$\frac{1}{L}\sum_{l=1}^{L} H(X_l) = \frac{1}{L}\sum_{l=1}^{L} H(X_{W_l^{(1)}}) \tag{2.66}$$

$$\geq \frac{1}{L}\sum_{l=1}^{L} \frac{H(X_{W_l^{(m)}}|S_1^n, \ldots, S_m^n)}{m} + n\sum_{\alpha=1}^{m} \frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{m} \frac{\delta_\alpha^{(n)}}{\alpha} \tag{2.67}$$

for any $m = 1, \ldots, L$.

Consider a proof via an induction on m. When $m = 1$, (2.67) can be written as

$$\frac{1}{L}\sum_{l=1}^{L} H(X_l) \geq \frac{1}{L}\sum_{l=1}^{L} H(X_l|S_1^n) + nH(S_1) - n\delta_1^{(n)} \tag{2.68}$$

which can be obtained via a uniform averaging of (2.32) for $\alpha = 2$ and $V = \{l\}$ for $l = 1, \ldots, L$. Now assume that the inequality (2.67) holds for $m = r-1$ for some

$r \in \{2, \dots, L\}$. We have

$$\frac{1}{L}\sum_{l=1}^{L} H(X_l) \geq \frac{1}{L}\sum_{l=1}^{L} \frac{H(X_{W_l^{(r-1)}}|S_1^n, \dots, S_{r-1}^n)}{r-1} + n\sum_{\alpha=1}^{r-1}\frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{r-1}\frac{\delta_\alpha^{(n)}}{\alpha} \tag{2.69}$$

$$\geq \frac{1}{L}\sum_{l=1}^{L} \frac{H(X_{W_l^{(r)}}|S_1^n, \dots, S_{r-1}^n)}{r} + n\sum_{\alpha=1}^{r-1}\frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{r-1}\frac{\delta_\alpha^{(n)}}{\alpha} \tag{2.70}$$

where (2.70) follows from the sliding-window subset entropy inequality (2.47) with $\alpha = r$. Letting $\alpha = r+1$ and $V = W_l^{(r)}$ in (2.32), we have

$$H(X_{W_l^{(r)}}|S_1^n, \dots, S_{r-1}^n) \geq H(X_{W_l^{(r)}}|S_1^n, \dots, S_r^n) + nH(S_r) - n\delta_r^{(n)}. \tag{2.71}$$

Substituting (2.71) into (2.70) gives

$$\frac{1}{L}\sum_{l=1}^{L} H(X_l) \geq \frac{1}{L}\sum_{l=1}^{L} \frac{H(X_{W_l^{(r)}}|S_1^n, \dots, S_r^n)}{r} + n\sum_{\alpha=1}^{r}\frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{r}\frac{\delta_\alpha^{(n)}}{\alpha}. \tag{2.72}$$

This completes the proof of the induction step and hence (2.67).

Now let $m = L$, and we have

$$\frac{1}{L}\sum_{l=1}^{L} H(X_l) \geq \frac{1}{L}\sum_{l=1}^{L} \frac{H(X_{W_l^{(L)}}|S_1^n, \dots, S_L^n)}{L} + n\sum_{\alpha=1}^{L}\frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{L}\frac{\delta_\alpha^{(n)}}{\alpha} \tag{2.73}$$

$$\geq n\sum_{\alpha=1}^{L}\frac{H(S_\alpha)}{\alpha} - n\sum_{\alpha=1}^{L}\frac{\delta_\alpha^{(n)}}{\alpha}. \tag{2.74}$$

Substituting (2.27) into (2.74) and dividing both sides of the inequality by n, we have

$$\frac{1}{L}\sum_{l=1}^{L}(R_l + \epsilon) \geq \sum_{\alpha=1}^{L}\frac{H(S_\alpha)}{\alpha} - \sum_{\alpha=1}^{L}\frac{\delta_\alpha^{(n)}}{\alpha}. \tag{2.75}$$

Finally, letting $n \to \infty$ and $\epsilon \to 0$ completes the proof of (2.26), i.e., superposition coding can achieve the minimum sum rate for the general SMDC problem.

Note that unlike the original proof of [3], which uses the classical subset entropy inequality of Han [4] and hence involves *all* nonempty subsets U of Ω_L, our proof relies on the sliding-window subset entropy inequality (2.47) and hence only involves the subsets U of a sliding-window type, i.e., $U = W_l^{(\alpha)}$ for some $l = 1, \ldots, L$ and $\alpha = 1, \ldots, L$. Therefore, based on our proof, the converse result (2.26) remains to be true even if we weaken the asymptotically perfect reconstruction requirement (2.4) to

$$\Pr\left\{d_U(X_U) \neq (S_1^n, \ldots, S_{|U|}^n)\right\} \leq \epsilon, \qquad \forall U \in \left\{W_l^{(\alpha)} : l = 1, \ldots, L \text{ and } \alpha = 1, \ldots, L\right\}. \tag{2.76}$$

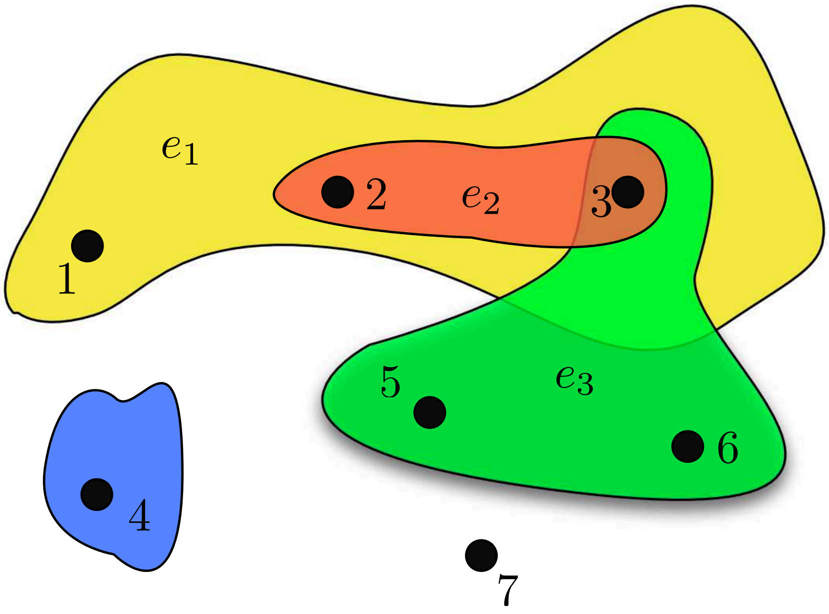
This is the *definitive* advantage of our proof over that based on the classical subset entropy inequality of Han [4].

2.3 The Subset Entropy Inequality Of Yeung And Zhang Revisited

In this section, we revisit the subset entropy inequality of Yeung and Zhang (2.39), which played a key in their proof [5] of the optimality of superposition coding for achieving the entire admissible rate region of the problem. As mentioned previously, in [5] the subset entropy inequality (2.39) was proved by combining the classical subset entropy inequality of Han [4] and a number of analysis results on the sequence of linear programs (2.19). However, the inequality, as stated in Theorem 2, does not even directly imply the classical subset entropy inequality of Han [4]. The reason is that Theorem 2 merely asserts the existence of a set of optimal solutions $c_{\boldsymbol{\lambda}}^{(\alpha)}$, $\alpha = 1, \ldots, L$, that satisfies the subset entropy inequality (2.39), rather than providing a sufficient condition for the inequality to hold. Below, we shall use a subset entropy inequality recently proved by Madiman and Tetali [6] to summarize the analysis results of [5] on the sequence of linear programs (2.19) into a succinct sufficient condition for the subset entropy inequality (2.39) to hold.

2.3.1 A Subset Entropy Inequality Of Madiman And Tetali

Consider a *hypergraph* $(U, \mathcal{V})$ where U is a finite ground set and $\mathcal{V}$ is a collection of subsets of U. An example is shown in Fig. 2.3.



$$U = \{1, 2, 3, 4, 5, 6, 7\}$$

$$\mathcal{V} = \{e_1, e_2, e_3, e_4\} = \{\{1, 2, 3\}, \{2, 3\}, \{3, 5, 6\}, \{4\}\}$$

Figure 2.3: An example of hypergraph representation of a set U and its collections of subsets $\mathcal{V}$. Here, the elements in U are vertices and the subsets are edges, each of which is represented by distinct color. Two vertices are connected if they belong to the same subset.

A function $g : \mathcal{V} \to \mathbb{R}^+$ is called a *fractional cover* of $(U, \mathcal{V})$ if it satisfies

$$\sum_{\{V \in \mathcal{V} : V \ni i\}} g(V) \geq 1, \quad \forall i \in U. \tag{2.77}$$

Theorem 4 (A subset entropy inequality of Madiman and Tetali [6])**.** *Let $(U, \mathcal{V})$ be a hypergraph, and let g be a fractional cover of $(U, \mathcal{V})$. Then*

$$\sum_{V \in \mathcal{V}} g(V) H(X_V) \geq H(X_U) \tag{2.78}$$

for any collection of jointly distributed random variables X_U.

The proof of Theorem 4 can be found in Appendix A. The following corollary provides a "chain" form of the subset entropy inequality (2.78). Let M be a positive integer, and let Σ be a finite ground set. Let $\Sigma^{(\alpha)}$ be a collection of subsets of Σ for each $\alpha = 1, \ldots, M,$. Assuming that $\Sigma^{(\alpha)}$, $\alpha = 1, \ldots, M$, are *mutually exclusive*, then $\{\Sigma^{(\alpha)} : \alpha = 1, \ldots, M\}$ *induces* a collection of hypergraphs $\{(U, \mathcal{V}_U) : U \in \cup_{\alpha=2}^{M} \Sigma^{(\alpha)}\}$ where

$$\mathcal{V}_U := \{V \in \Sigma^{(\alpha-1)} : V \subseteq U\}, \quad \forall U \in \Sigma^{(\alpha)}. \tag{2.79}$$

We shall term each subset $V \in \mathcal{V}_U$ a "child" of U. For convenience, we shall also define

$$\mathcal{U}_V := \{U \in \Sigma^{(\alpha)} : U \supseteq V\}, \quad \forall V \in \Sigma^{(\alpha-1)} \tag{2.80}$$

and term each subset $U \in \mathcal{U}_V$ a "parent" of V.

Corollary 1. *Let $c : \cup_{\alpha=1}^{M} \Sigma^{(\alpha)} \to \mathbb{R}^+$. For any $\alpha = 2, \ldots, M$, if there exists a* *collection of functions* $\{g_U : U \in \Sigma^{(\alpha)}\}$ *for which each* g_U *is a fractional cover of* $(U, \mathcal{V}_U)$ *and such that*

$$c(V) = \sum_{U \in \mathcal{U}_V} g_U(V) c(U), \quad \forall V \in \Sigma^{(\alpha-1)} \tag{2.81}$$

we have

$$\sum_{V \in \Sigma^{(\alpha-1)}} c(V) H(X_V) \geq \sum_{U \in \Sigma^{(\alpha)}} c(U) H(X_U) \tag{2.82}$$

for any collection of jointly distributed random variables X_Σ.

Proof. Fix $\alpha \in \{2, \ldots, M\}$. For any $U \in \Sigma^{(\alpha)}$, g_U is a fractional cover of $(U, \mathcal{V}_U)$. By the subset entropy inequality of Madiman and Tetali (2.78), we have

$$\sum_{V \in \mathcal{V}_U} g_U(V) H(X_V) \geq H(X_U), \quad \forall U \in \Sigma^{(\alpha)}. \tag{2.83}$$

Multiplying both sides of (2.83) by $c(U)$ and summing over $U \in \Sigma^{(\alpha)}$, we have

$$\sum_{U \in \Sigma^{(\alpha)}} \sum_{V \in \mathcal{V}_U} c(U) g_U(V) H(X_V) \geq \sum_{U \in \Sigma^{(\alpha)}} c(U) H(X_U). \tag{2.84}$$

Note that

$$\begin{aligned} \sum_{U \in \Sigma^{(\alpha)}} \sum_{V \in \mathcal{V}_U} c(U) g_U(V) H(X_V) &= \sum_{V \in \Sigma^{(\alpha-1)}} \left(\sum_{U \in \mathcal{U}_V} g_U(V) c(U) \right) H(X_V) &(2.85) \\ &= \sum_{V \in \Sigma^{(\alpha-1)}} c(V) H(X_V) &(2.86) \end{aligned}$$

where (2.86) follows (2.81). Substituting (2.86) into (2.84) completes the proof of the corollary. □

2.3.2 Connections To The Subset Entropy Inequalities Of Han And Yeung–Zhang

Specifying $\Sigma = \Omega_L$, $M = L$, and $\Sigma^{(\alpha)} = \Omega_L^{(\alpha)}$ for $\alpha = 1, \ldots, L$, the subset entropy inequality of Madiman and Tetali can be used to provide a *unifying* proof for both the subset entropy inequality of Han and the subset entropy inequality of Yeung and Zhang. Note that the choice $\{\Sigma^{(\alpha)} = \Omega_L^{(\alpha)} : \alpha = 1, \ldots, L\}$ is *regular* in that each subset $U \in \Omega_L^{(\alpha)}$ has exactly α children in $\Omega_L^{(\alpha-1)}$, and each subset $V \in \Omega_L^{(\alpha-1)}$ has exactly $L - (\alpha - 1)$ parents in $\Omega_L^{(\alpha)}$.

To see how the subset entropy inequality of Madiman and Tetali (2.78) implies the subset entropy inequality of Han (2.34), let

$$c(U) := \frac{1}{\alpha\binom{L}{\alpha}}, \quad \forall U \in \Omega_L^{(\alpha)} \text{ and } \alpha = 1, \ldots, L \tag{2.87}$$

and

$$g_U(V) := \frac{1}{\alpha - 1}, \quad \forall U \in \Omega_L^{(\alpha)},\ V \in \mathcal{V}_U, \text{ and } \alpha = 2, \ldots, L. \tag{2.88}$$

For any $\alpha = 2, \ldots, L$ and $U \in \Omega_L^{(\alpha)}$,

$$\sum_{\{V \in \mathcal{V}_U : V \ni i\}} g_U(V) = \frac{|\{V \in \mathcal{V}_U : V \ni i\}|}{\alpha - 1} = \frac{\alpha - 1}{\alpha - 1} = 1, \quad \forall i \in U \tag{2.89}$$

so g_U is a *uniform* fractional cover of $(U, \mathcal{V}_U)$. Furthermore, for any $\alpha = 2, \ldots, L$ and $V \in \Omega_L^{(\alpha-1)}$ we have

$$\sum_{U \in \mathcal{U}_V} g_U(V)c(U) = \frac{|\mathcal{U}_V|}{(\alpha - 1)\alpha\binom{L}{\alpha}} = \frac{L - (\alpha - 1)}{(\alpha - 1)\alpha\binom{L}{\alpha}} = \frac{1}{(\alpha - 1)\binom{L}{\alpha-1}} = c(V). \tag{2.90}$$

Substituting (2.87) into (2.82) immediately gives the subset entropy inequality of

Han (2.34).

To see how the subset entropy inequality of Madiman and Tetali (2.78) implies the subset entropy inequality of Yeung and Zhang (2.39), we shall need the following result, which is a synthesis of the analytical results on the sequence of linear programs (2.19) established in [5]. (For completeness, a sketched proof based on the results of [5] is included in Appendix B.)

Theorem 5 (A linear programing result of Yeung and Zhang [5])**.** *For any* $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$*, any* $\alpha = 2, \ldots, L$*, and any* $c_{\boldsymbol{\lambda}}^{(\alpha)}$ *which is an optimal solution to the linear program* (2.19) *with the optimal value* $f_\alpha(\boldsymbol{\lambda}) > 0$*, there exists a collection of functions* $\{g_U : U \in \Omega_L^{(\alpha)}\}$ *for which each* g_U *is a fractional cover of* $(U, \mathcal{V}_U)$ *and such that* $c_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{c_{\boldsymbol{\lambda}}(V) : V \in \Omega_L^{(\alpha-1)}\}$ *where*

$$c_{\boldsymbol{\lambda}}(V) := \sum_{U \in \mathcal{U}_V} g_U(V) c_{\boldsymbol{\lambda}}(U) \tag{2.91}$$

is an optimal solution to the linear program (2.19) *with* α *replaced by* $\alpha - 1$*.*

Now fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and consider the following construction of $c_{\boldsymbol{\lambda}} = \cup_{\alpha=1}^{L} c_{\boldsymbol{\lambda}}^{(\alpha)}$. For $\alpha = L$, choose $c_{\boldsymbol{\lambda}}^{(L)}$ to be an arbitrary *optimal* solution to the linear program (2.19). For $\alpha = 1, \ldots, L-1$, construct $c_{\boldsymbol{\lambda}}^{(\alpha)}$ iteratively as follows. Suppose that $c_{\boldsymbol{\lambda}}^{(\alpha)}$ is already in place for some $\alpha = 2, \ldots, L$ such that $c_{\boldsymbol{\lambda}}^{(\alpha)}$ is an optimal solution to the linear program (2.19). If the optimal value $f_\alpha(\boldsymbol{\lambda}) > 0$, construct $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ according to (2.81) so $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ is an optimal solution to the linear program (2.19) with α replaced by $\alpha - 1$. Moreover, by Corollary 1 $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ and $c_{\boldsymbol{\lambda}}^{(\alpha)}$ satisfy the subset entropy inequality of Yeung and Zhang (2.39). If, on the other hand, $f_\alpha(\boldsymbol{\lambda}) = 0$, we have $c_{\boldsymbol{\lambda}}(U) = 0$ for *all* $U \in \Omega_L^{(\alpha)}$. In this case, choose $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ to be an arbitrary *optimal* solution to the linear program (2.19) with α replaced by $\alpha - 1$, and $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ and $c_{\boldsymbol{\lambda}}^{(\alpha)}$ will trivially satisfy the subset entropy inequality of Yeung and Zhang (2.39). We have thus constructed for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, a sequence of $c_{\boldsymbol{\lambda}}^{(\alpha)}$, $\alpha = 1, \ldots, L$, such that each $c_{\boldsymbol{\lambda}}^{(\alpha)}$ is an optimal solution to the linear program (2.19), and the subset entropy inequality of Yeung and Zhang (2.39) holds for each $\alpha = 2, \ldots, L$.

We mention here that even though both the subset entropy inequality of Han and the subset entropy inequality of Yeung and Zhang can be directly established from the subset entropy inequality of Madiman and Tetali, this is *not* the case for the sliding-window subset entropy inequality (2.47) except for $\alpha = 2$ and L. This can be seen as follows.

Let $\Sigma = \Omega_L$, $M = L$, and $\Sigma^{(\alpha)} = \{W_l^{(\alpha)} : l = 1, \ldots, L\}$ for $\alpha = 1, \ldots, L$. Note that for any $\alpha = 1, \ldots, L-1$, each sliding window $W_l^{(\alpha)}$ represents a *different* subset for different l. (For $\alpha = L$, all sliding windows $W_l^{(L)}$, $l = 1, \ldots, L$, represent the *same* subset Ω_L.) Furthermore, for any $\alpha = 2, \ldots, L-1$ each sliding window $W_l^{(\alpha)}$ has only two children: $W_l^{(\alpha-1)}$ and $W_{\langle l+1 \rangle}^{(\alpha-1)}$, and each sliding window $W_l^{(\alpha-1)}$ has only two parents: $W_l^{(\alpha)}$ and $W_{\langle l-1 \rangle}^{(\alpha)}$. Now consider the elements l and $\langle l + \alpha - 1 \rangle$ from $W_l^{(\alpha)}$. Note that among the two children $W_l^{(\alpha-1)}$ and $W_{\langle l+1 \rangle}^{(\alpha-1)}$ of $W_l^{(\alpha)}$, l belongs only to $W_l^{(\alpha-1)}$, and $\langle l + \alpha - 1 \rangle$ belong only to $W_{\langle l+1 \rangle}^{(\alpha-1)}$. Thus, *any* fractional cover $g_{W_l^{(\alpha)}}$ of the hypergraph $(W_l^{(\alpha)}, \{W_l^{(\alpha-1)}, W_{\langle l+1 \rangle}^{(\alpha-1)}\})$ must satisfy

$$g_{W_l^{(\alpha)}}(W_l^{(\alpha-1)}) \geq 1 \quad \text{and} \quad g_{W_l^{(\alpha)}}(W_{\langle l+1 \rangle}^{(\alpha-1)}) \geq 1. \tag{2.92}$$

Now let $c(W_l^{(\alpha)}) := 1/\alpha$ for all $l = 1, \ldots, L$ and $\alpha = 1, \ldots, L-1$. We have

$$g_{W_l^{(\alpha)}}(W_l^{(\alpha-1)}) c(W_l^{(\alpha)}) + g_{W_{\langle l-1 \rangle}^{(\alpha)}}(W_l^{(\alpha-1)}) c(W_{\langle l-1 \rangle}^{(\alpha)}) \geq \frac{2}{\alpha} > \frac{1}{\alpha - 1} = c(W_l^{(\alpha-1)}) \tag{2.93}$$

for any $\alpha > 2$. We thus conclude that for any $2 < \alpha < L$, the sliding-window subset entropy inequality (2.47) cannot be directly inferred from the subset entropy inequality of Madiman and Tetali.

2.3.3 A Conditional Subset Entropy Inequality Of Yeung and Zhang

We conclude this section by providing a conditional extension of the subset entropy inequality of Yeung and Zhang, which will play a key role in proving the optimality of superposition coding for achieving the entire admissible rate region of the general S-SMDC problem. We shall start with the following generalization of Corollary 1.

Let Σ be a finite ground set, and let $\Sigma^{(\alpha)}$, $\alpha = 1, \ldots, M$, be a collection of subsets of Σ. As before, we shall assume that the collections $\Sigma^{(\alpha)}$, $\alpha = 1, \ldots, M$, are mutually exclusive, so $\{\Sigma^{(\alpha)} : \alpha = 1, \ldots, M\}$ induces a hypergraph $(U, \mathcal{V}_U)$ for every $U \in \cup_{\alpha=2}^{M} \Sigma^{(\alpha)}$. For each $U \in \Sigma^{(M)}$ let $\mathcal{A}_U$ be a collection of subsets of Σ, and let $\mathcal{A}^{(M)} := \{\mathcal{A}_U : U \in \Sigma^{(M)}\}$. For $\alpha = 1, \ldots, M-1$, define $\mathcal{A}^{(\alpha)} := \{\mathcal{A}_U : U \in \Sigma^{(\alpha)}\}$ iteratively as follows. Suppose that $\mathcal{A}^{(\alpha)}$ is already in place for some $\alpha = 2, \ldots, M$. Let $\mathcal{A}^{(\alpha-1)} = \{\mathcal{A}_V : V \in \Sigma^{(\alpha-1)}\}$ where

$$\mathcal{A}_V := \cup_{U \in \mathcal{U}_V} \mathcal{A}_U. \tag{2.94}$$

Proposition 1. *For each $U \in \cup_{\alpha=1}^{M} \Sigma^{(\alpha)}$, let $s(U, \cdot) : \mathcal{A}_U \to \mathbb{R}^+$. For any $\alpha = 2, \ldots, M$, if there exists a collection of functions $\{g_U : U \in \Sigma^{(\alpha)}\}$ for which each g_U is a fractional cover of $(U, \mathcal{V}_U)$ and such that*

$$s(V, A) = \sum_{\{U \in \mathcal{U}_V : \mathcal{A}_U \ni A\}} g_U(V) s(U, A), \quad \forall V \in \Sigma^{(\alpha-1)} \text{ and } A \in \mathcal{A}_V \tag{2.95}$$

we have

$$\sum_{V\in\Sigma^{(\alpha-1)}}\sum_{A\in\mathcal{A}_V} s(V,A)H(X_V|X_A) \geq \sum_{U\in\Sigma^{(\alpha)}}\sum_{A\in\mathcal{A}_U} s(U,A)H(X_U|X_A) \tag{2.96}$$

for any collection of jointly distributed random variables X_Σ.

Proof. Fix $\alpha \in \{2,\dots,M\}$. For any $U \in \Sigma^{(\alpha)}$, g_U is a fractional cover of $(U,\mathcal{V}_U)$. By the subset entropy inequality of Madiman and Tetali (2.78), we have

$$\sum_{V\in\mathcal{V}_U} g_U(V)H(X_V|X_A) \geq H(X_U|X_A), \quad \forall U \in \Sigma^{(\alpha)} \text{ and } A \in \mathcal{A}_U. \tag{2.97}$$

Multiplying both sides of (2.97) by $s(U,A)$ and summing over $A \in \mathcal{A}_U$ and $U \in \Sigma^{(\alpha)}$, we have

$$\sum_{U\in\Sigma^{(\alpha)}}\sum_{A\in\mathcal{A}_U}\sum_{V\in\mathcal{V}_U} s(U,A)g_U(V)H(X_V|X_A) \geq \sum_{U\in\Sigma^{(\alpha)}}\sum_{A\in\mathcal{A}_U} s(U,A)H(X_U|X_A). \tag{2.98}$$

Note that

$$\begin{aligned}
&\sum_{U\in\Sigma^{(\alpha)}}\sum_{A\in\mathcal{A}_U}\sum_{V\in\mathcal{V}_U} s(U,A)g_U(V)H(X_V|X_A)\\
&\qquad= \sum_{U\in\Sigma^{(\alpha)}}\sum_{V\in\mathcal{V}_U}\sum_{A\in\mathcal{A}_U} s(U,A)g_U(V)H(X_V|X_A) &(2.99)\\
&\qquad= \sum_{V\in\Sigma^{(\alpha-1)}}\sum_{U\in\mathcal{U}_V}\sum_{A\in\mathcal{A}_U} s(U,A)g_U(V)H(X_V|X_A) &(2.100)\\
&\qquad= \sum_{V\in\Sigma^{(\alpha-1)}}\sum_{A\in\mathcal{A}_V}\left(\sum_{\{U\in\mathcal{U}_V:\mathcal{A}_U\ni A\}} s(U,A)g_U(V)\right)H(X_V|X_A) &(2.101)\\
&\qquad= \sum_{V\in\Sigma^{(\alpha-1)}}\sum_{A\in\mathcal{A}_V} s(V,A)H(X_V|X_A) &(2.102)
\end{aligned}$$

where (2.102) follows from (2.95). Substituting (2.102) into (2.98) completes the proof of the proposition. □

Theorem 6 (A conditional subset entropy inequality of Yeung and Zhang)**.** *For any* $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ *and* $N = 0, \ldots, L-1$*, there exists for each* $U \in \cup_{\alpha=1}^{L-N} \Omega_L^{(\alpha)}$ *a collection of subsets* $\mathcal{A}_U$ *of* Ω_L *such that:*

$$|A| = N \quad \textit{and} \quad A \cap U = \emptyset, \quad \forall A \in \mathcal{A}_U \tag{2.103}$$

and a function $s_{\boldsymbol{\lambda}}(U, \cdot) : \mathcal{A}_U \to \mathbb{R}^+$ *such that:*

1) for each $\alpha = 1, \ldots, L-N$, $c_{\boldsymbol{\lambda}}^{(\alpha)} = \{c_{\boldsymbol{\lambda}}(U) : U \in \Omega_L^{(\alpha)}\}$ *where*

$$c_{\boldsymbol{\lambda}}(U) := \sum_{A \in \mathcal{A}_U} s_{\boldsymbol{\lambda}}(U, A) \tag{2.104}$$

is an optimal *solution to the linear program* (2.19)*; and*

2) for each $\alpha = 2, \ldots, L-N$,

$$\sum_{V \in \Omega_L^{(\alpha-1)}} \sum_{A \in \mathcal{A}_V} s(V, A) H(X_V | X_A) \geq \sum_{U \in \Omega_L^{(\alpha)}} \sum_{A \in \mathcal{A}_U} s(U, A) H(X_U | X_A) \tag{2.105}$$

for any *collection of* L *jointly distributed random variables* $(X_1, \ldots, X_L)$.

Proof. Fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ and $N \in \{0, \ldots, L-1\}$, and let $\Sigma = \Omega_L$, $M = L - N$, and $\Sigma^{(\alpha)} = \Omega_L^{(\alpha)}$ for $\alpha = 1, \ldots, L-N$. Consider the following construction of $\mathcal{A}^{(\alpha)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)} := \{s(U, \cdot) : U \in \Omega_L^{(\alpha)}\}$, $\alpha = 1, \ldots, L-N$.

For $\alpha = L - N$, let $\mathcal{A}^{(L-N)} = \{\mathcal{A}_U : U \in \Omega_L^{(L-N)}\}$ where $\mathcal{A}_U := \{\Omega_L \setminus U\}$, i.e., each $\mathcal{A}_U$ contains a single subset $A = \Omega_L \setminus U$ of size $|A| = L - (L-N) = N$ and such that $A \cap U = \emptyset$. Furthermore, let $c_{\boldsymbol{\lambda}}^{(L-N)} = \{c_{\boldsymbol{\lambda}}(U) : U \in \Omega_L^{(L-N)}\}$ be an *optimal* solution to the linear program (2.19) for $\alpha = L - N$, and let

$$s_{\boldsymbol{\lambda}}(U, \Omega_L \setminus U) := c_{\boldsymbol{\lambda}}(U), \quad \forall U \in \Omega_L^{(L-N)}. \tag{2.106}$$

Since by construction each $\mathcal{A}_U$, $U \in \Omega_L^{(L-N)}$, contains a *single* subset $A = \Omega_L \setminus U$, we trivially have

$$\sum_{A \in \mathcal{A}_U} s_{\boldsymbol{\lambda}}(U, A) = c_{\boldsymbol{\lambda}}(U), \quad \forall U \in \Omega_L^{(L-N)}. \tag{2.107}$$

For $\alpha = 1, \ldots, L - N - 1$, let us construct $\mathcal{A}^{(\alpha)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)}$ iteratively as follows. Suppose that $\mathcal{A}^{(\alpha)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)}$ are already in place for some $\alpha = 2, \ldots, L - N$ such that $|A| = N$ and $A \cap U = \emptyset$ for any $U \in \Omega_L^{(\alpha)}$ and $A \in \mathcal{A}_U$, and $c_{\boldsymbol{\lambda}}^{(\alpha)} = \{c_{\boldsymbol{\lambda}}(U) : U \in \Omega_L^{(\alpha)}\}$ where $c_{\boldsymbol{\lambda}}(U)$ is given by (2.104) is an *optimal* solution to the linear program (2.19). First, construct $\mathcal{A}^{(\alpha-1)}$ according to (2.94). Based on this construction, for any $V \in \Omega_L^{(\alpha-1)}$ and $A \in \mathcal{A}_V$ we have $A \in \mathcal{A}_U$ for some $U \in \mathcal{U}_V \subseteq \Omega_L^{(\alpha)}$. Therefore, by the induction assumption we must have $|A| = N$ and

$$A \cap V \subseteq A \cap U = \emptyset \tag{2.108}$$

for any $V \in \Omega_L^{(\alpha-1)}$ and $A \in \mathcal{A}_V$. Next, construct $s_{\boldsymbol{\lambda}}^{(\alpha-1)}$ as follows. If the optimal value $f_\alpha(\boldsymbol{\lambda}) > 0$, by Theorem 5 there exists a collection of functions $\{g_U : U \in \Omega_L^{(\alpha)}\}$ for which each g_U is a fractional cover of $(U, \mathcal{V}_U)$ and such that $c_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{c_{\boldsymbol{\lambda}}(V) : V \in \Omega_L^{(\alpha-1)}\}$ where $c_{\boldsymbol{\lambda}}(V)$ is given by (2.91) is an *optimal* solution to the linear program (2.19) with α replaced by $\alpha - 1$. In this case, let $s_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{s_{\boldsymbol{\lambda}}(V, \cdot) : V \in \Omega_L^{(\alpha-1)}\}$

where

$$s_{\boldsymbol{\lambda}}(V,A) := \sum_{\{U\in\mathcal{U}_V:\mathcal{A}_U\ni A\}} g_U(V)s_{\boldsymbol{\lambda}}(U,A). \tag{2.109}$$

Thus, for each $V \in \Omega_L^{(\alpha-1)}$ we have

$$\begin{align}\sum_{A\in\mathcal{A}_V} s_{\boldsymbol{\lambda}}(V,A) &= \sum_{A\in\mathcal{A}_V}\left[\sum_{\{U\in\mathcal{U}_V:\mathcal{A}_U\ni A\}} g_U(V)s_{\boldsymbol{\lambda}}(U,A)\right] \tag{2.110}\\ &= \sum_{U\in\mathcal{U}_V} g_U(V)\left[\sum_{A\in\mathcal{A}_U} s_{\boldsymbol{\lambda}}(U,A)\right] \tag{2.111}\\ &= \sum_{U\in\mathcal{U}_V} g_U(V)c_{\boldsymbol{\lambda}}(U) \tag{2.112}\\ &= c_{\boldsymbol{\lambda}}(V) \tag{2.113}\end{align}$$

Furthermore, by Proposition 1 $s_{\boldsymbol{\lambda}}^{(\alpha-1)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)}$ satisfy the subset entropy inequality (2.105). If, on the other hand, $f_\alpha(\boldsymbol{\lambda}) = 0$, we have $s_{\boldsymbol{\lambda}}(U,A) = 0$ for *all* $U \in \Omega_L^{(\alpha)}$ and $A \in \mathcal{A}_U$. In this case, choose an arbitrary $s_{\boldsymbol{\lambda}}^{(\alpha-1)}$ such that $c_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{c_{\boldsymbol{\lambda}}(V) : V \in \Omega_L^{(\alpha-1)}\}$ where

$$c_{\boldsymbol{\lambda}}(V) := \sum_{A\in\mathcal{A}_V} s_{\boldsymbol{\lambda}}(V,A) \tag{2.114}$$

is an optimal solution to the linear program (2.19) with α being replaced by $\alpha-1$, and $s_{\boldsymbol{\lambda}}^{(\alpha-1)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)}$ will trivially satisfy the subset entropy inequality (2.105).

We have thus constructed for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ and $N = 0,\ldots,L-1$, a sequence of $\mathcal{A}^{(\alpha)}$ and $c_{\boldsymbol{\lambda}}^{(\alpha)}$, $\alpha = 1,\ldots,L-N$, such that all conditions of Theorem 6 are met simultaneously. This completes the proof of the theorem. □

3. SYMMETRICAL MULTILEVEL DIVERSITY CODING WITH AN ALL-ACCESS ENCODER

3.1 Problem Statement

As illustrated in Figure 3.1, the problem of SMDC-A consists of:

- a total of L *independent* discrete memoryless sources $\{S_\alpha[t]\}_{t=1}^{\infty}$, where $\alpha = 1, \ldots, L$ and t is the time index;
- a set of $L+1$ encoders (encoder 0 to L);
- a decoder who has access to a subset $\{0\} \cup U$ of the encoder outputs for some nonempty $U \subseteq \Omega_L$.

The realization of U is *unknown* a priori at the encoders. However, no matter which U actually materializes, the decoder needs to nearly perfectly reconstruct the sources $(S_1, \ldots, S_\alpha)$ whenever $|U| \geq \alpha$.

Formally, an $(n, (M_0, M_1, \ldots, M_L))$ code is defined by a collection of $L+1$ encoding functions

$$e_l : \prod_{\alpha=1}^{L} \mathcal{S}_\alpha^n \rightarrow \{1, \ldots, M_l\}, \quad \forall l = 0, 1, \ldots, L \tag{3.1}$$

and $2^L - 1$ decoding functions

$$d_U : \{1, \ldots, M_0\} \times \prod_{l \in U} \{1, \ldots, M_l\} \rightarrow \prod_{\alpha=1}^{|U|} \mathcal{S}_\alpha^n, \quad \forall U \subseteq \Omega_L \text{ s.t. } U \neq \emptyset. \tag{3.2}$$

A nonnegative rate tuple $(R_0, R_1, \ldots, R_L)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block length n, an $(n, (M_0, M_1, \ldots, M_L))$ code such

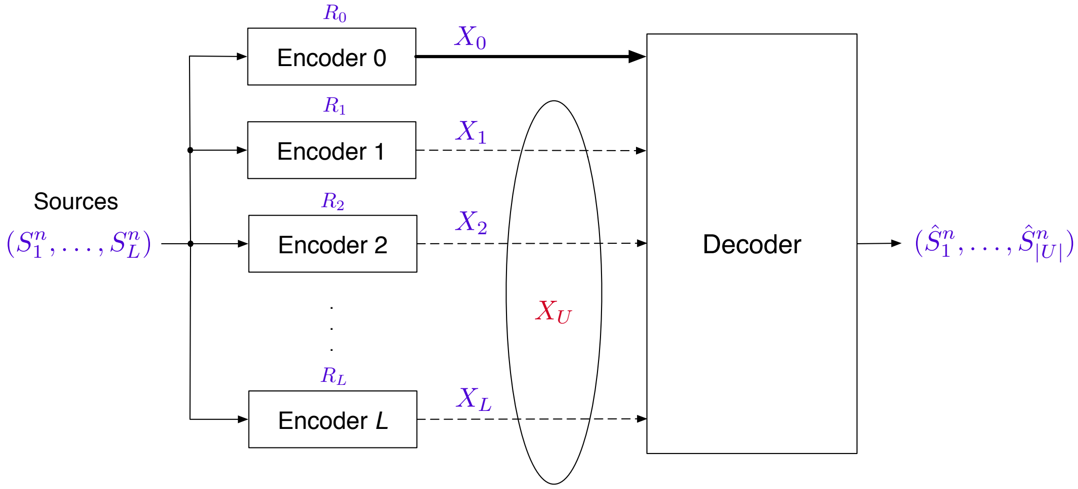


Figure 3.1: SMDC with an all-access encoder 0 and L randomly accessible encoders 1 to L. A total of L independent discrete memoryless sources $(S_1, \ldots, S_L)$ are to be encoded at the encoders. The decoder, which has access to encoder 0 and a subset U of the randomly accessible encoders, needs to nearly perfectly reconstruct the sources $(S_1, \ldots, S_{|U|})$ no matter what the realization of U is.

that:

- (Rate constraints at the encoders)

$$\frac{1}{n} \log M_l \leq R_l + \epsilon, \qquad \forall l = 0, 1, \ldots, L; \tag{3.3}$$

- (Asymptotically perfect reconstructions at the decoder)

$$\Pr\left\{d_U(X_{\{0\}\cup U}) \neq (S_1^n, \ldots, S_{|U|}^n)\right\} \leq \epsilon, \qquad \forall U \subseteq \Omega_L \text{ s.t. } U \neq \emptyset \tag{3.4}$$

  where $S_\alpha^n := \{S_\alpha[t]\}_{t=1}^n$, $X_l := e_l(S_1^n, \ldots, S_L^n)$ is the output of encoder l, and $X_{\{0\}\cup U} := \{X_l : l \in \{0\} \cup U\}$.

The *admissible rate region* $\mathcal{R}$ is the collection of *all* admissible rate tuples $(R_0, R_1, \ldots, R_L)$.

3.2 Superposition Coding Rate Region

Similar to classical SMDC, a natural strategy for SMDC-A is superposition coding, i.e., to encode the sources separately at the encoders and there is no coding across different sources. Formally, the problem of encoding a single source S_α can be viewed as a special case of the general problem where the sources S_m are deterministic for all $m \neq \alpha$. In this case, the source S_α needs to be nearly perfectly reconstructed whenever the decoder can access at least α randomly accessible encoders in addition to the all-access encoder 0. The following scheme is a natural extension of the simple source-channel separation scheme considered previously for classical SMDC:

- First compress the source sequence S_α^n into a source message W_α using a *lossless* source code. It is well known [8, Ch. 5] that the rate of the source message W_α can be made arbitrarily close to the entropy rate $H(S_\alpha)$ for sufficiently large block length n.

- Next, divide the source message W_α into two independent sub-messages $W_\alpha^{(0)}$ and $W_\alpha^{(1)}$ so we have

$$H(W_\alpha) = H(W_\alpha^{(0)}) + H(W_\alpha^{(1)}). \tag{3.5}$$

  The sub-message $W_\alpha^{(0)}$ is stored at the all-access encoder 0 *without* any coding, which requires

$$R_0 \geq \frac{1}{n} H(W_\alpha^{(0)}). \tag{3.6}$$

  The sub-message $W_\alpha^{(1)}$ is encoded by the randomly accessible encoders 1 to L using a *maximum distance separable* code [9]. Clearly, the sub-message $W_\alpha^{(1)}$

can be perfectly recovered at the decoder whenever

$$\sum_{l \in U} R_l \geq \frac{1}{n} H(W_\alpha^{(1)}), \quad \forall U \in \Omega_L^{(\alpha)} \tag{3.7}$$

for sufficiently large block length n. Eliminating $H(W_\alpha^{(0)})$ and $H(W_\alpha^{(1)})$ from (3.5)–(3.7), we conclude that the source message W_α can be perfectly recovered at the decoder whenever

$$R_0 + \sum_{l \in U} R_l \geq \frac{1}{n} H(W_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \tag{3.8}$$

Combining the above two steps, we conclude that the rate region that can be achieved by the above source-channel separation scheme is given by the collection of all nonnegative rate tuples $(R_0, R_1, \ldots, R_L)$ satisfying

$$R_0 + \sum_{l \in U} R_l \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}. \tag{3.9}$$

Following the same footsteps as those for classical SMDC [1,2], it is straightforward to show that the above rate region is in fact the admissible rate region for encoding the single source S_α. By definition, the superposition coding rate region $\mathcal{R}_{sup}$ for SMDC-A is given by the collection of all nonnegative rate tuples $(R_0, R_1, \ldots, R_L)$ such that

$$R_l := \sum_{\alpha=1}^{L} r_l^{(\alpha)} \tag{3.10}$$

for some nonnegative $r_l^{(\alpha)}$, $\alpha = 1, \ldots, L$ and $l = 0, 1, \ldots, L$, satisfying

$$r_0^{(\alpha)} + \sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}. \tag{3.11}$$

Similar to classical SMDC, the superposition coding rate region $\mathcal{R}_{sup}$ for SMDC-A is a *polyhedron* with polyhedral cone being the nonnegative orthant in $\mathbb{R}^{L+1}$ and hence can be completely characterized by the supporting hyperplanes

$$\sum_{l=0}^{L} \lambda_l R_l \geq f(\lambda_0, \boldsymbol{\lambda}), \quad \forall \lambda_0 \geq 0 \text{ and } \boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_L) \in (\mathbb{R}^+)^L \tag{3.12}$$

where

$$f(\lambda_0, \boldsymbol{\lambda}) = \min_{(R_0, R_1, \ldots, R_L) \in \mathcal{R}_{sup}} \sum_{l=0}^{L} \lambda_l R_l \tag{3.13}$$

$$= \begin{array}{rl} \min & \sum_{l=0}^{L} \left(\sum_{\alpha=1}^{L} \lambda_l r_l^{(\alpha)} \right) \\ \text{subject to} & r_0^{(\alpha)} + \sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \text{ and } \alpha = 1, \ldots, L \\ & r_l^{(\alpha)} \geq 0, \quad \forall \alpha = 1, \ldots, L \text{ and } l = 0, \ldots, L. \end{array} \tag{3.14}$$

Clearly, the above optimization problem can be separated into the following L sub-optimization problems:

$$f(\lambda_0, \boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f'_\alpha(\lambda_0, \boldsymbol{\lambda}) \tag{3.15}$$

where

$$f'_\alpha(\lambda_0, \boldsymbol{\lambda}) = \begin{array}{rl} \min & \sum_{l=0}^{L} \lambda_l r_l^{(\alpha)} \\ \text{subject to} & r_0^{(\alpha)} + \sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)} \\ & r_l^{(\alpha)} \geq 0, \quad \forall l = 0, \ldots, L \end{array} \tag{3.16}$$

$$= \begin{array}{rl} \max & \left(\sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0, \boldsymbol{\lambda}}(U) \right) H(S_\alpha) \\ \text{subject to} & \sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0, \boldsymbol{\lambda}}(U) \leq \lambda_0 \\ & \sum_{\{U \in \Omega_L^{(\alpha)} : U \ni l\}} c_{\lambda_0, \boldsymbol{\lambda}}(U) \leq \lambda_l, \quad \forall l = 1, \ldots, L \\ & c_{\lambda_0, \boldsymbol{\lambda}}(U) \geq 0, \quad \forall U \in \Omega_L^{(\alpha)}. \end{array} \tag{3.17}$$

Here, (3.17) follows from the strong duality for linear programs. For any $\lambda_0 \geq 0$, $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and $\alpha = 1, \ldots, L$, let

$$f_\alpha(\lambda_0, \boldsymbol{\lambda}) \;:=\; \begin{array}{rl} \max & \sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0,\boldsymbol{\lambda}}(U) \\ \text{subject to} & \sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0,\boldsymbol{\lambda}}(U) \leq \lambda_0 \\ & \sum_{\{U \in \Omega_L^{(\alpha)} : U \ni l\}} c_{\lambda_0,\boldsymbol{\lambda}}(U) \leq \lambda_l, \quad \forall l = 1, \ldots, L \\ & c_{\lambda_0,\boldsymbol{\lambda}}(U) \geq 0, \quad \forall U \in \Omega_L^{(\alpha)}. \end{array} \tag{3.18}$$

We have $f'_\alpha(\lambda_0, \boldsymbol{\lambda}) = f_\alpha(\lambda_0, \boldsymbol{\lambda}) H(S_\alpha)$ and hence

$$f(\lambda_0, \boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f_\alpha(\lambda_0, \boldsymbol{\lambda}) H(S_\alpha) \tag{3.19}$$

for any $\lambda_0 \geq 0$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$.

Note that in the optimization problem (3.18), if the constraint $\sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0,\boldsymbol{\lambda}}(U) \leq \lambda_0$ is inactive, it can be removed from the program. In this case the optimal value $f_\alpha(\lambda_0, \boldsymbol{\lambda}) = f_\alpha(\boldsymbol{\lambda})$, where $f_\alpha(\boldsymbol{\lambda})$ is the optimal value of the linear program (2.19). On the other hand, if the constraint $\sum_{U \in \Omega_L^{(\alpha)}} c_{\lambda_0,\boldsymbol{\lambda}}(U) \leq \lambda_0$ is active, the optimal value $f_\alpha(\lambda_0, \boldsymbol{\lambda}) = \lambda_0$. Combing these two cases, we have

$$f_\alpha(\lambda_0, \boldsymbol{\lambda}) = \min(f_\alpha(\boldsymbol{\lambda}), \lambda_0), \quad \forall \alpha = 1, \ldots, L. \tag{3.20}$$

Substituting (3.19) and (3.20) into (3.12), we conclude that the superposition coding rate region $\mathcal{R}_{sup}$ for SMDC with an all-access encoder is given by the collection of all nonnegative rate tuples $(R_0, R_1, \ldots, R_L)$ satisfying

$$\sum_{l=0}^{L} \lambda_l R_l \geq \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) H(S_\alpha), \quad \forall \lambda_0 \geq 0 \text{ and } \boldsymbol{\lambda} \in (\mathbb{R}^+)^L. \tag{3.21}$$

As mentioned previously, the superposition coding rate region $\mathcal{R}_{sup}$ is a polyhedron, so among all $\lambda_0 \geq 0$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, most of the inequalities in (3.21) are *redundant*. Identifying those which define the *faces* of the superposition coding rate region $\mathcal{R}_{sup}$ appears to be very difficult. Note, however, that for any given $(R_0, R_1, \ldots, R_L) \in (\mathbb{R}^+)^{L+1}$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, the left-hand side of (3.21) is a linear, nondecreasing function of λ_0, and the right-hand side of (3.21) is a piecewise linear, nondecreasing, and concave function of λ_0. Thus, the left-hand side of (3.21) will dominate the right-hand side for *every* $\lambda_0 \geq 0$ if and only if it dominates the right-hand side at its boundary points $\lambda_0 = f_m(\boldsymbol{\lambda})$, $m = 1, \ldots, L$, between the adjacent line segments. See Figure 3.2 for an illustration.

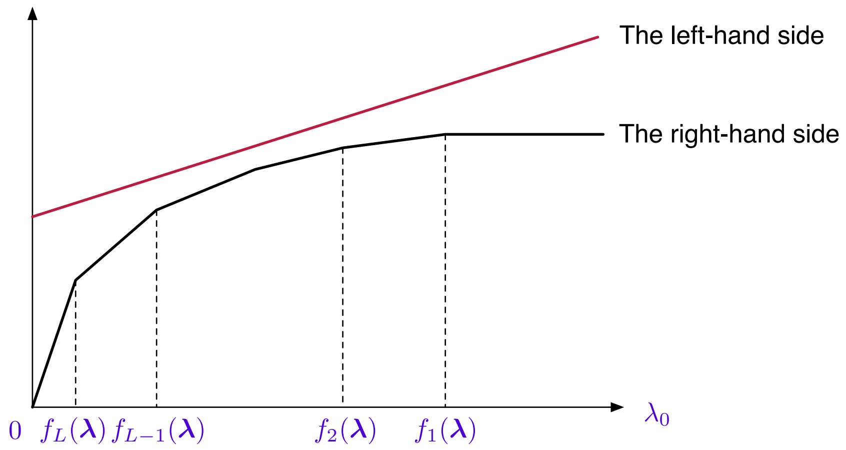


Figure 3.2: The left-hand and right-hand sides of (3.21) as a function of λ_0 for a fixed $(R_0, R_1, \ldots, R_L) \in (\mathbb{R}^+)^{L+1}$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$.

Formally, we have the following proposition, which plays a key role next in proving the optimality of superposition coding for achieving the entire admissible rate region of SMDC-A.

Proposition 2. *The superposition coding rate region $\mathcal{R}_{sup}$ for SMDC-A is given by the collection of all nonnegative rate tuples $(R_0, R_1, \ldots, R_L)$ satisfying*

$$\begin{aligned} & f_m(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \\ & \geq \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), f_m(\boldsymbol{\lambda}))H(S_\alpha) && (3.22) \\ & = f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha), \quad \forall m = 1, \ldots, L \textit{ and } \boldsymbol{\lambda} \in (\mathbb{R}^+)^L && (3.23) \end{aligned}$$

where $f_\alpha(\boldsymbol{\lambda})$ is the optimal value of the linear program (2.19).

Proof. Let us first recall the following results from [5]: for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ we have

$$f_1(\boldsymbol{\lambda}) \geq f_2(\boldsymbol{\lambda}) \geq \cdots \geq f_L(\boldsymbol{\lambda}) \geq 0. \tag{3.24}$$

It follows that

$$\begin{aligned} \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), f_m(\boldsymbol{\lambda}))H(S_\alpha) &= \sum_{\alpha=1}^{m} f_m(\boldsymbol{\lambda})H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha) && (3.25) \\ &= f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha). && (3.26) \end{aligned}$$

It remains to show that for any given $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, the set of inequalities (3.21) over all $\lambda_0 \geq 0$ is dominated by that over $\lambda_0 = f_m(\boldsymbol{\lambda})$ for $m = 1, \ldots, L$.

Fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and consider the following three cases separately.

Case 1: $\lambda_0 \geq f_1(\boldsymbol{\lambda})$. By (3.24), we have $\lambda_0 \geq f_\alpha(\boldsymbol{\lambda})$ and hence $\min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) = f_\alpha(\boldsymbol{\lambda})$ for any $\alpha = 1, \ldots, L$. For $m = 1$, the inequality (3.23) can be written as

$$f_1(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha) \tag{3.27}$$

which implies that

$$\lambda_0 R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_1(\boldsymbol{\lambda}) R_0 + \sum_{l=1}^{L} \lambda_l R_l \tag{3.28}$$

$$\geq \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.29}$$

$$= \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) H(S_\alpha) \tag{3.30}$$

for any $\lambda_0 \geq f_1(\boldsymbol{\lambda})$.

Case 2: $0 \leq \lambda_0 < f_L(\boldsymbol{\lambda})$. By (3.24), we have $\lambda_0 < f_\alpha(\boldsymbol{\lambda})$ and hence $\min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) = \lambda_0$ for any $\alpha = 1, \ldots, L$. For $m = L$, the inequality (3.23) can be written as

$$f_L(\boldsymbol{\lambda}) R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_L(\boldsymbol{\lambda}) \sum_{\alpha=1}^{L} H(S_\alpha) \tag{3.31}$$

which implies that

$$\lambda_0 R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq \frac{\lambda_0}{f_L(\boldsymbol{\lambda})} \left(f_L(\boldsymbol{\lambda}) R_0 + \sum_{l=1}^{L} \lambda_l R_l \right) \tag{3.32}$$

$$\geq \frac{\lambda_0}{f_L(\boldsymbol{\lambda})} \left(f_L(\boldsymbol{\lambda}) \sum_{\alpha=0}^{L} H(S_\alpha) \right) \tag{3.33}$$

$$= \lambda_0 \sum_{\alpha=0}^{L} H(S_\alpha) \tag{3.34}$$

$$= \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) H(S_\alpha) \tag{3.35}$$

for any $0 \leq \lambda_0 < f_L(\boldsymbol{\lambda})$.

Case 3: $f_{r+1}(\boldsymbol{\lambda}) \leq \lambda_0 < f_r(\boldsymbol{\lambda})$ for some $r = 1, \ldots, L-1$. By (3.24), we have $\lambda_0 < f_\alpha(\boldsymbol{\lambda})$ and hence $\min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) = \lambda_0$ for $\alpha = 1, \ldots, r$, and $\lambda_0 \geq f_\alpha(\boldsymbol{\lambda})$ and hence $\min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) = f_\alpha(\boldsymbol{\lambda})$ for $\alpha = r+1, \ldots, L$. For $m = r$ and $r+1$, the inequality (3.23) can be written as

$$f_r(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_r(\boldsymbol{\lambda}) \sum_{\alpha=1}^{r} H(S_\alpha) + \sum_{\alpha=r+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.36}$$

and

$$f_{r+1}(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_{r+1}(\boldsymbol{\lambda}) \sum_{\alpha=1}^{r+1} H(S_\alpha) + \sum_{\alpha=r+2}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.37}$$

respectively, which together imply that

$$\begin{aligned} \lambda_0 R_0 + \sum_{l=1}^{L} \lambda_l R_l &= \frac{\lambda_0 - f_{r+1}(\boldsymbol{\lambda})}{f_r(\boldsymbol{\lambda}) - f_{r+1}(\boldsymbol{\lambda})} \left(f_r(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \right) + \\ &\quad \frac{f_r(\boldsymbol{\lambda}) - \lambda_0}{f_r(\boldsymbol{\lambda}) - f_{r+1}(\boldsymbol{\lambda})} \left(f_{r+1}(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \right) \\ &\geq \frac{\lambda_0 - f_{r+1}(\boldsymbol{\lambda})}{f_r(\boldsymbol{\lambda}) - f_{r+1}(\boldsymbol{\lambda})} \left(f_r(\boldsymbol{\lambda}) \sum_{\alpha=1}^{r} H(S_\alpha) + \sum_{\alpha=r+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \right) + \\ &\quad \frac{f_r(\boldsymbol{\lambda}) - \lambda_0}{f_r(\boldsymbol{\lambda}) - f_{r+1}(\boldsymbol{\lambda})} \left(f_{r+1}(\boldsymbol{\lambda}) \sum_{\alpha=1}^{r+1} H(S_\alpha) + \sum_{\alpha=r+2}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \right) \\ &= \lambda_0 \sum_{\alpha=1}^{r} H(S_\alpha) + \sum_{\alpha=r+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \\ &= \sum_{\alpha=1}^{L} \min(f_\alpha(\boldsymbol{\lambda}), \lambda_0) H(S_\alpha) \end{aligned} \tag{3.38}$$

for any $f_{r+1}(\boldsymbol{\lambda}) \leq \lambda_0 < f_r(\boldsymbol{\lambda})$.

Combining these three cases completes the proof of the proposition. □

3.3 Optimality Of Superposition Coding

The main result of this section is that superposition coding remains optimal in terms of achieving the entire admissible rate region for SMDC-A, as summarized in the following theorem.

Theorem 7. *For the general SMDC-A problem, the admissible rate region*

$$\mathcal{R} = \mathcal{R}_{sup}. \tag{3.39}$$

Proof. Based on the discussions from Section 3.2, we naturally have $\mathcal{R}_{sup} \subseteq \mathcal{R}$. Thus, to show $\mathcal{R}_{sup} = \mathcal{R}$ we only need to show that $\mathcal{R} \subseteq \mathcal{R}_{sup}$. In light of Proposition 2, it is sufficient to show that *any* admissible rate tuple $(R_0, R_1, \ldots, R_L)$ must satisfy

$$f_m(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.40}$$

for *all* $m = 1, \ldots, L$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$.

Let $(R_0, R_1, \ldots, R_L)$ be an admissible rate tuple. By definition, for any sufficiently large block-length n there exists an $(n, (M_0, M_1, \ldots, M_L))$ code satisfying the rate constraints (3.3) for the admissible rate tuple $(R_0, R_1, \ldots, R_L)$ and the asymptotically perfect reconstruction requirement (3.4). Fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and let $\{c_{\boldsymbol{\lambda}}^{(\alpha)} : \alpha = 1, \ldots, L\}$ be a set of *optimal* solutions that satisfies the subset entropy inequality of Yeung and Zhang (2.39).

Note that for $\alpha = 1$, the optimal solution for the linear program (2.19) is *unique* and is given by

$$c_{\boldsymbol{\lambda}}(\{l\}) = \lambda_l, \quad \forall l = 1, \ldots, L. \tag{3.41}$$

We thus have for any $m = 1, \ldots, L$

$$\begin{aligned}
& n\left(f_m(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l\right) \\
& = f_m(\boldsymbol{\lambda})nR_0 + \sum_{l=1}^{L} c_{\boldsymbol{\lambda}}(\{l\})nR_l && (3.42) \\
& \geq f_m(\boldsymbol{\lambda})(H(X_0) - n\epsilon) + \sum_{l=1}^{L} c_{\boldsymbol{\lambda}}(\{l\})(H(X_l) - n\epsilon) && (3.43) \\
& = f_m(\boldsymbol{\lambda})H(X_0) + \sum_{l=1}^{L} c_{\boldsymbol{\lambda}}(\{l\})H(X_l) - n(f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon && (3.44) \\
& \geq f_m(\boldsymbol{\lambda})H(X_0) + \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U)H(X_U) - n(f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon && (3.45) \\
& = \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U)(H(X_0) + H(X_U)) - n(f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon && (3.46) \\
& \geq \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U)H(X_0, X_U) - n(f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon && (3.47)
\end{aligned}$$

where (3.43) follows from the rate constraint (3.3), (3.44) and (3.46) are due to the fact that $c_{\boldsymbol{\lambda}}^{(1)}$ and $c_{\boldsymbol{\lambda}}^{(m)}$ are optimal so we have

$$\sum_{l=1}^{L} c_{\boldsymbol{\lambda}}(\{l\}) = f_1(\boldsymbol{\lambda}) \quad \text{and} \quad \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U) = f_m(\boldsymbol{\lambda}) \tag{3.48}$$

(3.45) follows from the subset entropy inequality of Yeurng and Zhang (2.39) so we have

$$\sum_{l=1}^{L} c_{\boldsymbol{\lambda}}(\{l\})H(X_l) \geq \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U)H(X_U) \tag{3.49}$$

and (3.47) follows from the independence bound on entropy.

For any $U \in \Omega_L^{(m)}$ and $m = 1, \ldots, L$, by the asymptotically perfect reconstruction requirement (3.4) and the well-known Fano's inequality we have

$$H(S_1^n, \ldots, S_m^n | X_0, X_U) \leq n\delta_m^{(n)} \tag{3.50}$$

where $\delta_m^{(n)} \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$. By the chain rule for entropy,

$$\begin{aligned}
&H(X_0, X_U) \\
&= H(X_0, X_U, S_1^n, \ldots, S_m^n) - H(S_1^n, \ldots, S_m^n | X_0, X_U) && (3.51) \\
&= H(S_1^n, \ldots, S_m^n) + H(X_0, X_U | S_1^n, \ldots, S_m^n) - H(S_1^n, \ldots, S_m^n | X_0, X_U) && (3.52) \\
&= n \sum_{\alpha=1}^{m} H(S_\alpha) + H(X_0, X_U | S_1^n, \ldots, S_m^n) - H(S_1^n, \ldots, S_m^n | X_0, X_U) && (3.53) \\
&\geq n \sum_{\alpha=1}^{m} H(S_\alpha) + H(X_0, X_U | S_1^n, \ldots, S_m^n) - n\delta_m^{(n)} && (3.54)
\end{aligned}$$

where (3.53) is due to the fact that $S_1, \ldots, S_L$ are independent memoryless sources, and (3.54) follows from (3.50). Substituting (3.54) into (3.47), we have

$$\begin{aligned}
&n \left(f_m(\boldsymbol{\lambda}) R_0 + \sum_{l=1}^{L} \lambda_l R_l \right) \\
&\geq \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U) \left(n \sum_{\alpha=1}^{m} H(S_\alpha) + H(X_0, X_U | S_1^n, \ldots, S_m^n) - n\delta_m^{(n)} \right) - n(f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon \\
& && (3.55) \\
&= n f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U) H(X_0, X_U | S_1^n, \ldots, S_m^n) - \\
&\quad n(f_m(\boldsymbol{\lambda})\delta_m^{(n)} + (f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon). && (3.56)
\end{aligned}$$

Next, we show, via an induction on m, that for any $m = 1, \ldots, L$ we have

$$\sum_{U \in \Omega_L^{(m)}} c_{\boldsymbol{\lambda}}(U) H(X_0, X_U | S_1^n, \ldots, S_m^n) \geq n \left(\sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)} \right). \tag{3.57}$$

First consider the base case with $m = L$. In this case, the inequality (3.57) is trivial as the right-hand side of the inequality is zero. Next, assume that the inequality (3.57) holds for $m = l$ for some $l = 2, \ldots, L$, i.e,

$$\sum_{U \in \Omega_L^{(l)}} c_{\boldsymbol{\lambda}}(U) H(X_0, X_U | S_1^n, \ldots, S_l^n) \geq n \left(\sum_{\alpha=l+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - \sum_{\alpha=l+1}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)} \right). \tag{3.58}$$

For any $U \in \Omega_L^{(l)}$, we have

$$\begin{aligned}
& H(X_0, X_U | S_1^n, \ldots, S_l^n) \\
&\quad = H(X_0, X_U | S_1^n, \ldots, S_{l-1}^n) - I(S_l^n; X_0, X_U | S_1^n, \ldots, S_{l-1}^n) && (3.59) \\
&\quad = H(X_0, X_U | S_1^n, \ldots, S_{l-1}^n) - H(S_l^n | S_1^n, \ldots, S_{l-1}^n) + H(S_l^n | X_0, X_U, S_1^n, \ldots, S_{l-1}^n) && (3.60) \\
&\quad \leq H(X_0, X_U | S_1^n, \ldots, S_{l-1}^n) - H(S_l^n | S_1^n, \ldots, S_{l-1}^n) + H(S_l^n | X_0, X_U) && (3.61) \\
&\quad \leq H(X_0, X_U | S_1^n, \ldots, S_{l-1}^n) - H(S_l^n | S_1^n, \ldots, S_{l-1}^n) + \delta_l^{(n)} && (3.62) \\
&\quad = H(X_0, X_U | S_1^n, \ldots, S_{l-1}^n) - nH(S_l) + \delta_l^{(n)} && (3.63)
\end{aligned}$$

where (3.61) follows from the fact that conditioning reduces entropy, (3.62) follows the fact that

$$H(S_l^n | X_0, X_U) \leq H(S_1^n, \ldots, S_l^n | X_0, X_U) \leq n\delta_l^{(n)} \tag{3.64}$$

and (3.63) follows from the fact that $S_1, \ldots, S_L$ are independent memoryless sources.

Multiplying both sides of the inequality (3.63) by $c_{\boldsymbol{\lambda}}(U)$ and summing over all $U \in \Omega_L^{(l)}$, we have

$$\begin{aligned}
&\sum_{U\in\Omega_L^{(l)}} c_{\boldsymbol{\lambda}}(U)H(X_0,X_U|S_1^n,\dots,S_{l-1}^n)\\
&\qquad\geq \sum_{U\in\Omega_L^{(l)}} c_{\boldsymbol{\lambda}}(U)\left(H(X_0,X_U|S_1^n,\dots,S_l^n)+nH(S_l)-n\delta_l^{(n)}\right) && (3.65)\\
&\qquad= \sum_{U\in\Omega_L^{(l)}} c_{\boldsymbol{\lambda}}(U)H(X_0,X_U|S_1^n,\dots,S_l^n)+n\left(f_l(\boldsymbol{\lambda})H(S_l)-f_l(\boldsymbol{\lambda})n\delta_l^{(n)}\right) && (3.66)\\
&\qquad\geq n\left(\sum_{\alpha=l+1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-\sum_{\alpha=l+1}^{L} f_\alpha(\boldsymbol{\lambda})\delta_\alpha^{(n)}\right)+n\left(f_l(\boldsymbol{\lambda})H(S_l)-f_l(\boldsymbol{\lambda})\delta_l^{(n)}\right) && (3.67)\\
&\qquad= n\left(\sum_{\alpha=l}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-\sum_{\alpha=l}^{L} f_\alpha(\boldsymbol{\lambda})\delta_\alpha^{(n)}\right) && (3.68)
\end{aligned}$$

where (3.67) follows from the induction assumption (3.58). Finally, by the subset entropy inequality of Yeung and Zhang (2.39) we have

$$\begin{aligned}
\sum_{U\in\Omega_L^{(l-1)}} c_{\boldsymbol{\lambda}}(U)H(X_0,X_U|S_1^n,\dots,S_{l-1}^n) &\geq \sum_{U\in\Omega_L^{(l)}} c_{\boldsymbol{\lambda}}(U)H(X_0,X_U|S_1^n,\dots,S_{l-1}^n) && (3.69)\\
&\geq n\left(\sum_{\alpha=l}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-\sum_{\alpha=l}^{L} f_\alpha(\boldsymbol{\lambda})\delta_\alpha^{(n)}\right). && (3.70)
\end{aligned}$$

This proves that the inequality (3.57) also holds for $m = l-1$ and hence completes the proof of (3.57).

Substituting (3.57) into (3.56) and dividing both sides of the inequality by n, we have

$$f_m(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) - \left(\sum_{\alpha=m}^{L} f_\alpha(\boldsymbol{\lambda}) \delta_\alpha^{(n)} + (f_1(\boldsymbol{\lambda}) + f_m(\boldsymbol{\lambda}))\epsilon \right). \tag{3.71}$$

Letting $n \to \infty$ and $\epsilon \to 0$, we have from (3.71) that

$$f_m(\boldsymbol{\lambda})R_0 + \sum_{l=1}^{L} \lambda_l R_l \geq f_m(\boldsymbol{\lambda}) \sum_{\alpha=1}^{m} H(S_\alpha) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.72}$$

for any admissible rate tuple $(R_0, R_1, \ldots, R_L)$. This proves that $\mathcal{R} \subseteq \mathcal{R}_{sup}$ and hence completes the proof of the theorem. □

3.4 Rate Allocation At The All-access Encoder

In this section, we conclude our discussion on SMDC-A by focusing on a *greedy* rate allocation policy at the all-access encoder. Based on our previous discussion in Section 3.2, the output of the all-access encoder 0 consists of only *uncoded* information bits for the source messages $W_1, \ldots, W_L$. Hence, its storage efficiency is the same for each of the information sources $S_1, \ldots, S_L$. On the other hand, for the randomly accessible encoders 1 to L, S_1 has the highest reconstruction requirement and hence is the least efficient source to encode, and S_L has the lowest reconstruction requirement and hence is the most efficient source to encode. Therefore, intuitively, the greedy policy that assigns the remaining rate budget of the all-access encoder 0 to the least efficient source should be optimal.

More specifically, suppose that the rate budget R_0 of the all-access encoder 0 satisfies

$$\sum_{\alpha=1}^{q-1} H(S_\alpha) \leq R_0 < \sum_{\alpha=1}^{q} H(S_\alpha) \tag{3.73}$$

for some $q = 1, \ldots, L$. The greedy policy stores the source messages $W_1, \ldots, W_{q-1}$ in their entireties (without any coding) at the all-access encoder 0, and the residual rate budget $R_0 - \sum_{\alpha=1}^{q-1} H(S_\alpha)$ is then committed in full to the source message W_q. The residual source messages are W_q, with a residual rate

$$H(S_q) - \left(R_0 - \sum_{\alpha=1}^{q-1} H(S_\alpha)\right) = \sum_{\alpha=1}^{q} H(S_\alpha) - R_0 \tag{3.74}$$

and $W_{q+1}, \ldots, W_L$ with respective rates $H(S_{q+1}), \ldots, H(S_L)$. The residual source messages are encoded at the randomly accessible encoders using superposition coding, and the corresponding rate region $\mathcal{R}'_{sup}(R_0)$ is given by

$$\mathcal{R}'_{sup}(R_0) = \Bigg\{(R_1, \ldots, R_L) \in (\mathbb{R}^+)^L : \sum_{l=1}^{L} \lambda_l R_l \geq f_q(\boldsymbol{\lambda}) \left(\sum_{\alpha=1}^{q} H(S_\alpha) - R_0\right) + \sum_{\alpha=q+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha), \quad \forall \boldsymbol{\lambda} \in (\mathbb{R}^+)^L \Bigg\}. \tag{3.75}$$

Of course, when

$$R_0 \geq \sum_{\alpha=1}^{L} H(S_\alpha) \tag{3.76}$$

all source messages $W_1, \ldots, W_L$ can be stored at the all-access encoder 0 (without any coding), and there is no need to use the randomly access encoders 1 to L. In this case, we have $\mathcal{R}'_{sup}(R_0) = (\mathbb{R}^+)^L$.

To show that the aforementioned greedy rate allocation policy at the all-access encoder 0 is optimal, we need to show that $\mathcal{R}'_{sup}(R_0)$ *matches* the R_0-slice of the superposition coding rate region

$$\mathcal{R}_{sup}(R_0) := \left\{(R_1, \ldots, R_L) \in (\mathbb{R}^+)^L : (R_0, R_1, \ldots, R_L) \in \mathcal{R}_{sup}\right\} \tag{3.77}$$

for *all* $R_0 \geq 0$. By Proposition 2, for any $R_0 \geq 0$ the R_0-slice of the superposition coding rate region can be written as

$$\mathcal{R}_{sup}(R_0) = \left\{(R_1, \ldots, R_L) \in (\mathbb{R}^+)^L : \sum_{l=1}^{L} \lambda_l R_l \geq \max_{m=1,\ldots,L} \{g_m(\boldsymbol{\lambda})\}, \quad \forall \boldsymbol{\lambda} \in (\mathbb{R}^+)^L\right\} \tag{3.78}$$

where

$$g_m(\boldsymbol{\lambda}) := f_m(\boldsymbol{\lambda}) \left(\sum_{\alpha=1}^{m} H(S_\alpha) - R_0\right) + \sum_{\alpha=m+1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{3.79}$$

For any $m = 1, \ldots, L-1$, it is straightforward to calculate that

$$g_{m+1}(\boldsymbol{\lambda}) - g_m(\boldsymbol{\lambda}) = (f_{m+1}(\boldsymbol{\lambda}) - f_m(\boldsymbol{\lambda})) \left(\sum_{\alpha=1}^{m} H(S_\alpha) - R_0\right). \tag{3.80}$$

By (3.24), $f_{m+1}(\boldsymbol{\lambda}) - f_m(\boldsymbol{\lambda}) \leq 0$ for any $m = 1, \ldots, L-1$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$. Thus, when $\sum_{\alpha=1}^{q-1} H(S_\alpha) \leq R_0 < \sum_{\alpha=1}^{q} H(S_\alpha)$ for some $q = 1, \ldots, L$, we have $\sum_{\alpha=1}^{m} H(S_\alpha) - R_0 \geq 0$ and hence $g_{m+1}(\boldsymbol{\lambda}) - g_m(\boldsymbol{\lambda}) \leq 0$ for all $m = q, \ldots, L-1$, and $\sum_{\alpha=1}^{m} H(S_\alpha) - R_0 \leq 0$ and hence $g_{m+1}(\boldsymbol{\lambda}) - g_m(\boldsymbol{\lambda}) \geq 0$ for all $m = 1, \ldots, q-1$. We conclude that in this case,

$$\max_{m=1,\ldots,L}\{g_m(\boldsymbol{\lambda})\} = g_q(\boldsymbol{\lambda}) = f_q(\boldsymbol{\lambda})\left(\sum_{\alpha=1}^{q} H(S_\alpha) - R_0\right) + \sum_{\alpha=q+1}^{L} f_\alpha(\boldsymbol{\lambda})H(S_\alpha) \tag{3.81}$$

for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$ and hence $\mathcal{R}_{sup}(R_0) = \mathcal{R}'_{sup}(R_0)$. When$R_0 \geq \sum_{\alpha=1}^{L} H(S_\alpha)$, we have $\sum_{\alpha=1}^{m} H(S_\alpha) - R_0 \leq 0$ and hence $g_{m+1}(\boldsymbol{\lambda}) - g_m(\boldsymbol{\lambda}) \geq 0$ for all $m = 1, \ldots, L-1$. In this case,

$$\max_{m=1,\ldots,L}\{g_m(\boldsymbol{\lambda})\} = g_L(\boldsymbol{\lambda}) = f_L(\boldsymbol{\lambda})\left(\sum_{\alpha=1}^{L} H(S_\alpha) - R_0\right) \leq 0 \tag{3.82}$$

for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and we once again have $\mathcal{R}_{sup}(R_0) = \mathcal{R}'_{sup}(R_0)$. We summarize the above results in the following theorem.

Theorem 8. *Greedy rate allocation at the all-access encoder combined with superposition coding at the randomly accessible encoders can achieve the entire admissible rate region for the general SMDC-A problem.*

4. SECURE SYMMETRICAL MULTILEVEL DIVERSITY CODING

4.1 Problem Statement

Let L be a positive integer, and let $N \in \{0, \ldots, L-1\}$. Let $\{S_1[t], \ldots, S_{L-N}[t]\}_{t=1}^{\infty}$ be a collection of $L-N$ *independent* discrete memoryless sources with time index t, and let $S_\alpha^n := (S_\alpha[1], \ldots, S_\alpha[n])$ for $\alpha = 1, \ldots, L-N$. As illustrated in Figure 4.1, an (L, N) S-SMDC problem consists of a set of L encoders, a legitimate receiver who has access to a subset U of the encoder outputs, and an eavesdropper who has access to a subset A of the encoder outputs. Which subsets of the encoder outputs are available at the legitimate receiver and the eavesdropper are *unknown* a priori at the encoders. However, no matter which subsets U and A actually occur, the legitimate receiver must be able to asymptotically perfectly reconstruct the sources $(S_1, \ldots, S_\alpha)$ whenever $|U| \geq N + \alpha$, and all sources $(S_1, \ldots, S_{L-N})$ must be kept perfectly secure from the eavesdropper as long as $|A| \leq N$.

Formally, an $(n, (M_1, \ldots, M_L))$ code is defined by a collection of L encoding functions

$$e_l : \prod_{\alpha=1}^{L-N} \mathcal{S}_\alpha^n \times \mathcal{K} \rightarrow \{1, \ldots, M_l\}, \quad \forall l = 1, \ldots, L \tag{4.1}$$

and $\sum_{\alpha=N+1}^{L} \binom{L}{\alpha}$ decoding functions

$$d_U : \prod_{l \in U} \{1, \ldots, M_l\} \rightarrow \prod_{\alpha=1}^{|U|-N} \mathcal{S}_\alpha^n, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq N+1 \tag{4.2}$$

where $\mathcal{K}$ is the key space accessible to all L encoders. A nonnegative rate tuple $(R_1, \ldots, R_L)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block length n, an $(n, (M_1, \ldots, M_L))$ code such that:

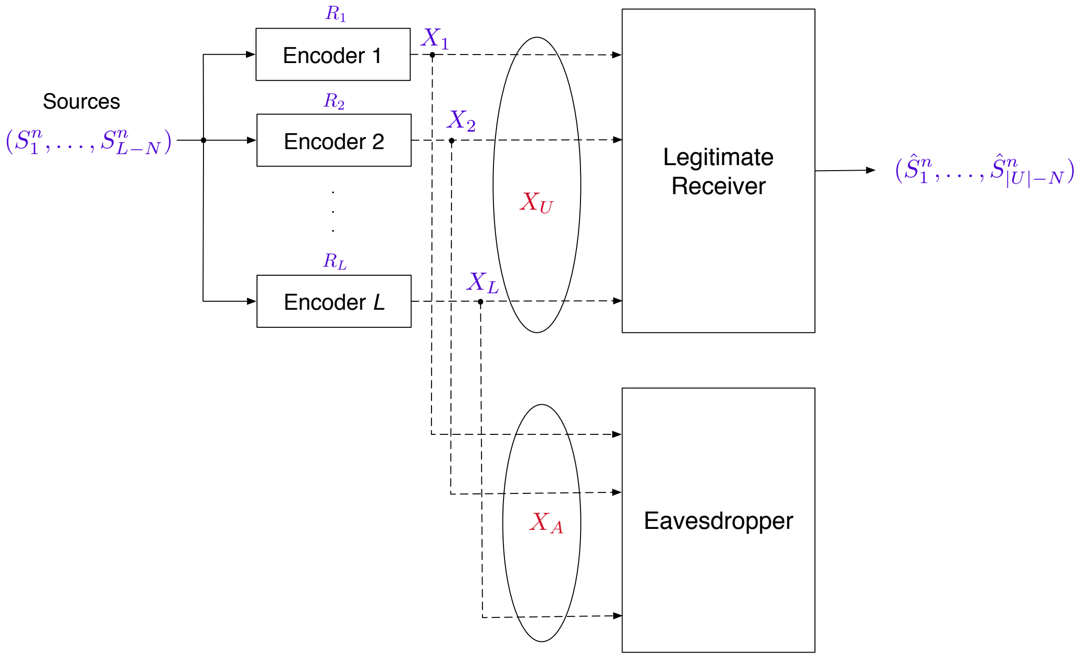


Figure 4.1: S-SMDC with L randomly accessible encoders 1 to L. A total of $L - N$ independent discrete memoryless sources $(S_1, \ldots, S_{L-N})$ are to be encoded at the encoders. The legitimate receiver, which has access to a subset U of the encoder outputs, needs to nearly perfectly reconstruct the sources $(S_1, \ldots, S_{|U|-N})$ whenever $|U| \geq N + 1$. The eavesdropper has access to a subset A of the encoder ouputs. All sources $(S_1, \ldots, S_{L-N})$ need to be kept perfectly secret from the eavesdropper whenever $|A| \leq N$.

- (Rate constraints)

$$\frac{1}{n} \log M_l \leq R_l + \epsilon, \quad \forall l = 1, \ldots, L; \tag{4.3}$$

- (Asymptotically perfect reconstruction at the legitimate receiver)

$$\Pr\{d_U(X_U) \neq (S_1^n, \ldots, S_{|U|-N}^n)\} \leq \epsilon, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq N + 1 \tag{4.4}$$

where $X_l := e_l((S_1^n, \ldots, S_{L-N}^n), K)$ is the output of the lth encoder, and K is the secret key shared by all L encoders; and

- (Perfect secrecy at the eavesdropper)

$$H(S_1^n, \ldots, S_{L-N}^n | X_A) = H(S_1^n, \ldots, S_{L-N}^n), \quad \forall A \subseteq \Omega_L \text{ s.t. } |A| \leq N \tag{4.5}$$

  i.e., observing the encoder outputs X_A does not provide *any* information regarding to the sources $(S_1^n, \ldots, S_{L-N}^n)$.

The *admissible rate region* $\mathcal{R}$ is the collection of *all* admissible rate tuples $(R_1, \ldots, R_L)$.

4.2 Superposition Coding Rate Region

A simple strategy for S-SMDC is to encode each of the $L - N$ sources separately without coding across different sources. Formally, the problem of encoding a single source S_α can be viewed as a special case of the general S-SMDC problem with $H(S_m) = 0$ for all $m \neq \alpha$. When $\alpha = 1$, the problem of encoding the single source S_1 is the well-known $(L, N+1)$ *threshold secret sharing* problem, for which the admissible rate region was characterized in the classical works [10, 11]. For the general case with $\alpha \geq 1$, the admissible rate region for encoding the single source S_α was characterized in [7] via a connection to the problem of threshold *ramp-type* secret sharing [12, 13] and utilizing some basic polyhedral structure of the admissible rate region. The result is summarized in the following proposition.

Proposition 3. *Let $\mathcal{R}^{(\alpha)}$ be the collection of all admissible rate tuples for encoding the single source S_α. Then, $\mathcal{R}^{(\alpha)}$ is given by the collection of all nonnegative tuples $(r_1^{(\alpha)}, \ldots, r_L^{(\alpha)})$ such that*

$$\sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Omega_L^{(\alpha)}. \tag{4.6}$$

By definition, the superposition coding rate region $\mathcal{R}_{sup}$ for encoding the sources $S_1, \ldots, S_{L-N}$ is given by the collection of nonnegative rate tuples $(R_1, \ldots, R_L)$ such that

$$R_l := \sum_{\alpha=1}^{L-N} r_l^{(\alpha)}, \quad \forall (r_l^{(1)}, \ldots, r_l^{(L-N)}) \in \prod_{\alpha=1}^{L-N} \mathcal{R}^{(\alpha)}. \tag{4.7}$$

Note that $\mathcal{R}^{(\alpha)}$ is *identical* to the admissible rate region for encoding the single source S_α in classical SMDC (even though the reconstruction and secrecy requirements are different between these two settings). We thus conclude that the superposition coding rate region $\mathcal{R}_{sup}$ for S-SMDC is given by the collection of nonnegative rate tuples $(R_1, \ldots, R_L)$ satisfying

$$\sum_{l=1}^{L} \lambda_l R_l \geq \sum_{\alpha=1}^{L-N} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha), \quad \forall \boldsymbol{\lambda} \in (\mathbb{R}^+)^L \tag{4.8}$$

where $f_\alpha(\boldsymbol{\lambda})$ is the optimal value of the linear program (2.19).

4.3 Optimality Of Superposition Coding

In [7], it was shown that superposition coding can achieve the minimum sum rate for the general S-SMDC problem. The proof was based on the trivial conditional version of the subset entropy inequality of Han. The main result of this section is to show that superposition coding can, in fact, achieve the entire admissible region for the general S-SMDC problem. Our main technical tool is the conditional extension of the subset entropy inequality of Yeung and Zhang proved in Theorem 6.

Theorem 9. *For the general S-SMDC problem, the admissible rate region*

$$\mathcal{R} = \mathcal{R}_{sup}. \tag{4.9}$$

Proof. Based on the discussions from Section 4.2, we naturally have $\mathcal{R}_{sup} \subseteq \mathcal{R}$. Thus, to show $\mathcal{R}_{sup} = \mathcal{R}$ we only need to show that $\mathcal{R} \subseteq \mathcal{R}_{sup}$, i.e., *any* admissible rate tuple $(R_1, \ldots, R_L)$ must satisfy (4.8).

Let $(R_1, \ldots, R_L)$ be an admissible rate tuple. By definition, for any sufficiently large block-length n there exists an $(n, (M_1, \ldots, M_L))$ code satisfying the rate constraints (4.3) for the admissible rate tuple $(R_1, \ldots, R_L)$, the asymptotically perfect reconstruction requirement (4.4), and the perfect secrecy requirement (4.5). Fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$, and choose $\mathcal{A}^{(\alpha)}$ and $s_{\boldsymbol{\lambda}}^{(\alpha)}$, $\alpha = 1, \ldots, L-N$, to satisfy all the requirement of Theorem 6.

First, let us show that

$$H(X_U|S_1^n, \ldots, S_{\alpha-1}^n, X_A) \geq nH(S_\alpha) - n\delta_n^{(\alpha)} + H(X_U|S_1^n, \ldots, S_\alpha^n, X_A) \tag{4.10}$$

for any $U \in \Omega_L^{(\alpha)}$, $A \in \mathcal{A}_U$, and $\alpha = 1, \ldots, L-N$, where $\delta_n^{(\alpha)} \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$.

Fix $U \in \Omega_L^{(\alpha)}$, $A \in \mathcal{A}_U$, and $\alpha = 1, \ldots, L-N$. By construction $|U| = \alpha$, $|A| = N$, and $A \cap U = \emptyset$, so we have $|U \cup A| = |U| + |A| = N + \alpha$. By the asymptotically perfect reconstruction requirement (4.4) and the well-known Fano's inequality, we have

$$H(S_1^n, \ldots, S_\alpha^n|X_U, X_A) \leq n\delta_n^{(\alpha)} \tag{4.11}$$

where $\delta_n^{(\alpha)} \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$. Furthermore, by the perfect secrecy requirement (4.5) we have

$$H(S_1^n, \ldots, S_\alpha^n|X_A) = H(S_1^n, \ldots, S_\alpha^n). \tag{4.12}$$

We thus have

$$H(X_U|S_1^n,\ldots,S_{\alpha-1}^n,X_A)+n\delta_n^{(\alpha)}$$

$$\geq H(X_U|S_1^n,\ldots,S_{\alpha-1}^n,X_A)+H(S_1^n,\ldots,S_\alpha^n|X_U,X_A) \tag{4.13}$$

$$\geq H(X_U|S_1^n,\ldots,S_{\alpha-1}^n,X_A)+H(S_\alpha^n|S_1^n,\ldots,S_{\alpha-1}^n,X_U,X_A) \tag{4.14}$$

$$= H(X_U,S_k^n|S_1^n,\ldots,S_{\alpha-1}^n,X_A) \tag{4.15}$$

$$= H(S_\alpha^n|S_1^n,\ldots,S_{k-1}^n,X_A)+H(X_V|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.16}$$

$$= H(S_1^n,\ldots,S_\alpha^n|X_A)-H(S_1^n,\ldots,S_{\alpha-1}^n|X_A)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.17}$$

$$= H(S_1^n,\ldots,S_\alpha^n)-H(S_1^n,\ldots,S_{\alpha-1}^n|X_A)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.18}$$

$$\geq H(S_1^n,\ldots,S_\alpha^n)-H(S_1^n,\ldots,S_{\alpha-1}^n)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.19}$$

$$= H(S_\alpha^n|S_1^n,\ldots,S_{\alpha-1}^n)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.20}$$

$$= H(S_\alpha^n)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.21}$$

$$= nH(S_\alpha)+H(X_U|S_1^n,\ldots,S_\alpha^n,X_A) \tag{4.22}$$

where (4.13) follows from (4.11), (4.18) follows from (4.12), (4.19) follows from the fact that conditioning reduces entropy, (4.21) follows from the fact that the sources $S_1,\ldots,S_\alpha$ are mutually independent, and (4.22) follows from the fact that the source S_α is memoryless. Moving $n\delta_n^{(\alpha)}$ to the right-hand side of the inequality completes the proof of (4.10).

Next, let us we show that

$$\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|X_A)$$
$$\geq n\sum_{\alpha=1}^{m}f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-n\sum_{\alpha=1}^{m}f_\alpha(\boldsymbol{\lambda})\delta_n^{(\alpha)}+\sum_{U\in\Omega_L^{(m)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,\ldots,S_m^n,X_A) \tag{4.23}$$

for any $m=1,\ldots,L-N$.

Consider a proof via an induction on m. First consider the base case with $m=1$. We have

$$\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}(U)}s_{\boldsymbol{\lambda}}(U,A)H(X_U|X_A)$$
$$\geq \sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}(U)}s_{\boldsymbol{\lambda}}(U,A)\left[nH(S_1)-n\delta_n^{(1)}+H(X_U|S_1^n,X_A)\right] \tag{4.24}$$
$$= nf_1(\boldsymbol{\lambda})H(S_1)-nf_1(\boldsymbol{\lambda})\delta_n^{(1)}+\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,X_A) \tag{4.25}$$

where (4.24) follows from (4.10) with $\alpha=1$.

Next, assume that the inequality (4.23) holds for $m=k-1$ for some $k=2,\ldots,L-N$, i.e.,

$$\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|X_A)$$
$$\geq n\sum_{\alpha=1}^{k-1}f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-n\sum_{\alpha=1}^{k-1}f_\alpha(\boldsymbol{\lambda})\delta_n^{(\alpha)}+\sum_{V\in\Omega_L^{(k-1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,\ldots,S_{k-1}^n,X_A). \tag{4.26}$$

We have

$$
\begin{aligned}
&\sum_{U\in\Omega_L^{(k-1)}}\sum_{A\in\mathcal{A}_U} s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,\ldots,S_{k-1}^n,X_A)\\
&\geq \sum_{U\in\Omega_L^{(k)}}\sum_{A\in\mathcal{A}_U} s_k(U,A)H(X_U|S_1^n,\ldots,S_{k-1}^n,X_A) && (4.27)\\
&\geq \sum_{U\in\Omega_L^{(k)}}\sum_{A\in\mathcal{A}_U} s_k(U,A)\left[nH(S_k)-n\delta_n^{(k)}+H(X_U|S_1^n,\ldots,S_k^n,X_A)\right] && (4.28)\\
&\geq nf_k(\boldsymbol{\lambda})H(S_k)-nf_k(\boldsymbol{\lambda})\delta_n^{(k)}+\sum_{U\in\Omega_L^{(k)}}\sum_{A\in\mathcal{A}_U} s_k(U,A)H(X_U|S_1^n,\ldots,S_k^n,X_A) && (4.29)
\end{aligned}
$$

where (4.27) follows from (2.105), and (4.28) follows from (4.10) with $\alpha = k$. Substituting (4.29) into (4.26) gives

$$
\begin{aligned}
&\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U} s_{\boldsymbol{\lambda}}(U,A)H(X_U|X_A)\\
&\geq n\sum_{\alpha=1}^{k} f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-n\sum_{\alpha=1}^{k} f_\alpha(\boldsymbol{\lambda})\delta_n^{(\alpha)}+\sum_{U\in\Omega_L^{(k)}}\sum_{A\in\mathcal{A}_U} s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,\ldots,S_k^n,X_A) && (4.30)
\end{aligned}
$$

i.e., the inequality (4.23) also holds for $m = k$. This completes the induction step and hence the proof of (4.23).

Finally, note that for $\alpha = 1$ the optimal solution for the linear program (2.19) is *unique* and is given by

$$
c_{\boldsymbol{\lambda}}(\{l\}) = \lambda_l, \quad \forall l = 1,\ldots,L. \tag{4.31}
$$

We thus have

$$n\left(\sum_{l=1}^{L}\lambda_l R_l\right)$$

$$=\sum_{l=1}^{L}c_{\boldsymbol{\lambda}}(\{l\})nR_l \tag{4.32}$$

$$\geq\sum_{l=1}^{L}c_{\boldsymbol{\lambda}}(\{l\})(H(X_l)-n\epsilon) \tag{4.33}$$

$$=\sum_{l=1}^{L}c_{\boldsymbol{\lambda}}(\{l\})H(X_l)-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.34}$$

$$=\sum_{U\in\Omega_L^{(1)}}c_{\boldsymbol{\lambda}}(U)H(X_U)-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.35}$$

$$=\sum_{U\in\Omega_L^{(1)}}\left[\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)\right]H(X_U)-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.36}$$

$$=\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U)-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.37}$$

$$\geq\sum_{U\in\Omega_L^{(1)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|X_A)-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.38}$$

$$\begin{aligned}\geq&\left[n\sum_{\alpha=1}^{L-N}f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-n\sum_{\alpha=1}^{L-N}f_\alpha(\boldsymbol{\lambda})\delta_n^{(\alpha)}+\right.\\&\left.\sum_{U\in\Omega_L^{(L-N)}}\sum_{A\in\mathcal{A}_U}s_{\boldsymbol{\lambda}}(U,A)H(X_U|S_1^n,\ldots,S_{L-N}^n,X_A)\right]-nf_1(\boldsymbol{\lambda})\epsilon\end{aligned} \tag{4.39}$$

$$\geq n\sum_{\alpha=1}^{L-N}f_\alpha(\boldsymbol{\lambda})H(S_\alpha)-n\sum_{\alpha=1}^{L-N}f_\alpha(\boldsymbol{\lambda})\delta_n^{(\alpha)}-nf_1(\boldsymbol{\lambda})\epsilon \tag{4.40}$$

where (4.33) follows from the rate constraint (4.3), (4.34) follows from the fact that $c_{\boldsymbol{\lambda}}^{(1)}$ is optimal so $f_1(\boldsymbol{\lambda})=\sum_{l=1}^{L}c_{\boldsymbol{\lambda}}(\{l\})$, (4.38) follows from the fact that conditioning reduce entropy, and (4.39) follows from (4.23) with $m=L-N$. Divide both sides of (4.40) by n and let $n\to\infty$ and $\epsilon\to 0$. Note that $\delta_n^{(\alpha)}\to 0$ in the limit as $n\to\infty$

and $\epsilon \to 0$ for all $\alpha = 1, \ldots, L - N$. We have thus proved that (4.8) holds for any admissible rate tuple $(R_1, \ldots, R_L)$. This completes the proof of the theorem. □

5. HIERARCHICAL MULTILEVEL DIVERSITY CODING

5.1 Problem Statement

As shown in Figure 5.1, the problem of HMDC consists of:

- a total of L independent discrete memoryless sources $\{S_\alpha[t]\}_{t=1}^{\infty}$, where $\alpha = 1, \ldots, L$ and t is the time index;

- The J encoders consist of two hierarchical encoder sets A and B, where $|A| + |B| = J$ and $|A| + \frac{|B|}{r} = L$ for some positive integer r. In other words, the sources are to be encoded by two sets of encoders A and B. The encoders in set A have higher rank than the ones in B, therefore we call encoders in A strong encoders and those in B weak encoders.

- a decoder has access to a subset U of the encoders outputs for some nonempty $U \subseteq A \cup B$.

For each $U \in \Omega_J = \{1, \ldots, J\}$, there is a corresponding J-dimensional 0-1 vector u such that $u_i = 1$ if and only if $i \in U$. Define the reliability scores for all encoders as a J-dimensional weight vector w, as the following

$$w = (\underbrace{1, \ldots, 1}_{|A|}, \underbrace{\frac{1}{r}, \ldots, \frac{1}{r}}_{|B|}). \tag{5.1}$$

The realization of U is unknown a priori at the encoders. However, no matter which U actually materializes, the decoder needs to nearly perfectly reconstruct the sources $S_1, \ldots, S_\alpha$ whenever $\lfloor u^T w \rfloor \geq \alpha$. The goal of encoding is to ensure that the number of sources that can be nearly perfectly reconstructed grows with the total weight

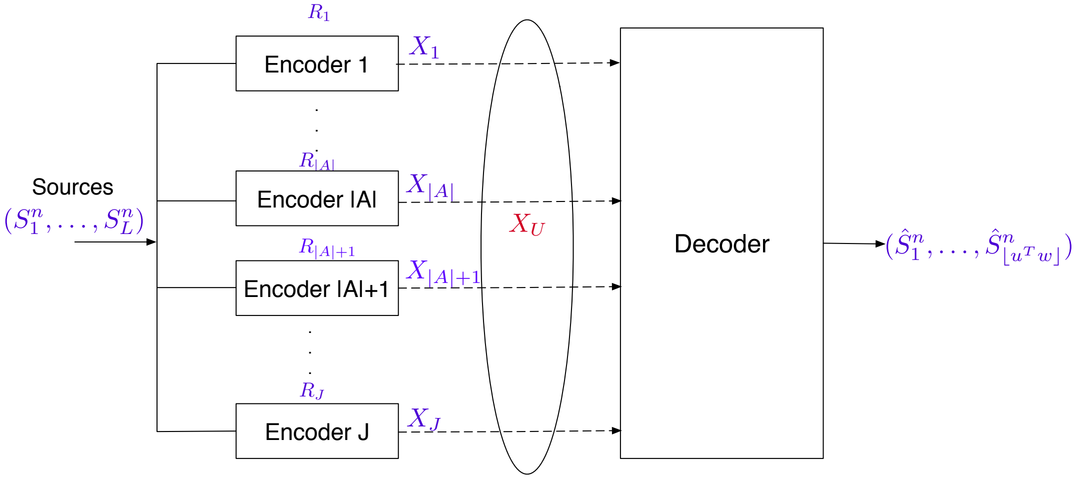


Figure 5.1: The HMDC problem where a total of L independent discrete memoryless sources $S_1, \ldots, S_L$ are to be encoded by a total of $J = |A| + |B|$ encoders of two types. Encoders from 1 to $|A|$ are strong encoders and the others are weak encoders. The decoder, which has access to a subset U of the encoder outputs, needs to nearly perfectly reconstruct the sources $S_1, \ldots, S_{\lfloor u^T w \rfloor}$ no matter what the realization of U is.

of accessible encoders. Note that when w is the all one vector of dimension L, the problem is reduced to the classical SMDC problem.

Formally, an $(n, (M_1, \ldots, M_J))$ code is defined by a collection of J encoding functions:

$$e_l : \prod_{\alpha=1}^{J} \mathcal{S}_\alpha^n \to \{1, \ldots, M_l\}, \quad \forall l = 1, \ldots, J \tag{5.2}$$

and $2^J - 1$ decoding functions:

$$d_U : \prod_{l \in U} \{1, \ldots, M_l\} \to \prod_{\alpha=1}^{\lfloor u^T w \rfloor} \mathcal{S}_\alpha^n, \quad \forall U \subseteq \Omega_J \text{ s.t. } U \neq \emptyset. \tag{5.3}$$

A nonnegative rate tuple $(R_1, \ldots, R_J)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block-length n, an $(n, (M_1, \ldots, M_J))$ code such that:

- (Rate constraints at the encoders)

$$\frac{1}{n}\log M_l \leq R_l + \epsilon, \qquad \forall l = 1, \ldots, J; \tag{5.4}$$

- (Asymptotically perfect reconstructions at the decoder)

$$\Pr\left\{d_U(X_U) \neq (S_1^n, \ldots, S_{\lfloor u^T w \rfloor}^n)\right\} \leq \epsilon, \qquad \forall U \subseteq \Omega_J \text{ s.t. } U \neq \emptyset \tag{5.5}$$

  where $S_\alpha^n := \{S_\alpha[t]\}_{t=1}^n$, $X_l := e_l(S_1^n, \ldots, S_L^n)$ is the output of encoder l, and $X_U := \{X_l : l \in U\}$.

The *admissible rate region* $\mathcal{R}$ is the collection of *all* admissible rate tuples $(R_1, \ldots, R_J)$. The *minimum sum rate* R_{ms} is defined as

$$R_{ms} := \min_{(R_1,\ldots,R_J)\in\mathcal{R}} \sum_{l=1}^{J} R_l. \tag{5.6}$$

5.2 Superposition Coding Rate Region

A simple strategy for HMDC is to encode each of the L sources separately without coding across different sources. Formally, the problem of encoding a single source S_α can be viewed as a special case of the HMDC problem with $H(S_m) = 0$ for all $m \neq \alpha$. In this case, the source S_α needs to be nearly perfectly reconstructed whenever the decoder can access a encoder set with weight at least α. This is essentially the same problem of multicast a single source through a 3-layer acyclic network to multiple destinations [14]. Therefore, any network multicast code can be used to encoder S_α. An example of such network coding problem is given in Figure 5.2.

The result on this network coding problem is summarized in the following proposition [14].

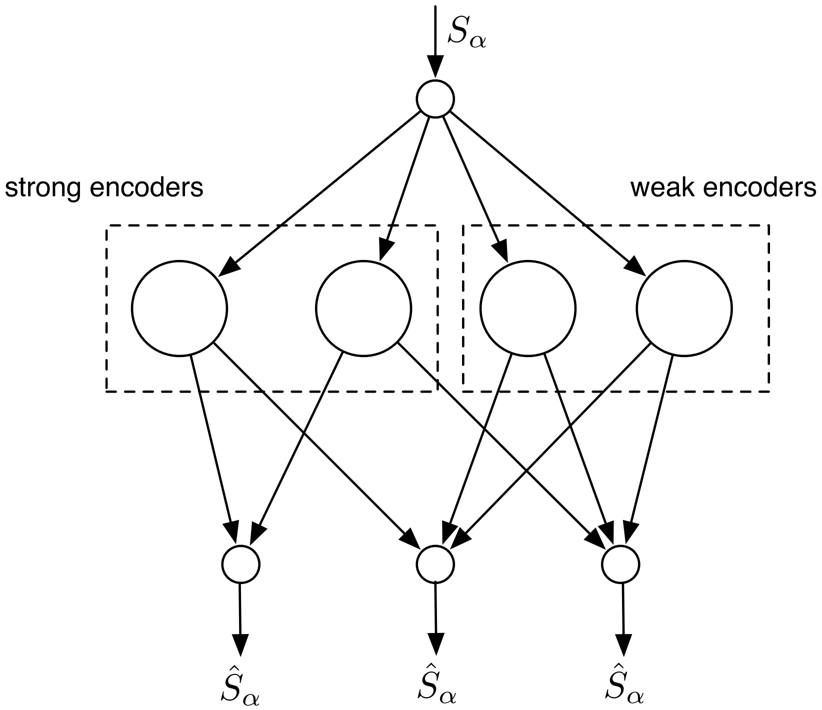


Figure 5.2: A single source network multicast example for coding a single source S_α ($\alpha = 2$) with 2 strong encoders and 2 weak encoders.

Proposition 4. *Let $\mathcal{R}^{(\alpha)}$ be the collection of all admissible rate tuples for encoding a single source S_α. Then, $\mathcal{R}^{(\alpha)}$ is given by the collection of all nonnegative tuples $(R_1^{(\alpha)}, \ldots, R_L^{(\alpha)})$ such that*

$$\sum_{l \in U} R_l^{(\alpha)} \geq H(S_\alpha), \forall U \in \Theta_J^{(\alpha)}. \tag{5.7}$$

where $\Theta_J^{(\alpha)} := \{U : U \in \Omega_J, u^T w = \alpha\}$.

Note that for $U' \in \Omega_J$ with $(u')^T w \in (\alpha, \alpha + 1)$, U' can be easily reduced to some U with $u^T w = \alpha$.

By definition, the superposition encoding rate region $\mathcal{R}_{sup}$ for HMDC is given by the collection of all nonnegative rate tuples $(R_1, \ldots, R_J)$ such that

$$R_l := \sum_{\alpha=1}^{L} r_l^{(\alpha)} \tag{5.8}$$

for some nonnegative $r_l^{(\alpha)}$, $\alpha = 1, \ldots, L$ and $l = 1, \ldots, J$, satisfying

$$\sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Theta_J^{(\alpha)}. \tag{5.9}$$

Note that we discard those decoding constraints on the set U' with $(u')^T w \in (\alpha, \alpha+1)$, which are dominated by (5.9).

In principle, an explicit characterization of the superposition coding rate region $\mathcal{R}_{sup}$ can be obtained by eliminating $r_l^{(\alpha)}$, $\alpha = 1, \ldots, L$ and $l = 1, \ldots, J$, via a Fourier-Motzkin elimination from (5.8) and (5.9). However, the elimination process is *unmanageable* even for moderate L and J, as there are simply too many equations involved. On the other hand, note that the superposition coding rate region $\mathcal{R}_{sup}$ is a convex polyhedron with polyhedral cone being $(\mathbb{R}^+)^J$, so an equivalent characterization is to characterize the supporting hyperplanes:

$$\sum_{l=1}^{J} \lambda_l R_l \geq f(\boldsymbol{\lambda}), \quad \forall \boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_J) \in (\mathbb{R}^+)^J \tag{5.10}$$

where

$$\begin{aligned} f(\boldsymbol{\lambda}) &= \min_{(R_1, \ldots, R_J) \in \mathcal{R}_{sup}} \sum_{l=1}^{J} \lambda_l R_l & (5.11) \\ &= \begin{array}{rl} \min & \sum_{l=1}^{J} \left(\sum_{\alpha=1}^{L} \lambda_l r_l^{(\alpha)} \right) \\ \text{subject to} & \sum_{l \in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Theta_J^{(\alpha)} \text{ and } \alpha = 1, \ldots, L \\ & r_l^{(\alpha)} \geq 0, \quad \forall \alpha = 1, \ldots, L \text{ and } l = 1, \ldots, J. \end{array} & (5.12) \end{aligned}$$

Clearly, the above optimization problem can be separated into the following L sub-optimization problems:

$$f(\boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f'_\alpha(\boldsymbol{\lambda}) \tag{5.13}$$

where

$$\begin{array}{rcll} & & \min & \sum_{l=1}^{J} \lambda_l r_l^{(\alpha)} \\ f'_\alpha(\boldsymbol{\lambda}) & = & \text{subject to} & \sum_{l\in U} r_l^{(\alpha)} \geq H(S_\alpha), \quad \forall U \in \Theta_J^{(\alpha)} \\ & & & r_l^{(\alpha)} \geq 0, \quad \forall l = 1, \ldots, J \end{array} \tag{5.14}$$

$$\begin{array}{rcll} & & \max & \left(\sum_{U\in\Theta_J^{(\alpha)}} c_{\boldsymbol{\lambda}}(U)\right) H(S_\alpha) \\ & = & \text{subject to} & \sum_{\{U\in\Theta_J^{(\alpha)}:U\ni l\}} c_{\boldsymbol{\lambda}}(U) \leq \lambda_l, \quad \forall l = 1, \ldots, J \\ & & & c_{\boldsymbol{\lambda}}(U) \geq 0, \quad \forall U \in \Theta_J^{(\alpha)}. \end{array} \tag{5.15}$$

and (5.15) follows from the strong *duality* for linear programs. For any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^J$ and any $\alpha = 1, \ldots, L$, let

$$\begin{array}{rcll} & & \max & \sum_{U\in\Theta_J^{(\alpha)}} c_{\boldsymbol{\lambda}}(U) \\ f_\alpha(\boldsymbol{\lambda}) & := & \text{subject to} & \sum_{\{U\in\Theta_J^{(\alpha)}:U\ni l\}} c_{\boldsymbol{\lambda}}(U) \leq \lambda_l, \quad \forall l = 1, \ldots, J \\ & & & c_{\boldsymbol{\lambda}}(U) \geq 0, \quad \forall U \in \Theta_J^{(\alpha)}. \end{array} \tag{5.16}$$

Then, we have $f'_\alpha(\boldsymbol{\lambda}) = f_\alpha(\boldsymbol{\lambda}) H(S_\alpha)$ and hence

$$f(\boldsymbol{\lambda}) = \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha) \tag{5.17}$$

for any $\boldsymbol{\lambda} \in (\mathbb{R}^+)^J$. Substituting (5.17) into (5.10), we conclude that the superposition coding rate region $\mathcal{R}_{sup}$ is given by the collection of nonnegative rate tuples

$(R_1, \ldots, R_J)$ satisfying

$$\sum_{l=1}^{J} \lambda_l R_l \geq \sum_{\alpha=1}^{L} f_\alpha(\boldsymbol{\lambda}) H(S_\alpha), \quad \forall \boldsymbol{\lambda} \in (\mathbb{R}^+)^L. \tag{5.18}$$

5.3 Optimality Of Superposition Coding For Minimum Sum Rate

In this section, we show that superposition coding is optimal in achieving the minimum sum rate. For a general $\boldsymbol{\lambda}$, the linear program (5.16) does not admit a *closed-form* solution. When considering the sum rate, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J) = \mathbf{1}$, the following lemma gives a sufficient condition for the optimal solution.

Lemma 2. *For $\boldsymbol{\lambda} = \mathbf{1}$, $c_{\mathbf{1}}^{(\alpha)}$ is an optimal solution to* (5.16) *if*

$$\sum_{U \in \Theta_J^{(\alpha)} : U \ni l} c_{\mathbf{1}}(U) = 1, \quad \forall l = 1, \ldots, J. \tag{5.19}$$

Meanwhile,

$$f_\alpha(\mathbf{1}) = \frac{L}{\alpha}. \tag{5.20}$$

Proof. Write the constraints of the linear program to $\sum_{U \in \Theta_J^\alpha} c_{\mathbf{1}}(U) u \leq \mathbf{1}$. Then for the J-dimensional weight vector w, we have

$$\sum_{U \in \Theta_J^\alpha} c_{\mathbf{1}}(U)(u \cdot w) \leq \mathbf{1} \cdot w. \tag{5.21}$$

Since $u \cdot w = \alpha$ for all $U \in \Theta_J^{(\alpha)}$, we have

$$\sum_{U \in \Theta_J^{\alpha}} c_{\mathbf{1}}(U) \leq \frac{L}{\alpha}. \tag{5.22}$$

Therefore, $f_\alpha(\mathbf{1}) \leq \frac{L}{\alpha}$. And the equality holds if and only if (5.19) holds. □

As discussed before, in order to show that superposition coding achieves the minimum sum rate, we need to show for any coding scheme, its sum rate should satisfy

$$\sum_{l=1}^{J} R_l \geq \sum_{\alpha=1}^{L} \frac{L}{\alpha} H(S_\alpha). \tag{5.23}$$

As in the classical SMDC, we need some sort of subset entropy inequalities to prove (5.23). For $\lambda = \mathbf{1}$, the following proposition gives a way to construct an optimal solution to (5.16) with α replaced by $\alpha - 1$ via the given optimal solution to (5.16).

Proposition 5. *Suppose that* $c_{\mathbf{1}}^{(\alpha)}$*,which satisfies* (5.19)*, is an optimal solution to linear program* (5.16) *and for each* $U \in \Theta_J^{(\alpha)}$*, there are* $s(U)$ *strong encoders and* $t(U)$ *weak encoders, i.e.,* $s(U) + \frac{t(U)}{r} = \alpha$*. Then for* $V \in \Theta_L^{\alpha-1}$*,*

$$c_{\mathbf{1}}(V) = \sum_{\substack{U \in \Theta_J^{(\alpha)}, U \supset V \\ s(U)=s(V)+1}} p(U) c_{\mathbf{1}}(U) + \sum_{\substack{U \in \Theta_J^{(\alpha)}, U \supset V \\ t(U)=t(V)+r}} q(U) c_{\mathbf{1}}(U) \tag{5.24}$$

$$p(U) = \frac{t(U) - r}{s(U)t(U) + r - s(U)r} \tag{5.25}$$

$$q(U) = \frac{1}{\binom{t(U)-1}{r-1}} \frac{r}{(s(U)t(U) + r - s(U)r)} \tag{5.26}$$

is an optimal solution to linear program (5.16) *with* α *replaced by* $\alpha - 1$*.*

Proof. We prove by induction on α. Initially, for $U \in \Theta_J^{(L)}$, $U = \Omega_J$ is the unique subset. Thus we have a unique optimal solution $c_{\mathbf{1}}(U) = 1$ by Lemma 2.

Suppose $c_{\mathbf{1}}^{(\alpha)}$ is an optimal solution to linear program (5.16) and the following equality holds.

$$\sum_{U \in \Theta_J^{(\alpha)}, U \ni l} c_{\mathbf{1}}(U) = 1, \forall l = 1, \ldots, J. \tag{5.27}$$

A key observation for providing the "chain" form of the subset entropy inequality is the following. For any subset $U \in \Theta_J^{(\alpha)}$ as the parent set, there are two ways to construct the child subset. One is to reduce any one strong encoders from U, the other is to reduce any r weak encoders from U. The combinations to do such reductions are captured in (5.25) and (5.26), which can be easily checked by simple counting argument.

Hence,

$$\begin{aligned}
&\sum_{V \in \Theta_J^{(\alpha-1)}, V \ni l} c_{\mathbf{1}}(V) \\
&= \sum_{V \in \Theta_J^{(\alpha-1)}, V \ni l} \sum_{\substack{U \in \Theta_J^{(\alpha)}, U \supset V \\ s(U) = s(V)+1}} p(U) c_{\mathbf{1}}(U) + \sum_{V \in \Theta_J^{(\alpha-1)}, V \ni l} \sum_{\substack{U \in \Theta_J^{(\alpha)}, U \supset V \\ t(U) = t(V)+r}} q(U) c_{\mathbf{1}}(U) && (5.28) \\
&= \sum_{U \in \Theta_J^{(\alpha)}, U \ni l} \sum_{\substack{V \in \Theta_J^{(\alpha-1)}, U \supset V \\ s(U) = s(V)+1}} p(U) c_{\mathbf{1}}(U) + \sum_{U \in \Theta_J^{(\alpha)}, U \ni l} \sum_{\substack{V \in \Theta_J^{(\alpha-1)}, U \supset V \\ t(U) = t(V)+r}} q(U) c_{\mathbf{1}}(U) && (5.29) \\
&= \sum_{U \in \Theta_J^{(\alpha)}, U \ni l} c_{\mathbf{1}}(U) && (5.30) \\
&= 1. && (5.31)
\end{aligned}$$

This implies that when $\boldsymbol{\lambda} = \mathbf{1}$, there is always an α-optimal solution. □

Now we can establish the inequality chain in the following theorem to prove the optimality of separate encoding in terms of minimum sum rate.

Theorem 10. *If $\boldsymbol{\lambda} = \mathbf{1}$, for any optimal solution $c_{\mathbf{1}}^{(\alpha)}$ to linear program (5.16), there exists an optimal solution $c_{\mathbf{1}}^{(\alpha-1)}$ to linear program (5.16) with α replaced by $\alpha - 1$ such that*

$$\sum_{U \in \Theta_J^{(\alpha)}} c_{\mathbf{1}}(U) H(X_U) \leq \sum_{V \in \Theta_J^{(\alpha-1)}} c_{\mathbf{1}}(V) H(X_V) \tag{5.32}$$

Proof. This is simply true by invoking the Madiman-Tetali subset inequality. To show that, we need to verify

$$g_U(V) := \begin{cases} p(U), & \text{if } s(U) = s(V) + 1; \\ q(U), & \text{if } t(U) = t(V) + r. \end{cases} \tag{5.33}$$

is indeed a fractional cover.

If l belongs to strong encoders,

$$\sum_{V \in \mathcal{V}_U : V \ni l} g_U(V) = (s(U) - 1) p(U) + \binom{t(U)}{r} q(U) = 1; \tag{5.34}$$

otherwise

$$\sum_{V \in \mathcal{V}_U : V \ni l} g_U(V) = s(U) p(U) + \binom{t(U) - 1}{r - 1} q(U) = 1. \tag{5.35}$$

Therefore, we arrives at

$$\sum_{V \in \mathcal{V}_U : V \ni l} g_U(V) = 1, \forall l \in U. \tag{5.36}$$

Thus, by the same argument as shown in Corollary 1, we complete the proof. □

Now we are ready to show the optimality of superposition for achieving the minimum sum rate.

Theorem 11. *For the minimum sum rate of HMDC problem,*

$$\mathcal{R}_{ms} = \sum_{\alpha=1}^{L} \frac{L}{\alpha} H(S_\alpha). \tag{5.37}$$

Proof. Iteratively applying the decoding requirements and (5.32), we may obtain

$$\sum_{V \in \Theta_L^{(1)}} c_{\mathbf{1}}(V) H(V) \geq \sum_{U \in \Theta_L^{(m)}} c_{\mathbf{1}}(U) H(X_U | S_1^n, \ldots, S_m^n) \\ + n \sum_{\alpha=1}^{m} f_\alpha(\mathbf{1}) H(S_\alpha) - n \sum_{\alpha=1}^{m} f_\alpha(\mathbf{1}) \delta_\alpha^{(n)} \tag{5.38}$$

for any $m = 1, \ldots, L$. In particular, let $m = L$, and there exists an optimal solution to the linear program (5.16) with $\alpha = 1$ such that

$$\sum_{U \in \Theta_L^{(1)}, U \ni l} c_{\mathbf{1}}(U) = 1. \tag{5.39}$$

We have

$$\begin{aligned}
\sum_{l=1}^{J} H(X_l) &= \left(\sum_{U\in\Theta_L^{(1)}, U\ni l} c_{\mathbf{1}}(U) \right) \sum_{l=1}^{J} H(X_l) \\
&\geq \sum_{U\in\Theta_L^{(1)}} c_{\mathbf{1}}(U) H(X_U) \qquad (5.40) \\
&\geq \sum_{U\in\Theta_L^{(L)}} c_{\mathbf{1}}(U) H(X_U|S_1^n, \ldots, S_L^n) + n \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) H(S_\alpha) - n \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) \delta_\alpha^{(n)} \\
&\geq n \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) H(S_\alpha) - n \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) \delta_\alpha^{(n)}, \qquad (5.41)
\end{aligned}$$

where (5.40) is by the entropy independence bound.

Substituting and dividing both sides of the inequality by n, we have

$$\sum_{l=1}^{J} (R_l + \epsilon) \geq \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) H(S_\alpha) - \sum_{\alpha=1}^{L} f_\alpha(\mathbf{1}) \delta_\alpha^{(n)}. \qquad (5.42)$$

Finally, letting $n \to \infty$ and $\epsilon \to 0$ completes the proof of (5.37), i.e., superposition coding can achieve the minimum sum rate for the HMDC problem. □

5.4 Optimality Of Superposition Coding Beyond Minimum Sum Rate

Based on the result of Theorem 11, it is very tempting to conjecture that superposition coding can in fact achieve the entire admissible rate region for the HMDC problem. However, even if we only consider the points on the boundary of the rate region where the supporting hyperplanes are characterized by two coefficients in the sense that

$$\boldsymbol{\lambda} = (\underbrace{\lambda_1, \ldots, \lambda_1}_{|A|}, \underbrace{\lambda_2, \ldots, \lambda_2}_{|B|}),$$

the problem seems nontrivial. In Appendix C, we verify that superposition coding is indeed optimal for this case where $r = 2$ and there are 4 discrete memoryless sources to be encoded by 2 strong encoders and 4 weak encoders. Our proof relies on an explicit characterization of the optimal solutions to the linear programs (5.16). The optimality of superposition coding is then proved by carefully establishing the subset inequality in the same form as (5.32).

Extending such a proof strategy to the general HMDC problem, however, faces a number of challenges. To begin with, explicitly finding the optimal solution to the linear programs for the general $\boldsymbol{\lambda}$ seems extremely complicated, because the relationship between the optimal solutions to linear programs with different α is unclear due to the vast number of combinations of strong and weak encoders that satisfy the decoding constraints. To understand the structure of the linear program (5.16) is currently under the investigation. Moreover, even knowing the optimal solution, it is not straightforward to establish the subset inequality in the same form of (5.32) for all possible supporting hyperplanes of the entire rate region, since we need to verify whether the optimal solutions to the linear programs of different α can be connected with the notion of fractional covering. Finally, it is possible that the potentially necessary inequalities to establish the optimality is beyond the framework of hypergraph covering as the sliding-window subset inequality (2.47).

6. CONCLUSION

SMDC is a classical model for coding over distributed storage. In this setting, a simple separate encoding strategy known as superposition coding was shown to be optimal in terms of achieving the minimum sum rate [3] and the entire admissible rate region [5] of the problem. The proofs utilized carefully constructed induction arguments, for which the classical subset entropy inequality of Han [4] played a key role.

This thesis includes two parts. In the first part the existing optimality proofs for classical SMDC were revisited, with a focus on their connections to subset entropy inequalities. First, a new sliding-window subset entropy inequality was introduced and then used to establish the optimality of superposition coding for achieving the minimum sum rate under a weaker source-reconstruction requirement. Second, a subset entropy inequality recently proved by Madiman and Tetali [6] was used to develop a new structural understanding to the proof of Yeung and Zhang [5] on the optimality of superposition coding for achieving the entire admissible rate region. Building on the connections between classical SMDC and the subset entropy inequalities developed in the first part, in the second part the optimality of superposition coding was further extended to the cases where there is an additional all-access encoder (SMDC-A), an additional secrecy constraint (S-SMDC) or an encoder hierarchy (HMDC). However, we are only able to show that superposition coding is optimal in achieving the minimum sum rate in HMDC case. The optimality for the entire admissible rate region is under further research.

Finally, we mention here that an “asymmetric” setting of the multilevel diversity coding problem was considered in the recent work [15], where the sources that need to be asymptotically perfectly reconstructed depend on, not only the cardinality, but the actual subset of the encoder outputs available at the decoder. Unlike the symmetrical setting considered in [1–3, 5] and in this thesis, as demonstrated in [15] for the case with three encoders, coding across different sources is generally needed to achieve the entire admissible rate region of the problem.

REFERENCES

[1] J. R. Roche, "Distributed information storage," *Ph.D. Dissertation*, Stanford University, Stanford, CA, Mar. 1992.

[2] R. W. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inf. Theory*, vol. 41, pp. 412–422, Mar. 1995.

[3] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1059–1064, May 1997.

[4] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Inf. Control*, vol. 36, no. 2, pp. 133–156, Feb. 1978.

[5] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 45, pp. 609–621, Mar. 1999.

[6] M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2699–2713, June 2010.

[7] A. Balasubramanian, H. D. Ly, S. Li, T. Liu, and S. L. Miller, "Secure symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, submitted for publication.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd ed.* Hoboken, NJ: John Wiley & Sons, 2006.

[9] R. C. Singleton, "Maximum distance q-nary codes," *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 116–118, Apr. 1964.

[10] A. Shamir, "How to share a secret," *Comm. ACM*, vol. 22, pp. 612–613, Nov. 1979.

[11] G. R. Blakley, "Safeguarding cryptographic keys," in *Proc. National Computer Conference*, New York, NY, June 1979, vol. 48, pp. 313–317.

[12] H. Yamamoto, "Secret sharing system using (k, L, n) threshold scheme," *IEICE Trans. Fundamentals (Japanese Edition)*, vol. J68-A, pp. 945–952, Sept. 1985 (English Translation: Scripta Technica, Inc., Electronics and Comm. in Japan, Part I, vol. 69, pp. 46–54, 1986).

[13] G. R. Blakley and C. Meadows, "Security of ramp scheme," in *Advances in Cryptology - CRYPTO '84, LNCS 196*, pp. 242–269, 1985.

[14] R. W. Yeung, *Information Theory and Network Coding.* New York, NY: Springer, 2008.

[15] S. Mohajer, C. Tian, and S. N. Diggavi, "Asymmetric multilevel diversity coding and asymmetric Gaussian multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4367–4387, Sept. 2010.

APPENDIX A

PROOFS OF SUBSET ENTROPY INEQUALITIES

In this part, the differences in the proof techniques for subset entropy inequalities are highlighted. Before proceeding to the proof, we shall revisit the entropy function. The entropy $H(X)$ of a discrete random variable X with possible values $x_1, \ldots, x_n$ and probability mass function $p(X)$ is defined by

$$H(X) = E[\log p(X)] = -\sum_{i=1}^{n} p(x_i) \log p(x_i). \tag{A.1}$$

Similarly, we can define joint entropy, conditional entropy, mutual information and so on. The key properties of entropy function includes the following. For random variables X,Y and Z,

- Chain rule: $H(XY) = H(X) + H(Y|X)$,
- Nonnegativity of mutual information: $I(X;Y|Z) \geq 0$,
- Conditioning reduces entropy: $H(X|Y) \geq H(X|Z)$ if Y is some function of Z.

Meanwhile, we know that entropy function belongs to a more general class of set function: *submodular* set functions. A function $f : 2^{\Omega} \to \mathbf{R}$ is defined by satisfying one of the following inequalities.

- Submodularity: $f(S) + f(T) \geq f(S \cap T) + f(S \cup T)$,
- Diminishing return: $f(S \cup X) - f(S) \geq f(T \cup X) - f(T)$,$\forall S \subseteq T$,

where S,T and X are subsets of Ω.

A.1 Proof Of Han's Subset Inequality

For L jointly distributed random variables $(X_1, \ldots, X_L)$, consider the average joint entropy over all subsets of a fixed size. First prove the right end of the Han's subset inequality chain, i.e.,

$$H(X_1, \ldots, X_L) \leq \sum_{U \in \Omega_L^{(L-1)}} \frac{X_U}{L-1} \tag{A.2}$$

Using the chain rule to break the joint entropy into two terms, we have

$$\begin{aligned}
H(X_1, \ldots, X_L) &= H(X_1, \ldots, X_{L-1}) + H(X_L | X_1, \ldots, X_{L-1}) && \text{(A.3)}\\
H(X_1, \ldots, X_L) &= H(X_1, \ldots, X_{L-2}, X_L) + H(X_{L-1} | X_1, \ldots, X_{L-2}, X_L) && \text{(A.4)}\\
&\leq H(X_1, \ldots, X_{L-2}, X_L) + H(X_{L-1} | X_1, \ldots, X_{L-2}) && \text{(A.5)}\\
&\vdots && \text{(A.6)}\\
H(X_1, \ldots, X_L) &\leq H(X_2, \ldots, X_L) + H(X_1). && \text{(A.7)}
\end{aligned}$$

Adding these L inequalities and using the chain rule to combine conditional entropy terms,

$$LH(X_1, \ldots, X_L) \leq \sum_{l=1}^{L} H(X_1, \ldots, X_{l-1}, X_{l+1}, \ldots, X_L) + H(X_1, \ldots, X_L) \tag{A.8}$$

Rearranging the terms in (A.8) leads to (A.2).

For each $|U|$-element subset U, taking a uniform average over its $(|U|-1)$-element subsets, one has

$$H(X_U) \leq \sum_{V \subset U, |V| = |U|-1} \frac{H(X_V)}{|U|-1}, \ \forall U \in \Omega_L. \tag{A.9}$$

Taking a uniform average over all subsets of the same size as U, we arrive at Han's subset inequality.

A.2 Proof Of Madiman–Tetali's Inequality

Equip the elements in any set U with indices of natural increasing order. For $j \in U$, denote by $< j$ as the subset of elements with indices smaller than j.

$$\begin{aligned}
H(X_U) &= \sum_{j \in U} H(X_j | X_{<j}) && \text{(A.10)} \\
&\leq \sum_{j \in U} H(X_j | X_{<j}) \sum_{V \in \mathcal{V}_U} g(V) I_{\{j \in V\}} && \text{(A.11)} \\
&= \sum_{j \in U} \sum_{V \in \mathcal{V}_U} g(V) I_{\{j \in V\}} H(X_j | X_{<j}) && \text{(A.12)} \\
&= \sum_{V \in \mathcal{V}_U} g(V) \sum_{j \in V} H(X_j | X_{<j}) && \text{(A.13)} \\
&\leq \sum_{V \in \mathcal{V}_U} g(V) \sum_{j \in V} H(X_j | X_{<j \cap V}) && \text{(A.14)} \\
&= \sum_{V \in \mathcal{V}_U} g(V) H(X_V) && \text{(A.15)}
\end{aligned}$$

where (A.10) is due to chain rule of entropy functions, (A.11) is due to the factional cover, (A.13) is by interchanging the sum order and (A.14) is by the fact that conditioning reduces entropy.

It is not hard to see that the proof of Madiman-Tetali's inequality follows the same steps as the one of Han's inequality. The only difference the averaging coefficients. The average coefficients for Han's inequality are uniform, while the coefficients for Madiman-Tetali's inequality are non-uniform in general and form fractional covers .

Furthermore, the proofs for both Han's subset inequality and Madiman-Tetali's inequality only depend on the properties of condtional entropy, i.e., the chain rule and the fact that conditioning reduces entropy. However, in order to show the sliding-window subset entropy inequality, we rely on the submodularity of entropy functions. In fact, submodularity can induce the properties of conditional entropy.

APPENDIX B

PROOF OF THEOREM 5

Consider a proof via an induction on the total number of encoders L. Fix $\boldsymbol{\lambda} \in (\mathbb{R}^+)^L$. Without loss of generality, let us assume that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0. \tag{B.1}$$

First consider the base case with $L = 2$. In this case, the optimal solution to the linear program (2.19) is unique and is given by

$$c_{\boldsymbol{\lambda}}(\{l\}) = \lambda_l, \quad l = 1, 2 \quad \text{and} \quad c_{\boldsymbol{\lambda}}(\{1, 2\}) = \lambda_2. \tag{B.2}$$

When $f_2(\boldsymbol{\lambda}) = \lambda_2 > 0$, it is straightforward to verify that

$$g_{\{1,2\}}(\{l\}) = \lambda_l / \lambda_2, \quad l = 1, 2 \tag{B.3}$$

is a fractional cover of $(\{1, 2\}, \{\{1\}, \{2\}\})$ and such that

$$c_{\boldsymbol{\lambda}}(\{l\}) = g_{\{1,2\}}(\{l\}) c_{\boldsymbol{\lambda}}(\{1, 2\}), \quad l = 1, 2. \tag{B.4}$$

Now, assume that the theorem holds for $L = N - 1$ for some integer $N \geq 3$. Fix $\alpha \in \{2, \ldots, N\}$, and let $c_{\boldsymbol{\lambda}}^{(\alpha)}$ be an optimal solution to the linear program to (2.19) with the optimal value $f_\alpha(\boldsymbol{\lambda}) > 0$. Next, we show that we can always find a collection of functions $\{g_U : U \in \Omega_L^{(\alpha)}\}$ for which each g_U is a fractional cover of $(U, \mathcal{V}_U)$ and such that $c_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{c_{\boldsymbol{\lambda}}(V) : V \in \Omega_L^{(\alpha-1)}\}$ where $c_{\boldsymbol{\lambda}}(V)$ is given by (2.91) is an optimal solution to the linear program (2.19) with α replaced by $\alpha - 1$.

We shall consider the following three cases separately.

Case 1: $\lambda_1 \leq \frac{\lambda_2+\cdots+\lambda_N}{\alpha-1}$. In this case, it is sufficient to consider for any $U \in \Omega_N^{(\alpha)}$, the *uniform* fractional cover

$$g_U(V) = \frac{1}{\alpha - 1}, \quad \forall V \in \mathcal{V}_U \tag{B.5}$$

for the hypergraph $(U, \mathcal{V}_U)$ so we have

$$c_{\boldsymbol{\lambda}}(V) = \sum_{U \in \mathcal{U}_V} \frac{c_{\boldsymbol{\lambda}}(U)}{\alpha - 1}, \quad \forall V \in \Omega_N^{(\alpha-1)}. \tag{B.6}$$

By [5, Eq. (39)], $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ constructed as such is an optimal solution to the linear program (2.19) with α replaced by $\alpha - 1$.

Case 2: $\lambda_1 > \frac{\lambda_2+\cdots+\lambda_N}{\alpha-2}$. In this case, by [5, Lemma 6] $c_\alpha(U) > 0$ implies that $U \ni 1$. Furthermore, by [5, Lemma 8] $\tilde{c}_{\boldsymbol{\lambda}}^{(\alpha-1)} = \{\tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) : \tilde{U} \subseteq \tilde{\Omega}_{N-1} := \{2, \ldots, N\}\}$ where

$$\tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) = c_{\boldsymbol{\lambda}}(\{1\} \cup \tilde{U}) \tag{B.7}$$

is an optimal solution to the linear program

$$\begin{array}{rl} \max & \sum_{\tilde{U} \in \tilde{\Omega}_{N-1}^{(\alpha-1)}} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) \\ \text{subject to} & \sum_{\tilde{U} \in \tilde{\Omega}_{N-1}^{(\alpha-1)}, \tilde{U} \ni l} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) \leq \lambda_l, \quad \forall l = 2, \ldots, N \\ & \tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) \geq 0, \quad \forall \tilde{U} \in \tilde{\Omega}_{N-1}^{(\alpha-1)} \end{array} \tag{B.8}$$

with the optimal solution $\tilde{f}_{\alpha-1}(\boldsymbol{\lambda}) = f_\alpha(\boldsymbol{\lambda}) > 0$. Thus, by the induction assumption there exists a collection of functions $\{\tilde{g}_{\tilde{U}} : \tilde{U} \in \tilde{\Omega}_{N-1}^{(\alpha-1)}\}$ such that each $\tilde{g}_{\tilde{U}}$ is a fractional cover of $(\tilde{U}, \tilde{\mathcal{V}}_{\tilde{U}})$ and $\tilde{c}_{\boldsymbol{\lambda}}^{(\alpha-2)} = \{\tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) : \tilde{V} \in \tilde{\Omega}_{N-1}^{(\alpha-2)}\}$ where

$$\tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) := \sum_{\tilde{U} \in \tilde{\mathcal{U}}_{\tilde{V}}} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) \tilde{g}_{\tilde{U}}(\tilde{V}) \tag{B.9}$$

is an optimal solution to the linear program

$$\begin{array}{rl} \max & \sum_{\tilde{V} \in \tilde{\Omega}_{N-1}^{(\alpha-2)}} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) \\ \text{subject to} & \sum_{\tilde{V} \in \tilde{\Omega}_{N-1}^{(\alpha-2)}, \tilde{V} \ni l} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) \leq \lambda_l, \quad \forall l = 2, \ldots, N \\ & \tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) \geq 0, \quad \forall \tilde{V} \in \tilde{\Omega}_{N-1}^{(\alpha-2)}. \end{array} \tag{B.10}$$

For any $U \in \Omega_N^{(\alpha)}$ such that $U \ni 1$, let $\tilde{U} = U \setminus \{1\}$, and let

$$g_U(V) := \begin{cases} \tilde{g}_{\tilde{U}}(\tilde{V}), & \text{if } V = \{1\} \cup \tilde{V} \text{ for some } \tilde{V} \in \tilde{\mathcal{V}}_{\tilde{U}} \\ 0, & \text{otherwise.} \end{cases} \tag{B.11}$$

For any $U \in \Omega_N^{(\alpha)}$ such that $1 \notin U$, let us choose g_U to be an *arbitrary* fractional cover of $(U, \mathcal{V}_U)$. Then, for any $V \in \Omega_N^{(\alpha-1)}$ such that $V \ni 1$ we have

$$\begin{aligned} c_{\boldsymbol{\lambda}}(V) &= \sum_{U \in \mathcal{U}_V} c_{\boldsymbol{\lambda}}(U) g_U(V) && \text{(B.12)} \\ &= \sum_{\tilde{U} \in \tilde{\mathcal{U}}_{\tilde{V}}} c_{\boldsymbol{\lambda}}(\{1\} \cup \tilde{U}) \tilde{g}_{\tilde{U}}(\tilde{V}) && \text{(B.13)} \\ &= \sum_{\tilde{U} \in \tilde{\mathcal{U}}_{\tilde{V}}} \tilde{c}_{\boldsymbol{\lambda}}(\tilde{U}) \tilde{g}_{\tilde{U}}(\tilde{V}) && \text{(B.14)} \\ &= \tilde{c}_{\boldsymbol{\lambda}}(\tilde{V}) && \text{(B.15)} \end{aligned}$$

where $\tilde{V} = V \setminus \{1\}$, and for any $V \in \Omega_N^{(\alpha-1)}$ such that $1 \notin V$

$$c_{\boldsymbol{\lambda}}(V) = \sum_{U \in \mathcal{U}_V} c_{\boldsymbol{\lambda}}(U) g_U(V) = 0. \tag{B.16}$$

By [5, Eq. (46)], $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ constructed as such is an optimal solution to the linear program (2.19) with α replaced by $\alpha-1$. It remains to show that g_U is a fractional cover of $(U, \mathcal{V}_U)$ for any $U \in \Omega_N^{(\alpha)}$ such that $U \ni 1$.

Fix $U \in \Omega_N^{(\alpha)}$ such that $U \ni 1$. For any $i \in U \setminus \{1\}$, we have

$$\sum_{\{V \in \mathcal{V}_U : V \ni i\}} g_U(V) = \sum_{\{V \in \mathcal{V}_U : V \supseteq \{1,i\}\}} g_U(V) = \sum_{\{\tilde{V} \in \tilde{\mathcal{V}}_{\tilde{U}} : \tilde{V} \ni i\}} \tilde{g}_{\tilde{U}}(\tilde{V}) \geq 1 \tag{B.17}$$

and

$$\sum_{\{V \in \mathcal{V}_U : V \ni 1\}} g_U(V) \geq \sum_{\{V \in \mathcal{V}_U : V \supseteq \{1,i\}\}} g_U(V) \geq 1. \tag{B.18}$$

This completes the proof of Case 2.

Case 3: $\frac{\lambda_2+\cdots+\lambda_N}{\alpha-1} < \lambda_1 \leq \frac{\lambda_2+\cdots+\lambda_N}{\alpha-2}$. In this case, we shall need the following notations. For any $U \in \Omega_N^{(\alpha)}$ and $\tau \in \{1, \ldots, \alpha\}$, denote by $a_U(\tau)$ the *smallest* positive integer l such that

$$|\{1, \ldots, l\} \cap U| = \tau. \tag{B.19}$$

Let

$$W_\tau(U) := U \setminus \{a_U(\tau)\} \tag{B.20}$$

so $W_\tau(U) \in \Omega_N^{(\alpha-1)}$. For each $U \in \Omega_N^{(\alpha)}$, $m \in \{2, \ldots, \alpha\}$, and $\tau \in \{m, \ldots, \alpha\}$, let $\xi_{U,m,\tau} : \Omega_N^{(\alpha-1)} \to \mathbb{R}^+$ where

$$\xi_{U,m,\tau}(V) := \begin{cases} \frac{b_{m-1}^{(\alpha)} - b_m^{(\alpha)}}{f_\alpha(\boldsymbol{\lambda})}, & \text{if } V = W_\tau(U) \\ 0, & \text{otherwise} \end{cases} \tag{B.21}$$

$$b_l^{(\alpha)} := \lambda_l - \tilde{\lambda}_l \tag{B.22}$$

$$\text{and} \quad \tilde{\lambda}_l := \sum_{\{U \in \Omega_N^{(\alpha)}, U \ni l\}} c_{\boldsymbol{\lambda}}(U), \quad \forall l = 1, \ldots, L. \tag{B.23}$$

Let

$$\beta := \sum_{m=2}^{\alpha-1} (b_1^{(\alpha)} - b_m^{(\alpha)}). \tag{B.24}$$

Consider the collection of functions $\{g_U : U \in \Omega_N^{(\alpha)}\}$ where

$$g_U(V) := \left(1 - \frac{\beta}{f_\alpha(\boldsymbol{\lambda})}\right) \frac{1}{\alpha - 1} + \sum_{m=2}^{\alpha} \sum_{\tau=m}^{\alpha} \xi_{U,m,\tau}(V), \quad \forall V \in \mathcal{V}_U. \tag{B.25}$$

This gives

$$c_{\boldsymbol{\lambda}}(V) = \left(1 - \frac{\beta}{f_\alpha(\boldsymbol{\lambda})}\right) \sum_{U \in \mathcal{U}_V} \frac{c_{\boldsymbol{\lambda}}(U)}{\alpha - 1} + \sum_{U \in \mathcal{U}_V} \sum_{m=2}^{\alpha} \sum_{\tau=m}^{\alpha} \xi_{U,m,\tau}(V) c_{\boldsymbol{\lambda}}(U), \quad \forall V \in \Omega_N^{(\alpha-1)}. \tag{B.26}$$

By [5, Eq. (55)], $c_{\boldsymbol{\lambda}}^{(\alpha-1)}$ constructed as such is an optimal solution to the linear program (2.19) with α replaced by $\alpha - 1$. It remains to show that g_U is a fractional cover of $(U, \mathcal{V}_U)$ for any $U \in \Omega_N^{(\alpha)}$

Note that for any $i \in U$,

$$\sum_{\{V \in \mathcal{V}_U, V \ni i\}} \left(1 - \frac{\beta}{f_\alpha(\boldsymbol{\lambda})}\right) \frac{1}{\alpha - 1} = 1 - \frac{\beta}{f_\alpha(\boldsymbol{\lambda})} \tag{B.27}$$

and

$$\begin{aligned}
\sum_{\{V \in \mathcal{V}_U, V \ni i\}} \sum_{m=2}^{\alpha} \sum_{\tau=m}^{\alpha} \xi_{U,m,\tau}(V) &= \sum_{m=2}^{\alpha} \sum_{\tau=m}^{\alpha} \left(\sum_{\{V \in \mathcal{V}_U, V \ni i\}} \xi_{U,m,\tau}(V) \right) & \text{(B.28)} \\
&= \sum_{m=2}^{\alpha} \sum_{\tau=m}^{\alpha} \frac{b_{m-1}^{(\alpha)} - b_m^{(\alpha)}}{f_\alpha(\boldsymbol{\lambda})} 1_{\{a_U(\tau) \neq i\}} & \text{(B.29)} \\
&= \sum_{m=2}^{\alpha} \frac{b_{m-1}^{(\alpha)} - b_m^{(\alpha)}}{f_\alpha(\boldsymbol{\lambda})} \left(\sum_{\tau=m}^{\alpha} 1_{\{a_U(\tau) \neq i\}} \right) & \text{(B.30)}
\end{aligned}$$

$$\geq \sum_{m=2}^{\alpha} \frac{b_{m-1}^{(\alpha)} - b_m^{(\alpha)}}{f_\alpha(\boldsymbol{\lambda})}(\alpha - m) \tag{B.31}$$

$$= \frac{\beta}{f_\alpha(\boldsymbol{\lambda})} \tag{B.32}$$

where (B.32) follows from [5, Eq. (66)]. Combing (B.27) and (B.32) gives

$$\sum_{\{V \in \mathcal{V}_U, V \ni i\}} g_U(V) \geq 1 - \frac{\beta}{f_\alpha(\boldsymbol{\lambda})} + \frac{\beta}{f_\alpha(\boldsymbol{\lambda})} = 1. \tag{B.33}$$

We thus conclude that g_U as defined in (B.25) is indeed a fractional cover of $(U, \mathcal{V}_U)$ for any $U \in \Omega_N^{(\alpha)}$. This completes the proof of Case 3.

APPENDIX C

HMDC: 2 STRONG ENCODERS AND 4 WEAK ENCODERS

In this part, consider the case where $r = 2$ and there are 4 discrete memoryless sources to be encoded by 2 strong encoders and 4 weak encoders. We briefly sketch the computation that is used to show superposition coding is optimal in achieving the points on the boundary of the entire rate region where the supporting hyperplanes are only characterized by two coefficients in the sense that

$$\boldsymbol{\lambda} = (\underbrace{\lambda_1, \ldots, \lambda_1}_{|A|}, \underbrace{\lambda_2, \ldots, \lambda_2}_{|B|}).$$

For this particular problem, we discuss the optimal solution for each $\alpha = \{1, 2, 3, 4\}$.

- $\alpha = 1$.

  We can rewrite the constraints in (5.15) explicitly as

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} c_{\boldsymbol{\lambda}}^{(1)}(U) \leq \boldsymbol{\lambda}^T \tag{C.1}$$

It is not hard to see that $f_1(\boldsymbol{\lambda}) = 2\lambda_1 + \lambda_2$ with an optimal solution such that

$$c_{\boldsymbol{\lambda}}^{(1)}(U) = \begin{cases} \lambda_1, & u^T w = 1,\, u_i = 1, i \in \{1,2\}; \\ \frac{\lambda_2}{3}, & u^T w = 1,\, u_i = 0, i = 1,2. \end{cases} \tag{C.2}$$

- $\alpha = 2$.

  Similarly, we can get $f_2(\boldsymbol{\lambda}) = \lambda_1 + \lambda_2$, where an optimal solution has the following form

  $$c_{\boldsymbol{\lambda}}^{(2)}(U) = \begin{cases} \lambda_1, & u^T w = 2,\, u_i = 1, i = 1,2; \\ \lambda_2, & u^T w = 2,\, u_i = 0, i = 1,2. \end{cases} \tag{C.3}$$

- $\alpha = 3$.

  This case is a little bit complicated since the optimal solution depends on the order of λ_1 and λ_2.

  - If $\frac{\lambda_2}{2} \le \lambda_1 \le 2\lambda_2$, $f_3(\boldsymbol{\lambda}) = \frac{2}{3}(\lambda_1 + \lambda_2)$ and an optimal solution is

    $$c_{\boldsymbol{\lambda}}^{(3)}(U) = \begin{cases} \frac{2\lambda_1 - \lambda_2}{9}, & u^T w = 3,\, u_i = 1, i = 1,2; \\ \frac{2\lambda_2 - \lambda_1}{3}, & u^T w = 3,\, u_i = 0, i \in \{1,2\}. \end{cases} \tag{C.4}$$

  - If $\lambda_1 < \frac{\lambda_2}{2}$, $f_3(\boldsymbol{\lambda}) = 2\lambda_1$ and an optimal solution is

    $$c_{\boldsymbol{\lambda}}^{(3)}(U) = \lambda_1, u^T w = 3,\, u_i = 0, i \in \{1,2\}. \tag{C.5}$$

  - If $\lambda_1 > 2\lambda_2$, $f_3(\boldsymbol{\lambda}) = 2\lambda_2$ and an optimal solution is

    $$c_{\boldsymbol{\lambda}}^{(3)}(U) = \frac{\lambda_2}{3}, u^T w = 3,\, u_i = 1, i = 1,2. \tag{C.6}$$

- $\alpha = 4$.

  In this case, $f_4(\boldsymbol{\lambda}) = \min\{\lambda_1, \lambda_2\}$ and the optimal solution is unique and

$$c_{\boldsymbol{\lambda}}^{(4)}(U) = \min\{\lambda_1, \lambda_2\},\ u_i = 1\,\forall\, i = \{1, \ldots, 6\}. \tag{C.7}$$

With all the optimal solutions for $\alpha = 1$ to 4 discussed above, for any λ_1 and λ_2, we can easily establish the same results as those in Corollary 1 and thus have the desired inequality chain to show the optimality of superposition coding.