On the Linear Codebook-Level Duality Between Slepian–Wolf Coding and Channel Coding

Jun Chen, *Member, IEEE*, Da-ke He, Ashish Jagmohan, Luis A. Lastras-Montaño, *Senior Member, IEEE*, and En-hui Yang, *Fellow, IEEE*

***Abstract*—In this paper, it is shown that each Slepian–Wolf coding problem is related to a dual channel coding problem in the sense that the sphere packing exponents, random coding exponents, and correct decoding exponents in these two problems are mirror-symmetrical to each other. This mirror symmetry is interpreted as a manifestation of the linear codebook-level duality between Slepian–Wolf coding and channel coding. Furthermore, this duality, in conjunction with a systematic analysis of the expurgated exponents, reveals that nonlinear Slepian–Wolf codes can strictly outperform linear Slepian–Wolf codes in terms of rate-error tradeoff at high rates. The linear codebook-level duality is also established for general sources and channels.**

***Index Terms*—Channel coding, duality, error exponent, linear code, reliability function, Slepian–Wolf coding.**

I. Introduction

CONSIDER the problem shown in Fig. 1. Here the encoder compresses its observation $X^n = (X_1, X_2, \ldots, X_n)$ and sends the compressed data at rate R to the decoder; the decoder tries to reconstruct X^n given the side information $Y^n = (Y_1, Y_2, \ldots, Y_n)$ and the data from the encoder. Let $\mathcal{X}$ (resp., $\mathcal{Y}$) be the alphabet of X_t (resp., Y_t) for all t. Throughout this paper, $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be finite with $\mathcal{X} = \mathbb{Z}_M$ and $\mathcal{Y} = \mathbb{Z}_N$, where $\mathbb{Z}_K \triangleq \{0, 1, \ldots, K-1\}$ for any positive integer K.

This problem was first studied by Slepian and Wolf [1]; therefore, it is often referred to as the Slepian–Wolf coding problem. In the case where the source $\{(X_t, Y_t)\}_{t=1}^{\infty}$ is an independent identically distributed (i.i.d.) process, Slepian and Wolf [1] proved a surprising result that the minimum rate for reconstructing X^n at the decoder with the decoding error probability decaying to zero asymptotically as the block length n goes to infinity is the same as the case where the side information Y^n is available at both the encoder and the decoder. The result was further generalized by Cover [2] to the stationary and ergodic sources. Slepian–Wolf coding for general sources (not necessarily stationary) was studied by Miyake and Kanaya [3].

For most of this paper, the source $\{(X_t, Y_t)\}_{t=1}^{\infty}$ is assumed to be an i.i.d. process with zeroth-order joint probability distribution P_{XY}. In this setting, the minimum achievable rate of Slepian–Wolf coding is $H(X|Y)$ (i.e., the conditional entropy of X given Y), which is often referred to as the Slepian–Wolf limit. We will only consider the case $H(X|Y) > 0$ since otherwise X^n can be perfectly reconstructed from Y^n. Let P_X and P_Y be the marginal probability distributions induced by P_{XY}. Without loss of generality, we will assume $P_Y(y) > 0$ for all $y \in \mathcal{Y}$. It is well known that Slepian–Wolf coding is closely related to channel coding. Intuitively, one can view the side information Y^n as the output generated by X^n via the virtual channel $P_{Y|X}$, where $P_{Y|X}$ is the conditional probability distribution of Y given X induced by P_{XY}. Indeed, this viewpoint was adopted in the seminal paper by Slepian and Wolf. It is worth noting that to justify this viewpoint, one generally needs to use nonlinear Slepian–Wolf codes.[1] In practice, linear codes are commonly used for Slepian–Wolf coding, which is supported by the classic result by Csiszár [5] that linear codes suffice to achieve the Slepian–Wolf limit. Actually in the case of the binary symmetric source, this fact was known to Slepian and Wolf [6], and inspired their proof for the general case. However, if linear codes are used for Slepian–Wolf coding, then except for some special cases the relation with channel coding for channel $P_{Y|X}$ breaks down. Indeed, this phenomenon was already evident in an example constructed by Wyner [7] in which Slepian–Wolf coding with source distribution P_{XY} is related to channel coding for channel $P_{X|Y}$ (not $P_{Y|X}$), where $P_{X|Y}$ is the conditional distribution of X given Y induced by P_{XY}. Unfortunately, this subtle issue is often overlooked, especially by the practitioners in the area of Slepian–Wolf code design. Moreover, although linear codes suffice to achieve the Slepian–Wolf limit, its performance relative to the optimum Slepian–Wolf codes in terms of rate-error tradeoff is still not well understood.

We will show that in the linear coding framework, each Slepian–Wolf coding problem is equivalent to a channel coding

Manuscript received March 23, 2008; revised April 24, 2009. Current version published November 20, 2009.

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1 Canada (e-mail: junchen@ece.mcmaster.ca).

D.-k. He was with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA. He is now with SlipStream Data, a subsidiary of Research In Motion, Waterloo, ON N2L 5Z5, Canada (e-mail: dhe@rim.com).

A. Jagmohan and L. A. Lastras-Montaño are with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: ashishja, lastrasl@us.ibm.com).

E.-h. Yang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: ehyang@uwaterloo.ca).

Communicated by E. Ordentlich, Associate Editor for Source Coding.

Color versions of Figures 2 and 3 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2032815

[1] In particular, Ahlswede and Dueck [4] established an intimate connection between Slepian–Wolf coding with source distribution P_{XY} and channel coding for channel $P_{Y|X}$ using constant composition codes; they further showed that this connection implies an interesting duality relationship between the error exponents of these two coding problems. It is instructive to compare the results in the present work with those in [4] to see the fundamental difference between linear Slepian–Wolf codes and nonlinear Slepian–Wolf codes.

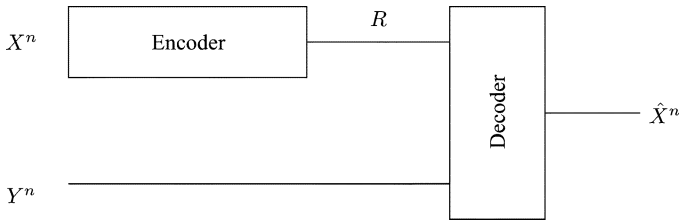

Fig. 1. Slepian–Wolf coding.

problem for a channel which is in general neither $P_{Y|X}$ nor $P_{X|Y}$. Specifically, given source distribution P_{XY}, we construct a dual channel $Q_{V|U}$ in which X serves as the additive channel noise while Y serves as the channel state (provided to the receiver). It is worth noting that such a construction is not completely new. Indeed, in the case where the side information is absent, it was observed by Ancheta [8] that each source coding problem can be converted to a channel coding problem for an additive noise channel if linear codes are used; moreover, Csiszár [5] showed that the reliability function of a discrete memoryless source is mirror-symmetrical to the sphere packing exponent of the corresponding additive noise channel while the expurgated exponent of this additive noise channel can be used to upper-bound the best error exponent attainable for linear codes at high rates in the original source coding problem. It will be seen that the construction of the dual channel $Q_{V|U}$ allows one to generalize the insights of Ancheta and Csiszár to the Slepian–Wolf coding scenario.

The main contributions of this work are the following. We establish a linear codebook-level duality and use it to develop a conceptual framework for interpreting many existing results in Slepian–Wolf coding and channel coding. Furthermore, we show that the linear codebook-level duality can be used to distinguish the performance limits of linear Slepian–Wolf codes and nonlinear Slepian–Wolf codes; specifically, it reveals that in the high rate regime nonlinear Slepian–Wolf codes can outperform linear Slepian–Wolf codes in terms of rate-error tradeoff.

The rest of this paper is divided into four sections. In Section II, we couple each Slepian–Wolf coding problem with a channel coding problem by establishing a linear codebook-level duality. This duality is leveraged to interpret the mirror symmetry exhibited by the error exponents in these two problems. A systematic analysis of the expurgated exponents is provided in Section III, which along with the linear codebook-level duality between Slepian–Wolf coding and channel coding reveals that nonlinear Slepian–Wolf codes can strictly outperform linear Slepian–Wolf codes in terms of rate-error tradeoff at high rates. A subtle difference between the roles played by linear codes in Slepian–Wolf coding and channel coding is observed. The difference in the performances between linear and nonlinear Slepian–Wolf coding is further illustrated via an example in Section IV. The linear codebook-level duality is extended to the general sources and channels in Section V. We conclude the paper in Section VI. Throughout this paper, the logarithm function is to the base e unless specified otherwise.

II. Mirror Symmetry and Linear Codebook-Level Duality

A Slepian–Wolf code consists of an encoding function $f_n : \mathcal{X}^n \to \mathcal{I}_n$ and a decoding function $g_n : \mathcal{Y}^n \times \mathcal{I}_n \to \mathcal{X}^n$. The rate of Slepian–Wolf code (f_n, g_n) is defined as

$$R(f_n) = \frac{1}{n} \log |\mathcal{I}_n|$$

and the decoding error probability is defined as

$$P_e(f_n, g_n, P_{XY}) = \Pr\{\hat{X}^n \neq X^n\}$$

where $\hat{X}^n = g_n(Y^n, f_n(X^n))$.

We will mainly use the following two decoding functions:

1) maximum *a posteriori* (MAP) decoding

$$\hat{X}^n = \arg \max_{x^n : f_n(x^n) = f_n(X^n)} \sum_{t=1}^{n} \log P_{X|Y}(x_t|Y_t)$$
$$= \arg \max_{x^n : f_n(x^n) = f_n(X^n)} \sum_{t=1}^{n} \log P_{XY}(x_t, Y_t);$$

2) minimum-entropy (ME) decoding

$$\hat{X}^n = \arg \min_{x^n : f_n(x^n) = f_n(X^n)} H(x^n, Y^n)$$

where $H(x^n, Y^n)$ is the entropy induced by the empirical joint distribution of (x^n, Y^n). For example, if the empirical joint distribution of (x^n, Y^n) is $P_{\tilde{X}\tilde{Y}}$, then we have $H(x^n, Y^n) = H(\tilde{X}, \tilde{Y})$.

In both cases, ties are broken in an arbitrary manner. Note that MAP decoding is the optimal decoding rule, and ME decoding is a universal decoding rule, i.e., where decoder does not need to know P_{XY}.

Definition 1: Given a joint probability distribution P_{XY}, we say rate R is achievable if given any $\delta > 0$, there exists a sequence of Slepian–Wolf codes $\{(f_n, g_n)\}$ such that for all sufficiently large n

$$R(f_n) \leq R + \delta$$
$$P_e(f_n, g_n, P_{XY}) \leq \delta.$$

The minimum achievable rate, denoted by $R(P_{XY})$, is referred to as the Slepian–Wolf limit.

Remark: It is well known [1] that $R(P_{XY}) = H(X|Y)$.

Definition 2: Given a joint probability distribution P_{XY}, we say an error exponent $E \geq 0$ is achievable at rate R if given any $\delta > 0$, there exists a sequence of Slepian–Wolf codes $\{(f_n, g_n)\}$ such that for all sufficiently large n

$$R(f_n) \leq R + \delta$$
$$P_e(f_n, g_n, P_{XY}) \leq e^{-n(E-\delta)}.$$

The maximum achievable error exponent at rate R is denoted by $E(P_{XY}, R)$, which is referred to as the reliability function of Slepian–Wolf coding. Similarly, we say a correct decoding exponent $E^* \geq 0$ is achievable at rate R if given any $\delta > 0$,

there exists a sequence of Slepian–Wolf codes $\{(f_n, g_n)\}$ such that for all sufficiently large n

$$R(f_n) \le R + \delta$$
$$P_e(f_n, g_n, P_{XY}) \le 1 - e^{-n(E^*+\delta)}.$$

The minimum achievable correct decoding exponent at rate R is denoted by $E^*(P_{XY}, R)$, which is referred to as the reliability function below the Slepian–Wolf limit since $E^*(P_{XY}, R)$ is positive if and only if $R < R(P_{XY})$ (as will be seen). We will simply call $E^*(P_{XY}, R)$ the correct decoding exponent when no confusion can arise.

It is known that the reliability function $E(P_{XY}, R)$ is lower bounded by the random coding exponent and upper bounded by the sphere packing exponent. Let $|t|^+ = \max\{t, 0\}$. The random coding exponent [5], [9] is defined as

$$E_{\rm rc}(P_{XY}, R) = \min_{P_{\tilde{X}\tilde{Y}}} [D(P_{\tilde{X}\tilde{Y}} \| P_{XY}) + |R - H(\tilde{X}|\tilde{Y})|^+] \quad (1)$$

where the minimization is over all probability distributions $P_{\tilde{X}\tilde{Y}}$ on $\mathcal{X} \times \mathcal{Y}$. Alternatively, the random coding exponent can be written in Gallager's form [10]

$$E_{\rm rc}(P_{XY}, R) = \max_{0 \le \rho \le 1} \left\{ -\log \sum_y \left[\sum_x P_{XY}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} + \rho R \right\}.$$

The sphere packing exponent is given by [9]

$$E_{\rm sp}(P_{XY}, R) = \min_{P_{\tilde{X}\tilde{Y}}: H(\tilde{X}|\tilde{Y}) \ge R} D(P_{\tilde{X}\tilde{Y}} \| P_{XY})$$

or, alternatively [10]

$$E_{\rm sp}(P_{XY}, R) = \sup_{\rho > 0} \left\{ -\log \sum_y \left[\sum_x P_{XY}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} + \rho R \right\}.$$

To make the connection between the random coding exponent and the sphere packing exponent explicit, we will write them in the parametric forms [10] as shown in (2) and (3) at the bottom of the page, where the joint distribution of $(X^{(\rho)}, Y^{(\rho)})$ is specified by

$$P_{Y^{(\rho)}}(y) = \frac{P_Y(y) \left[\sum_{x'} P_{X|Y}(x'|y)^{\frac{1}{1+\rho}} \right]^{1+\rho}}{\sum_{y'} P_Y(y') \left[\sum_{x'} P_{X|Y}(x'|y')^{\frac{1}{1+\rho}} \right]^{1+\rho}}, \quad y \in \mathcal{Y}, \quad (4)$$

$$P_{X^{(\rho)}|Y^{(\rho)}}(x|y) = \frac{P_{X|Y}(x|y)^{\frac{1}{1+\rho}}}{\sum_{x'} P_{X|Y}(x'|y)^{\frac{1}{1+\rho}}}, \quad x \in \mathcal{X}, \quad y \in \mathcal{Y}. \quad (5)$$

Define the critical rate $R_{\rm cr}(P_{XY}) = \max\{R : E_{\rm rc}(P_{XY}, R) = R_{\rm sp}(P_{XY}, R)\}$. It is clear that $R_{\rm cr}(P_{XY}) = H(X^{(\rho)}|Y^{(\rho)})|_{\rho=1}$. Note that $E_{\rm rc}(P_{XY}, R)$ and $E_{\rm sp}(P_{XY}, R)$ coincide when $R \in [H(X|Y), R_{\rm cr}(P_{XY})]$, and the reliability function $E(P_{XY}, R)$ is completely determined in this rate regime.

Define $R_\infty(P_{XY}) = \sup\{R : E_{\rm sp}(P_{XY}, R) < \infty\}$. It is clear that $R_\infty(P_{XY}) = H(X^{(\rho)}|Y^{(\rho)})|_{\rho=\infty}$, where $\rho = \infty$ should be interpreted as $\rho \to \infty$ (which corresponds to the scenario that the slope of $E_{\rm sp}(P_{XY}, R)$ goes to infinity). Let $\mathcal{A}_y = \{x : P_{X|Y}(x|y) > 0\}$ for $y \in \mathcal{Y}$, and $\mathcal{B} = \{y : |\mathcal{A}_y| = \max_{y'} |\mathcal{A}_{y'}|\}$. It can be verified that

$$P_{Y^{(\rho)}}(y)|_{\rho=\infty} = \begin{cases} \frac{P_Y(y)}{\sum_{y' \in \mathcal{B}}} P_Y(y'), & \text{if } y \in \mathcal{B} \\ 0, & \text{otherwise} \end{cases}$$

$$P_{X^{(\rho)}|Y^{(\rho)}}(x|y)|_{\rho=\infty} = \begin{cases} \frac{1}{|\mathcal{A}_y|}, & \text{if } x \in \mathcal{A}_y \\ 0, & \text{otherwise} \end{cases}$$

which implies

$$R_\infty(P_{XY}) = \max_y \log |\{x : P_{X|Y}(x|y) > 0\}|. \quad (6)$$

The sphere packing exponent $E_{\rm sp}(P_{XY}, R)$ continues to be an upper bound on the reliability function even if the side information is available at both the encoder and the decoder; moreover, in this case, it is also achievable [11], [10]. Therefore, we can interpret $R_\infty(P_{XY})$ as the minimum achievable zero-error rate by fixed-rate codes when the side information is known at both the encoder and the decoder, which is intuitively obvious from the expression of $R_\infty(P_{XY})$ in (6).

The correct decoding exponent $E^*(P_{XY}, R)$ has been completely characterized [12], and is given by

$$E^*(P_{XY}, R) = \min_{P_{\tilde{X}\tilde{Y}}} [D(P_{\tilde{X}\tilde{Y}} \| P_{XY}) + |H(\tilde{X}|\tilde{Y}) - R|^+].$$

An alternative expression of $E^*(P_{XY}, R)$ is

$$E^*(P_{XY}, R) = \max_{-1 \le \rho \le 0} \left\{ -\log \sum_y \left[\sum_x P_{XY}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} + \rho R \right\}.$$

We can also write $E^*(P_{XY}, R)$ in the parametric form as shown in the equation at the bottom of the next page, where $P_{X^{(\rho)}Y^{(\rho)}}$ is given in (4) and (5) with $\rho \in [-1, 0]$. Specifically, we define $R^*_{\rm cr}(P_{XY})$ as the largest R at which the convex curve $E^*(P_{XY}, R)$ meets its supporting line of slope -1, i.e.,

$$E_{\rm sp}(P_{XY}, R) = D\left(P_{X^{(\rho)}Y^{(\rho)}} \| P_{XY}\right), \qquad R = H\left(X^{(\rho)}|Y^{(\rho)}\right) \quad (2)$$

$$E_{\rm rc}(P_{XY}, R) = \begin{cases} D(P_{X^{(\rho)}Y^{(\rho)}} \| P_{XY}), & \text{if } H(X|Y) \le R \le H(X^{(\rho)}|Y^{(\rho)})|_{\rho=1} \\ -\log \sum_y \left[\sum_x \sqrt{P_{XY}(x,y)} \right]^2 + R, & \text{if } R > H(X^{(\rho)}|Y^{(\rho)})|_{\rho=1} \end{cases} \quad (3)$$

$R^*_{\rm cr}(P_{XY}) = H(X^{(\rho)}|Y^{(\rho)})|_{\rho=-1}$ with $\rho = -1$ interpreted as $\rho \downarrow -1$. It can be verified that

$$P_{Y^{(\rho)}}(y)|_{\rho=-1} = \frac{\max_{x'} P_{XY}(x',y)}{\sum_{y'} \max_{x'} P_{XY}(x',y')}, \qquad y \in \mathcal{Y}$$

$$P_{X^{(\rho)}|Y^{(\rho)}}(x|y)|_{\rho=-1} = \begin{cases} \dfrac{1}{|\mathcal{A}^*_y|}, & \text{if } x \in \mathcal{A}^*_y \\ 0, & \text{otherwise} \end{cases}$$

where $\mathcal{A}^*_y = \{x : P_{X|Y}(x|y) = \max_{x'} P_{X|Y}(x'|y)\}$. Note that for any $P_{\tilde{X}\tilde{Y}}$ satisfying

$$P_{\tilde{Y}}(y) = P_{Y^{(\rho)}}(y)|_{\rho=-1}, \qquad y \in \mathcal{Y} \tag{7}$$

$$P_{\tilde{X}|\tilde{Y}}(x|y) = 0, \qquad x \in \mathcal{X}, \quad x \notin \mathcal{A}^*_y \tag{8}$$

we have

$$\begin{aligned} &D(P_{\tilde{X}\tilde{Y}}||P_{XY}) + H(\tilde{X}|\tilde{Y}) \\ &\quad = D(P_{\tilde{Y}}||P_Y) - \sum_{x,y} P_{\tilde{X}\tilde{Y}}(x,y) \log P_{X|Y}(x|y) \\ &\quad = D\left(P_{Y^{(\rho)}}||P_Y\right) - \sum_{x,y} P_{X^{(\rho)}Y^{(\rho)}}(x,y) \log P_{X|Y}(x|y) \Big|_{\rho=-1} \\ &\quad = D\left(P_{X^{(\rho)}Y^{(\rho)}}||P_{XY}\right) + H(X^{(\rho)}|Y^{\rho}) \Big|_{\rho=-1} \\ &\quad = -\log \sum_y \max_x P_{XY}(x,y). \end{aligned}$$

Moreover, for any $R \in [0, R^*_{\rm cr}(P_{XY})]$, it is always possible to find a joint probability distribution $P_{\tilde{X}\tilde{Y}}$ satisfying (7), (8), and the condition $H(\tilde{X}|\tilde{Y}) = R$. Therefore, one can readily show that $E^*(P_{XY}, R)$ can be equivalently written as

$$E^*(P_{XY}, R) = \min_{P_{\tilde{X}\tilde{Y}}: H(\tilde{X}|\tilde{Y}) \le R} D(P_{\tilde{X}\tilde{Y}}||P_{XY})$$

which is exactly the expression derived in [12].

The following result can be proved by direct verification.

Theorem 1: The following four statements are equivalent[2]:

1) $R_{\rm cr}(P_{XY}) = R(P_{XY})$;
2) $R_{\infty}(P_{XY}) = R(P_{XY})$;
3) $R^*_{\rm cr}(P_{XY}) = R(P_{XY})$;
4) $|\mathcal{A}_y|$ does not depend on y, and $P_{X|Y}(x|y) = \frac{1}{|\mathcal{A}_y|}$ for $x \in \mathcal{A}_y$, where $\mathcal{A}_y = \{x : P_{X|Y}(x|y) > 0\}$.

We need to introduce a few definitions related to channel coding before proceeding to discuss its connection with Slepian–Wolf coding.

Let $Q_{V|U} : \mathcal{U} \to \mathcal{V}$ be a discrete memoryless channel with input alphabet $\mathcal{U}$ and output alphabet $\mathcal{V}$. A block code for channel $Q_{V|U}$ consists of a codebook $\mathcal{C}_n \subseteq \mathcal{U}^n$ and a decoding function $\phi_n : \mathcal{V}^n \to \mathcal{C}_n$. The rate of channel code $(\mathcal{C}_n, \phi_n)$ is defined as

$$R(\mathcal{C}_n) = \frac{1}{n} \log |\mathcal{C}_n|$$

and the decoding error probability is defined as

$$P_e(\mathcal{C}_n, \phi_n, Q_{V|U}) = \frac{1}{|\mathcal{C}_n|} \sum_{u^n \in \mathcal{C}_n} \Pr\{\hat{U}^n \ne u^n | u^n \text{ is transmitted}\} \tag{9}$$

with $\hat{U}^n = \phi_n(V^n)$, where V^n is the channel output. In particular, $\phi_n(\cdot)$ is a maximum-likelihood (ML) decoder if

$$\hat{U}^n = \arg \max_{u^n \in \mathcal{C}_n} \sum_{t=1}^{n} \log Q_{V|U}(V_t|u_t)$$

where ties are broken in an arbitrary manner. Note that ML decoding is the optimal decoding rule in the current setting since the transmitted codeword is assumed to have a uniform prior distribution [cf., (9)].

Definition 3: Given a channel $Q_{V|U}$, we say rate R is achievable if given any $\delta > 0$, there exists a sequence of channel codes $\{(\mathcal{C}_n, \phi_n)\}$ such that for all sufficiently large n

$$R(\mathcal{C}_n) \ge R - \delta$$
$$P_e(\mathcal{C}_n, \phi_n, Q_{V|U}) \le \delta.$$

The maximum achievable rate, denoted by $C(Q_{V|U})$, is referred to as the channel capacity.

Remark: It is well known (see, e.g., [13]) that $C(Q_{V|U}) = \max_{Q_U} I(U;V)$, where the maximization is over all probability distributions Q_U on $\mathcal{U}$.

Definition 4: Given a channel $Q_{V|U}$, we say an error exponent $E \ge 0$ is achievable at rate R if given any $\delta > 0$, there exists a sequence of channel codes $\{(\mathcal{C}_n, \phi_n)\}$ such that for all sufficiently large n

$$R(\mathcal{C}_n) \ge R - \delta$$
$$P_e(\mathcal{C}_n, \phi_n, Q_{V|U}) \le e^{-n(E-\delta)}.$$

The maximum achievable error exponent at rate R is denoted by $E(Q_{V|U}, R)$, which is referred to as the reliability function of channel $Q_{V|U}$. Similarly, we say a correct decoding exponent $E^* \ge 0$ is achievable at rate R if given any $\delta > 0$, there exists a sequence of channel codes $\{(\mathcal{C}_n, \phi_n)\}$ such that for all sufficiently large n

$$R(\mathcal{C}_n) \ge R - \delta$$
$$P_e(\mathcal{C}_n, \phi_n, Q_{V|U}) \le 1 - e^{-n(E^*+\delta)}.$$

The minimum achievable correct decoding exponent at rate R is denoted by $E^*(Q_{V|U}, R)$, which is referred to as the reliability function above the channel capacity since $E^*(Q_{V|U}, R)$ is positive if and only if $R > C(Q_{V|U})$ (as will be seen). We will simply call $E^*(Q_{V|U}, R)$ the correct decoding exponent when no confusion can arise.

[2]Here we assume $P_Y(y) > 0$ for all $y \in \mathcal{Y}$; otherwise, statement 4) has to be modified by restricting y in the set $\{y' : P_Y(y') > 0\}$.

$$E^*(P_{XY}, R) = \begin{cases} D(P_{X^{(\rho)}Y^{(\rho)}}||P_{XY}), & \text{if } H\left(X^{(\rho)}|Y^{(\rho)}\right)|_{\rho=-1} \le R \le H(X|Y) \\ -\log \sum_y \max_x P_{XY}(x,y) - R, & \text{if } R < H\left(X^{(\rho)}|Y^{(\rho)}\right)|_{\rho=-1} \end{cases}$$

An instructive way to understand the connection between Slepian–Wolf coding and channel coding is to couple each Slepian–Wolf coding problem with a channel coding problem. Specifically, given a joint probability distribution $P_{XY}: \mathcal{X} \times \mathcal{Y}$, we define a dual channel $Q_{V|U}: \mathcal{U} \to \mathcal{V}$ with $V_t = (U_t +_M X_t, Y_t)$, where $\mathcal{U} = \mathcal{X}$, $\mathcal{V} = \mathcal{X} \times \mathcal{Y}$, and $+_M$ (resp., $-_M$) denotes modulo-M addition (resp., subtraction). The channel input U_t is assumed to be independent of (X_t, Y_t). Intuitively, one may view X_t as the additive channel noise and Y_t as the channel state (provided to the receiver) at time t. It can be shown that the class of dual channels $Q_{V|U}$ induced by all possible joint distributions P_{XY} is equivalent to the class of cyclic-symmetric channels defined in [14].

The following result shows the connection between the Slepian–Wolf limit associated with the joint probability distribution P_{XY} and the capacity of the dual channel $Q_{V|U}$.

Theorem 2: For any joint probability distribution P_{XY} and its dual channel $Q_{V|U}$, we have $R(P_{XY}) = \log M - C(Q_{V|U})$.

Remark: It is well known that $R(P_{XY})$ [i.e., $H(X|Y)$] continues to be the minimum achievable rate even if the side information is available at the encoder. Interestingly, it can be verified that the capacity of the dual channel $Q_{V|U}$ is also unaffected even if the channel state is provided to the transmitter.

Proof: By the cyclic symmetry of $Q_{V|U}$, the capacity-achieving input distribution is the uniform distribution on $\mathcal{U}$. Now it can be readily verified that

$$\begin{aligned} C(Q_{V|U}) &= \max_{Q_U} I(U;V) \\ &= \max_{Q_U} I(U; U +_M X, Y) \\ &= \max_{Q_U} H(U +_M X, Y) - H(X,Y) \\ &= \log M + H(Y) - H(X,Y) \\ &= \log M - R(P_{XY}). \end{aligned}$$

The proof is complete. □

The random coding exponent, sphere packing exponent, and correct decoding exponent for a discrete memoryless channel $Q_{V|U}$ are, respectively, defined as [15], [16]

$$\begin{aligned} &E_{\mathrm{rc}}(Q_{V|U}, R) \\ &\quad = \max_{Q_{\tilde{U}}} \min_{Q_{\tilde{V}|\tilde{U}}} [D(Q_{\tilde{V}|\tilde{U}} \| Q_{V|U} | Q_{\tilde{U}}) + |I(\tilde{U};\tilde{V}) - R|^+] \\ &E_{\mathrm{sp}}(Q_{V|U}, R) \\ &\quad = \max_{Q_{\tilde{U}}} \min_{Q_{\tilde{V}|\tilde{U}}: I(\tilde{U};\tilde{V}) \le R} D(Q_{\tilde{V}|\tilde{U}} \| Q_{V|U} | Q_{\tilde{U}}) \\ &E^*(Q_{V|U}, R) \\ &\quad = \min_{Q_{\tilde{U}}} \min_{Q_{\tilde{V}|\tilde{U}}} [D(Q_{\tilde{V}|\tilde{U}} \| Q_{V|U} | Q_{\tilde{U}}) + |R - I(\tilde{U};\tilde{V})|^+]. \end{aligned}$$

Alternatively, they can be written as [13], [17]

$$\begin{aligned} E_{\mathrm{rc}}(Q_{V|U}, R) &= \max_{0 \le \rho \le 1} \left[-\rho R + \max_{Q_U} E_0(\rho, Q_U, Q_{V|U}) \right] \\ E_{\mathrm{sp}}(Q_{V|U}, R) &= \sup_{\rho > 0} \left[-\rho R + \max_{Q_U} E_0(\rho, Q_U, Q_{V|U}) \right] \\ E^*(Q_{V|U}, R) &= \max_{-1 \le \rho \le 0} \left[-\rho R + \min_{Q_U} E_0(\rho, Q_U, Q_{V|U}) \right] \end{aligned}$$

where

$$E_0(\rho, Q_U, Q_{V|U}) = -\log \sum_v \left[\sum_u Q_U(u) Q_{V|U}(v|u)^{\frac{1}{1+\rho}} \right]^{1+\rho}.$$

It is well known that the reliability function $E(Q_{V|U}, R)$ is lower bounded by $E_{\mathrm{rc}}(Q_{V|U}, R)$ and upper bounded by $E_{\mathrm{sp}}(Q_{V|U}, R)$.

Interestingly, these exponents are mirror symmetrical to their counterparts in the dual Slepian–Wolf coding problem as shown by the following result, which can be viewed as a strengthened version of Theorem 2.

Theorem 3: Give any joint probability distribution P_{XY} and its dual channel $Q_{V|U}$, we have

$$E_{\mathrm{rc}}(P_{XY}, R) = E_{\mathrm{rc}}(Q_{V|U}, \log M - R) \tag{10}$$

$$E_{\mathrm{sp}}(P_{XY}, R) = E_{\mathrm{sp}}(Q_{V|U}, \log M - R) \tag{11}$$

$$E^*(P_{XY}, R) = E^*(Q_{V|U}, \log M - R). \tag{12}$$

Remark: Equation (1) can be viewed as a generalization of (A.1) in [5].

Proof: Define $F(\rho, Q_U, Q_{V|U}) = e^{-E_0(\rho, Q_U, Q_{V|U})}$. It is easy to see that the Q_U that minimizes (resp., maximizes) $F(\rho, Q_U, Q_{V|U})$ will maximize (resp., minimizes) $E_0(\rho, Q_U, Q_{V|U})$.

For any $\rho \ge 0$, $F(\rho, Q_U, Q_{V|U})$ is a convex function of Q_U. Necessary and sufficient conditions on Q_U that minimizes $F(\rho, Q_U, Q_{V|U})$ are [13]

$$\begin{aligned} &\sum_v Q_{V|U}(v|u)^{\frac{1}{1+\rho}} \left[\sum_{u'} Q_U(u') Q_{V|U}(v|u')^{\frac{1}{1+\rho}} \right]^{\rho} \\ &\qquad \ge \sum_v \left[\sum_{u'} Q_U(u') Q_{V|U}(v|u')^{\frac{1}{1+\rho}} \right]^{1+\rho} \end{aligned} \tag{13}$$

with equality for every $u \in \mathcal{U}$ such that $Q_U(u) > 0$. Let $Q_U(u) = \frac{1}{M}$ for all $u \in \mathcal{U}$. With this choice of Q_U, it is easy to verify that

$$\begin{aligned} &\sum_v Q_{V|U}(v|u)^{\frac{1}{1+\rho}} \left[\sum_{u'} Q_U(u') Q_{V|U}(v|u')^{\frac{1}{1+\rho}} \right]^{\rho} \\ &\quad = \sum_v \left[\sum_{u'} Q_U(u') Q_{V|U}(v|u')^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &\quad = \frac{1}{M^\rho} \sum_y \left[\sum_x P_{XY}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} \end{aligned} \tag{14}$$

which implies that the necessary and sufficient conditions in (13) are satisfied. Therefore, we have

$$\begin{aligned} &\max_{Q_U} E_0(\rho, Q_U, Q_{V|U}) \\ &\quad = -\log \sum_v \left[\sum_u \frac{1}{M} Q_{V|U}(v|u)^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &\quad = -\log \sum_y \left[\sum_x P_{XY}(x,y)^{\frac{1}{1+\rho}} \right]^{1+\rho} + \rho \log M \end{aligned}$$

from which (10) and (11) follow directly.

For any $\rho \in [-1,0]$, $F(\rho, Q_U, Q_{V|U})$ is a concave function of Q_U. Necessary and sufficient conditions on Q_U that maximizes $F(\rho, Q_U, Q_{V|U})$ are the same as (13) with "$\geq$" replaced by "$\leq$." It follows from (14) that the necessary and sufficient conditions are satisfied if Q_U is chosen to be the uniform distribution on $\mathcal{U}$. Now (12) can be easily verified. □

For any discrete memoryless channel $Q_{V|U}$, we define $R_{\mathrm{cr}}(Q_{V|U}) = \min\{R : E_{\mathrm{rc}}(Q_{V|U}, R) = E_{\mathrm{sp}}(Q_{V|U}, R)\}$, $R_{\infty}(Q_{V|U}) = \inf\{R : E_{\mathrm{sp}}(Q_{V|U}, R) < \infty\}$, and $R^*_{\mathrm{cr}}(Q_{V|U})$ as the smallest R at which the convex curve $E^*(Q_{V|U}, R)$ meets its supporting line of slope 1. In view of Theorem 3, it is clear that for each joint probability distribution P_{XY} and its dual channel $Q_{V|U}$

$$\begin{aligned} R_{\mathrm{cr}}(P_{XY}) &= \log M - R_{\mathrm{cr}}(Q_{V|U}) \\ R_{\infty}(P_{XY}) &= \log M - R_{\infty}(Q_{V|U}) \\ R^*_{\mathrm{cr}}(P_{XY}) &= \log M - R^*_{\mathrm{cr}}(Q_{V|U}). \end{aligned}$$

In particular, it is known [13] that for any discrete memoryless channel $Q_{V|U}$

$$R_{\infty}(Q_{V|U}) = -\min_{Q_U} \max_{v} \left[\log \sum_{u: Q_{V|U}(v|u) > 0} Q_U(u) \right]. \quad (15)$$

Since in the current setting the dual channel $Q_{V|U}$ is cyclic symmetric, the minimum in (15) is attained at the uniform distribution on $\mathcal{U}$. Now it can be readily verified that

$$\begin{aligned} &R_{\infty}(Q_{V|U}) \\ &\quad = \log M - \max_{v} \log |\{u : Q_{V|U}(v|u) > 0\}| \\ &\quad = \log M - \max_{x,y} \log |\{u : Q_{V|U}(u +_M x, y|u) > 0\}| \\ &\quad = \log M - \max_{y} \log |\{x : P_{XY}(x,y) > 0\}| \\ &\quad = \log M - R_{\infty}(P_{XY}). \end{aligned}$$

Note that $R_{\infty}(Q_{V|U})$ is equal to the zero-error feedback capacity of channel $Q_{V|U}$ [18], which can be interestingly compared with the operational interpretation of $R_{\infty}(P_{XY})$ (i.e., the minimum achievable zero-error rate by fixed-rate codes when the side information is known at both the encoder and the decoder). Furthermore, it is known [13] that the following statements are equivalent:

1) $R_{\mathrm{cr}}(Q_{V|U}) = C(Q_{V|U})$;
2) $R_{\infty}(Q_{V|U}) = C(Q_{V|U})$;
3) for some capacity-achieving input distribution Q_U

$$\log \frac{Q_{V|U}(v|u)}{\sum_u Q_U(u) Q_{V|U}(v|u)} = C(Q_{V|U})$$

is satisfied for all u, v such that $Q_U(u) Q_{V|U}(v|u) > 0$.

It is also easy to show that $R_{\mathrm{cr}}(Q_{V|U}) = C(Q_{V|U})$ if and only if $R^*_{\mathrm{cr}}(Q_{V|U}) = C(Q_{V|U})$. Therefore, Theorem 1 can be viewed as a direct consequence of Theorems 2 and 3.

Now we proceed to show that the formula-level mirror symmetry exhibited in Theorems 2 and 3 is a manifestation of the fundamental linear codebook-level duality between Slepian–Wolf coding and channel coding.

Definition 5 [19]: A linear block code $\mathcal{C}_n$ of length n over $\mathbb{Z}_M$ is a subgroup of $\mathbb{Z}_M^n$, where $\mathbb{Z}_M^n$ is the group of n-tuples of elements of $\mathbb{Z}_M$ with componentwise addition.

The subgroup $\mathcal{C}_n$ partitions the group $\mathbb{Z}_M^n$ into $\frac{M^n}{|\mathcal{C}_n|}$ disjoint cosets, each of size $|\mathcal{C}_n|$. We can see that only the group property is needed in order to define $\mathcal{C}_n$ and its cosets. However, it is often more convenient and desirable if we can define them using parity check matrix. Fortunately, this is possible due to the ring structure of $\mathbb{Z}_M$.

It is shown in [20] that for any linear code $\mathcal{C}_n$ of length n over $\mathbb{Z}_M$, there exists an $n \times k$ parity check matrix $\mathbf{H}$ such that

$$c^n \mathbf{H} = 0^k \Longleftrightarrow c^n \in \mathcal{C}_n. \quad (16)$$

We can define the syndrome group $\mathcal{S}_k = \mathrm{Im}(\mathbf{H})$ using the surjective homomorphism $\mathbf{H} : \mathbb{Z}_M^n \to \mathcal{S}_k$. The syndrome group $\mathcal{S}_k$ can be used to label the cosets of $\mathcal{C}_n$. Specifically, a coset of $\mathcal{C}_n$ is denoted by $\mathcal{C}_n(s^k)(s^k \in \mathcal{S}_k)$ if

$$c^n \mathbf{H} = s^k \quad (17)$$

for all c^n in this coset. In particular, we have $\mathcal{C}_n = \mathcal{C}_n(0^k)$. Conversely, given a parity check matrix $\mathbf{H}$, we can also define a linear code and its cosets via (16) and (17).

For any linear block code $\mathcal{C}_n$ over $\mathbb{Z}_M$ with parity check matrix $\mathbf{H}$, we define a linear Slepian–Wolf encoding function $f_n : \mathbb{Z}_M^n \to \mathcal{S}_k$ with $f_n(x^n) = x^n \mathbf{H}$ for all $x^n \in \mathbb{Z}_M^n$. Note that

$$\begin{aligned} R(f_n) &= \frac{1}{n} \log |\mathcal{S}_k| \\ &= \frac{1}{n} \log \frac{M^n}{|\mathcal{C}_n|} \\ &= \log M - R(\mathcal{C}_n). \end{aligned} \quad (18)$$

Moreover, given a joint probability distribution P_{XY} and its dual channel $Q_{V|U}$, we have

$$\begin{aligned} &\arg \max_{\hat{u}^n \in \mathcal{C}_n} \sum_{t=1}^{n} \log Q_{V|U}(u_t +_M x_t, y_t|\hat{u}_t) \\ &\quad = \arg \max_{\hat{u}^n \in \mathcal{C}_n} \sum_{t=1}^{n} \log P_{XY}(u_t +_M x_t -_M \hat{u}_t, y_t) \\ &\quad = u^n +_M x^n -_M \arg \max_{\hat{x}^n : f_n(\hat{x}^n) = f_n(x^n)} \sum_{t=1}^{n} \log P_{XY}(\hat{x}_t, y_t) \end{aligned}$$

for any $u^n \in \mathcal{C}_n$, $x^n \in \mathcal{X}^n$, and $y^n \in \mathcal{Y}^n$, where the second equality follows by setting $\hat{x}^n = u^n +_M x^n -_M \hat{u}^n$ and noticing that $(u^n +_M x^n -_M \hat{u}^n)\mathbf{H} = x^n \mathbf{H}$. We can see that whether a decoding error occurs depends on (x^n, y^n) but not on u^n; furthermore, any (x^n, y^n) that leads to a channel decoding error also causes a Slepian–Wolf decoding error, and *vice versa*. Therefore, we have

$$P_e(f_n, g_n, P_{XY}) = P_e(\mathcal{C}_n, \phi_n, Q_{V|U}) \quad (19)$$

where $g_n(\cdot)$ is a MAP Slepian–Wolf decoder, and $\phi_n(\cdot)$ is an ML channel decoder. Note that each linear code $\mathcal{C}_n$ can be associated with different parity check matrices $\mathbf{H}$; however, $P_e(\mathcal{C}_n, \phi_n, Q_{V|U})$ does not depend on the choice of $\mathbf{H}$, neither does $P_e(f_n, g_n, P_{XY})$ for the induced linear Slepian–Wolf code. It can be seen that (18) and (19) together reveal a fundamental linear codebook-level duality between Slepian–Wolf coding and channel coding.

Theorem 4: The random coding exponent $E_{\rm rc}(P_{XY},R)$ and the correct decoding exponent $E^*(P_{XY},R)$ are universally attainable by linear codes over $\mathbb{Z}_M$ if M is a prime number.

Remark:

1) It is not a real restriction to assume that M is a prime number since one can always make the alphabet size to be a prime number by adding symbols of zero probability.
2) Since the random coding exponent $E_{\rm rc}(P_{XY},R)$ and the correct decoding exponent $E^*(P_{XY},R)$ are universally attainable by linear codes, they must also be attainable by linear codes under MAP decoding. Now it follows from the linear codebook-level duality that $E_{\rm rc}(Q_{V|U},R)$ and $E^*(Q_{V|U},R)$ of the dual channel $Q_{V|U}$ are attainable by linear codes under ML decoding. Moreover, linear codes are rate-error tradeoff optimal (on the exponential scale) for rates below the critical rate $R_{\rm cr}(P_{XY})$ in Slepian–Wolf coding and for rates above the critical rate $R_{\rm cr}(Q_{V|U})$ in the dual channel coding problem. As a simple corollary, linear codes suffice to achieve the Slepian–Wolf limit as well as the capacity of the dual channel. Therefore, Theorems 2 and 3 can be viewed as a natural consequence of the operational duality between Slepian–Wolf coding and channel coding in the linear coding framework.

Proof: See Appendix I. □

The following theorem provides the second-order expansion of $E(P_{XY},R)$ and $E^*(P_{XY},R)$ at the Slepian–Wolf limit, as well as that of $E(Q_{V|U},R)$ and $E^*(Q_{V|U},R)$ at the capacity of the dual channel.

Theorem 5: For any joint probability distribution P_{XY} and its dual channel $Q_{V|U}$, if $R_{\rm cr}(P_{XY})>R(P_{XY})$ (or equivalently, $R_{\rm cr}(Q_{V|U})<C(Q_{V|U})$), then

$$\lim_{r\downarrow 0}\frac{E(P_{XY},R(P_{XY})+r)}{r^2}=\lim_{r\downarrow 0}\frac{E^*(P_{XY},R(P_{XY})-r)}{r^2}=\kappa$$

$$\lim_{r\downarrow 0}\frac{E(Q_{V|U},C(Q_{V|U})-r)}{r^2}=\lim_{r\downarrow 0}\frac{E^*(Q_{V|U},C(Q_{V|U})+r)}{r^2}=\kappa$$

where

$$\kappa=\frac{1}{2}\left[-H^2(X|Y)+\sum_{x,y}P_{XY}(x,y)\log^2 P_{X|Y}(x|y)\right]^{-1}.$$

Remark: In view of Theorems 1 and 3, if $R_{\rm cr}(P_{XY})=R(P_{XY})$, then we have

$$\begin{aligned}E_{\rm rc}(P_{XY},R)&=R-R(P_{XY}),&&R\ge R(P_{XY})\\ E^*(P_{XY},R)&=R(P_{XY})-R,&&R\le R(P_{XY})\\ E_{\rm rc}(Q_{V|U},R)&=C(Q_{V|U})-R,&&R\le C(Q_{V|U})\\ E^*(Q_{V|U},R)&=R-C(Q_{V|U}),&&R\ge C(Q_{V|U}).\end{aligned}$$

Therefore, in this case

$$\lim_{r\downarrow 0}\frac{E(P_{XY},R(P_{XY})+r)}{r^2}=\lim_{r\downarrow 0}\frac{E^*(P_{XY},R(P_{XY})-r)}{r^2}=\infty$$

$$\lim_{r\downarrow 0}\frac{E(Q_{V|U},C(Q_{V|U})-r)}{r^2}=\lim_{r\downarrow 0}\frac{E^*(Q_{V|U},C(Q_{V|U})+r)}{r^2}=\infty.$$

Note that a lower bound on κ was obtained by Gallager [13] for general discrete memoryless channels. It is also worth mentioning that the second-order expansion of $E(P_{XY},R)$ at the Slepian–Wolf limit yields the redundancy-error tradeoff constant of fixed-rate Slepian–Wolf coding derived in [21].

Proof: See Appendix II. □

III. Expurgated Exponents

In view of Theorem 3, it is natural to conjecture that the mirror symmetry holds for the reliability functions, i.e., $E(P_{XY},R)=E(Q_{V|U},\log M-R)$ for each joint distribution P_{XY} and its dual channel $Q_{V|U}$. We will show that this conjecture does not hold in general. It will be seen that such a symmetry-breaking phenomenon has interesting implications on the performance of linear Slepian–Wolf codes. To the end of disproving this conjecture, we will give a detailed analysis and comparison of the expurgated exponents for channel coding and Slepian–Wolf coding.

In channel coding, the random coding exponent can be improved at low rates by expurgating poor codewords, and the resulting exponent is referred to as the expurgated exponent. For any discrete memoryless channel $Q_{V|U}:\mathcal{U}\to\mathcal{V}$, define

$$E_{\rm ex}(Q,Q_{V|U},R)=\min[\mathbb{E}d_{Q_{V|U}}(U,\tilde{U})+I(U;\tilde{U})-R]$$

where $d_{Q_{V|U}}(u,\tilde{u})=-\log\sum_v\sqrt{Q_{V|U}(v|u)Q_{V|U}(v|\tilde{u})}$ and the minimization is over all joint distributions $Q_{U\tilde{U}}$ with $Q_U=Q_{\tilde{U}}=Q$ and $I(U;\tilde{U})\le R$. The expurgated exponent $E_{\rm ex}(Q_{V|U},R)$ is given by [15]

$$E_{\rm ex}(Q_{V|U},R)=\max_Q E_{\rm ex}(Q,Q_{V|U},R).$$

Alternatively, we can write $E_{\rm ex}(Q_{V|U},R)$ as [13]

$$E_{\rm ex}(Q_{V|U},R)=\sup_{\rho\ge 1}\left[-\rho R+\max_Q E_{\rm ex}(\rho,Q,Q_{V|U})\right] \quad (20)$$

with

$$E_{\rm ex}(\rho,Q,Q_{V|U})=-\rho\log\sum_{u,\tilde{u}}Q(u)Q(\tilde{u})e^{-\frac{d_{Q_{V|U}}(u,\tilde{u})}{\rho}}.$$

It is well known that the expurgated exponent $E_{\rm ex}(Q_{V|U},R)$ is a lower bound on the reliability function $E(Q_{V|U},R)$; moreover, it is asymptotically tight at low rates, i.e., $E_{\rm ex}(Q_{V|U},0)=\lim_{R\downarrow 0}E(Q_{V|U},R)$ [13].

Define $R_{\rm ex,\infty}(Q_{V|U})=\min\{R: E_{\rm ex}(Q_{V|U},R)<\infty\}$. It is shown in [13] that

$$R_{\rm ex,\infty}(Q_{V|U})=\max_Q-\log\left[\sum_{u,\tilde{u}}Q(u)Q(\tilde{u})a(u,\tilde{u})\right]$$

where

$$a(u,\tilde{u})=\begin{cases}1, & \text{if } \sum_v\sqrt{Q_{V|U}(v|u)Q_{V|U}(v|\tilde{u})}\ne 0\\ 0, & \text{otherwise.}\end{cases}$$

Furthermore, let $L(Q_{V|U})$ be the size of the largest set $\mathcal{S} \subseteq \mathcal{U}$ such that $a(u,\tilde{u}) = 0$ for all $u,\tilde{u} \in \mathcal{S}$, $u \neq \tilde{u}$. It is shown in [13] that

$$R_{\mathrm{ex},\infty}(Q_{V|U}) = \log L(Q_{V|U}).$$

The expurgation technique can also be used in Slepian–Wolf coding to improve the random coding exponent at high rates. It will be seen that different types of codes and decoding rules can lead to different expurgated exponents.

A. Csiszár and Körner's Expurgated Exponent

Via a graph decomposition argument, Csiszár and Körner [22] derived the following expurgated exponent for Slepian–Wolf coding:

$$E_{\mathrm{ex},\alpha}^{\mathrm{CK}}(P_{XY},R) = \min_{P_{\hat{X}\tilde{X}\tilde{Y}}} [D(P_{\tilde{X}\tilde{Y}}||P_{XY}) + |R - H(\hat{X}|\tilde{X},\tilde{Y})|^+] \tag{21}$$

where the minimization is over all probability distributions $P_{\hat{X}\tilde{X}\tilde{Y}}$ subject to the constraints $\alpha(P_{\hat{X}\tilde{Y}}) \leq \alpha(P_{\tilde{X}\tilde{Y}}), P_{\hat{X}} = P_{\tilde{X}}$, and $H(\hat{X}|\tilde{X}) \geq R$. Specifically, if MAP decoding is used, then $\alpha(P_{\hat{X}\tilde{Y}}) = D(P_{\hat{X}\tilde{Y}}||P_{XY}) + H(\hat{X},\tilde{Y})$, $\alpha(P_{\tilde{X}\tilde{Y}}) = D(P_{\tilde{X}\tilde{Y}}||P_{XY}) + H(\tilde{X},\tilde{Y})$, and we denote $E_{\mathrm{ex},\alpha}^{\mathrm{CK}}(P_{XY},R)$ by $E_{\mathrm{ex,MAP}}^{\mathrm{CK}}(P_{XY},R)$; if ME decoding is used, then $\alpha(P_{\hat{X}\tilde{Y}}) = H(\tilde{Y}|\hat{X})$, $\alpha(P_{\tilde{X}\tilde{Y}}) = H(\tilde{Y}|\tilde{X})$, and we denote $E_{\mathrm{ex},\alpha}^{\mathrm{CK}}(P_{XY},R)$ by $E_{\mathrm{ex,ME}}^{\mathrm{CK}}(P_{XY},R)$.

Define

$$E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY},R) = \min_{P_{\tilde{X}}} [D(P_{\tilde{X}}||P_X) + E_{\mathrm{ex}}(P_{\tilde{X}},P_{Y|X},H(\tilde{X}) - R)]. \tag{22}$$

It is shown in [22] that

$$E_{\mathrm{ex}}(P_{\tilde{X}},P_{Y|X},R) = \min_{P_{\hat{X}\tilde{Y}|\tilde{X}}} [D(P_{\tilde{Y}|\tilde{X}}||P_{Y|X}|P_{\tilde{X}}) + I(\hat{X},\tilde{Y};\hat{X}) - R] \tag{23}$$

where the minimization is over all $P_{\hat{X}\tilde{Y}|\tilde{X}}$ subject to the constraints $P_{\hat{X}} = P_{\tilde{X}}$, $D(P_{\tilde{Y}|\hat{X}}||P_{Y|X}|P_{\hat{X}}) + H(\tilde{Y}|\hat{X}) \leq D(P_{\tilde{Y}|\tilde{X}}||P_{Y|X}|P_{\hat{X}}) + H(\tilde{Y}|\tilde{X})$, and $H(\hat{X}|\tilde{X}) \geq R$. In view of (21)–(23), it is easy to verify that

$$E_{\mathrm{ex,MAP}}^{\mathrm{CK}}(P_{XY},R) \geq E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY},R).$$

Let P_{X^*} be the uniform distribution on $\mathcal{X}$. We have

$$\begin{aligned} E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY},\log M) &= D(P_{X^*}||P_X) + E_{\mathrm{ex}}(P_{X^*},P_{Y|X},0) \\ &= \sum_x \frac{1}{M} \log\left[\frac{1}{MP_X(x)}\right] \\ &\quad - \frac{1}{M^2}\sum_{x,\tilde{x}} \log\left[\sum_y \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\tilde{x})}\right] \\ &= -\log M - \frac{1}{M^2}\sum_{x,\tilde{x}} \log \sqrt{P_X(x)P_X(\tilde{x})} \\ &\quad - \frac{1}{M^2}\sum_{x,\tilde{x}} \log\left[\sum_y \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\tilde{x})}\right] \\ &= -\log M - \frac{1}{M^2}\sum_{x,\tilde{x}} \log \sum_y \sqrt{P_{XY}(x,y)P_{XY}(\tilde{x},y)}. \end{aligned}$$

Note that $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY},\log M) = \infty$ if and only if there exist x and $\tilde{x}$ such that

$$\sum_y \sqrt{P_{XY}(x,y)P_{XY}(\tilde{x},y)} = 0$$

i.e., $P_{XY}(x,y)P_{XY}(\tilde{x},y) = 0$ for all y. Intuitively, if $P_{XY}(x,y)P_{XY}(\tilde{x},y) = 0$ for all y, then given any y, at most one of x and $\tilde{x}$ is possible; therefore, the encoder can treat x and $\tilde{x}$ as the same symbol, which implies that zero decoding error probability is achievable with rate no greater than $\log(M-1)$.

B. Csiszár's Linear Coding Expurgated Exponent

Csiszár [5] derived the following expurgated exponent for Slepian–Wolf coding by exploiting the properties of linear codes:

$$E_{\mathrm{ex},\alpha}^{C}(P_{XY},R) = \min_{P_{\hat{X}\tilde{X}\tilde{Y}}} [D(P_{\tilde{X}\tilde{Y}}||P_{XY}) + |R - H(\hat{X}|\tilde{X},\tilde{Y})|^+]$$

where M is assumed to be a prime number, and the minimization is over all probability distributions $P_{\hat{X}\tilde{X}\tilde{Y}}$ subject to the constraints $H(\hat{X} -_M \tilde{X}) \geq R$ and $\alpha(P_{\hat{X}\tilde{Y}}) \leq \alpha(P_{\tilde{X}\tilde{Y}})$. We will denote $E_{\mathrm{ex},\alpha}^{C}(P_{XY},R)$ by $E_{\mathrm{ex,MAP}}^{C}(P_{XY},R)$ if MAP decoding is used, and by $E_{\mathrm{ex,ME}}^{C}(P_{XY},R)$ if ME decoding is used. Define

$$E_{\mathrm{ex,map}}^{C}(P_{XY},R) = \min_{P_{\tilde{X}}:H(\tilde{X})\geq R} [-\mathbb{E}\log b(\tilde{X}) + R - H(\tilde{X})]$$

where $b(\tilde{x}) = \sum_{x,y} \sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x},y)}$. It is shown in [5] that

$$E_{\mathrm{ex,map}}^{C}(P_{XY},R) \leq \min_{P_{\hat{X}\tilde{X}\tilde{Y}}} [D(P_{\tilde{X}\tilde{Y}}||P_{XY}) + R - H(\hat{X}|\tilde{X},\tilde{Y})] \tag{24}$$

where the minimization is over all probability distributions $P_{\hat{X}\tilde{X}\tilde{Y}}$ subject to the same constraints as those in the definition of $E_{\mathrm{ex,MAP}}^{C}(P_{XY},R)$. Therefore, we have

$$E_{\mathrm{ex,MAP}}^{C}(P_{XY},R) \geq E_{\mathrm{ex,map}}^{C}(P_{XY},R).$$

The following result can be proved via Lagrange duality.

Lemma 1: $E_{\mathrm{ex,map}}^{C}(P_{XY},R) = \sup_{\rho\geq 1}\{\rho R - \rho \log \sum_{\tilde{x}} b^{\frac{1}{\rho}}(\tilde{x})\}$.

Theorem 6: Assuming M is a prime number, we have

$$E_{\mathrm{ex,map}}^{C}(P_{XY},R) \leq E_{\mathrm{ex}}(Q_{V|U},\log M - R) \tag{25}$$

for each joint probability distribution P_{XY} and its dual channel $Q_{V|U}$, where the inequality in (25) becomes equality if the following condition holds:

C1) for any $\rho \geq 1$, the minimum of $\sum_{x,\tilde{x}} P(x)P(\tilde{x})\, b^{\frac{1}{\rho}}(\tilde{x} -_M x)$ over all probability distributions P on $\mathcal{X}$ is attained when P is the uniform distribution.

Remark: A sufficient condition for C1) to hold is that the dual channel $Q_{V|U}$ is equidistant.[3] It is worth noting that Theorem 6 can be viewed as a generalization of (A.2) in [5].

Proof: In view of (20), we have (26) shown at the bottom of the page, where the inequality becomes equality if C1) is satisfied. □

Note that $E^C_{\mathrm{ex,map}}(P_{XY}, R)$ is achievable by linear codes. Therefore, Theorem 6 can be viewed as a direct consequence of the linear codebook-level duality between Slepian–Wolf coding and channel coding. Furthermore, since $E_{\mathrm{ex}}(Q_{V|U}, 0) = \lim_{R\downarrow 0} E(Q_{V|U}, R)$, the following statements hold if C1) is satisfied:

1) linear codes can asymptotically achieve the reliability function of the dual channel $Q_{V|U}$ for rates approaching zero;
2) $E^C_{\mathrm{ex,map}}(P_{XY}, R)$ is asymptotically tight for linear Slepian–Wolf codes as R goes to $\log M$.

It is easy to verify that

$$\begin{aligned} &E^C_{\mathrm{ex,map}}(P_{XY}, \log M) \\ &\quad = -\frac{1}{M}\sum_{\tilde{x}} \log b(\tilde{x}) \\ &\quad = -\frac{1}{M}\sum_{\tilde{x}} \log \sum_{x,y} \sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x}, y)}. \end{aligned}$$

Therefore, $E^C_{\mathrm{ex,map}}(P_{XY}, \log M) = \infty$ if and only if $b(\tilde{x}) = 0$ for some $\tilde{x}$. Define

$$R^C_{\mathrm{ex},\infty}(P_{XY}) = \max\{R : E^C_{\mathrm{ex,map}}(P_{XY}, R) < \infty\}.$$

[3]A discrete memoryless channel $Q_{V|U}$ is called equidistant if there exists a number $\beta \geq 0$ such that $\sum_v \sqrt{Q_{V|U}(v|u_1)Q_{V|U}(v|u_2)} = \beta$ for all pairs of inputs $u_1 \neq u_2$ [23]. In particular, all binary input channels are equidistant. In the current setting, the dual channel $Q_{V|U}$ is equidistant if there exists a number $\beta \geq 0$ such that $b(\bar{x}) = \beta$ for all $\bar{x} \neq 0$.

It can be shown that $R^C_{\mathrm{ex},\infty}(P_{XY}) = \log L(P_{XY})$, where $L(P_{XY}) = \sum_{\tilde{x}} \varphi(\tilde{x})$, and

$$\varphi(\tilde{x}) = \begin{cases} 1, & \text{if } b(\tilde{x}) \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

For the dual channel $Q_{V|U}$, we have $E_{\mathrm{ex}}(Q_{V|U}, 0) = \infty$ if and only if there exist u and $\tilde{u}$ such that

$$\sum_v \sqrt{Q_{V|U}(v|u)Q_{V|U}(v|\tilde{u})} = 0.$$

Since $\sum_v \sqrt{Q_{V|U}(v|u)Q_{V|U}(v|\tilde{u})} = b(\tilde{u} -_M u)$, it is easy to see that $E^C_{\mathrm{ex,map}}(P_{XY}, \log M) = \infty$ implies $E_{\mathrm{ex}}(Q_{V|U}, 0) = \infty$, and *vice versa*. In view of Theorem 6, we have $R_{\mathrm{ex},\infty}(Q_{V|U}) \geq \log M - R^C_{\mathrm{ex},\infty}(P_{XY})$. Note that the inequality can be strict. Indeed, if M is not a multiple of $L(P_{XY})$ or $L(Q_{V|U})$, then we must have $R_{\mathrm{ex},\infty}(Q_{V|U}) + R^C_{\mathrm{ex},\infty}(P_{XY}) > \log M$. The inequality can be strict even if M is a multiple of $L(P_{XY})$ and $L(Q_{V|U})$. Consider the joint probability distribution $P_{XY} : \mathbb{Z}_6 \times \mathbb{Z}_3$ with $P_{Y|X}(0|0) = P_{Y|X}(1|0) = P_{Y|X}(2|1) = P_{Y|X}(3|1) = P_{Y|X}(4|2) = P_{Y|X}(5|2) = 1$, and $P_X(x) = \frac{1}{6}$ for all $x \in \mathbb{Z}_6$. It is easy to verify that for this distribution P_{XY} and its dual channel $Q_{V|U}$, we have $L(P_{XY}) = L(Q_{V|U}) = 3$ and $R_{\mathrm{ex},\infty}(Q_{V|U}) + R^C_{\mathrm{ex},\infty}(P_{XY}) = \log 9 > \log 6$. Therefore, C1) is not a redundant condition.

C. Oohama and Han's Expurgated Exponent

Based on an interesting observation on the cardinality of the set of sequences of the same marginal type, Oohama and Han [12] derived the following expurgated exponent for Slepian–Wolf coding by using ME decoding:

$$\begin{aligned} &E^{\mathrm{OH}}_{\mathrm{ex,ME}}(P_{XY}, R) \\ &\quad = \min_{P_{\tilde{X}\tilde{Y}} : H(\tilde{X}) \geq R} [D(P_{\tilde{X}\tilde{Y}} \| P_{XY}) + |R - H(\tilde{X}|\tilde{Y})|^+]. \end{aligned} \quad (27)$$

Comparing (27) with (1), it is clear that $E^{\mathrm{OH}}_{\mathrm{ex,ME}}(P_{XY}, R) \geq E_{\mathrm{rc}}(P_{XY}, R)$. Furthermore, one can readily see that the con-

$$\begin{aligned} E_{\mathrm{ex}}(Q_{V|U}, R) &= \sup_{\rho \geq 1} \left\{ -\rho R - \min_Q \rho \log \left[\sum_{u,\tilde{u}} Q(u)Q(\tilde{u}) \left(\sum_v Q_{V|U}(v|u)Q_{V|U}(v|\tilde{u}) \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \geq 1} \left\{ -\rho R - \min_Q \rho \log \left[\sum_{u,\tilde{u}} Q(u)Q(\tilde{u}) \left(\sum_{x,y} P_{XY}(x -_M u, y)P_{XY}(x -_M \tilde{u}, y) \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \geq 1} \left\{ -\rho R - \min_Q \rho \log \left[\sum_{u,\tilde{u}} Q(u)Q(\tilde{u}) b^{\frac{1}{\rho}}(\tilde{u} -_M u) \right] \right\} \\ &\geq \sup_{\rho \geq 1} \left\{ -\rho R - \rho \log \left[\frac{1}{M^2} \sum_{u,\tilde{u}} b^{\frac{1}{\rho}}(\tilde{u} -_M u) \right] \right\} \\ &= \sup_{\rho \geq 1} \left\{ -\rho R - \rho \log \left[\frac{1}{M} \sum_{\tilde{x}} b^{\frac{1}{\rho}}(\tilde{x}) \right] \right\} \\ &= E^C_{\mathrm{ex,map}}(P_{XY}, \log M - R) \end{aligned} \quad (26)$$

straint $H(\tilde{X}) \geq R$ in (27) is not active if $R \leq H(X^{(\rho)})|_{\rho=1}$ [cf., (2)]. Therefore, we have

$$E_{\mathrm{ex,ME}}^{\mathrm{OH}}(P_{XY}, R) = E_{\mathrm{rc}}(P_{XY}, R), \qquad R \leq H(X^{(\rho)})|_{\rho=1}$$
$$E_{\mathrm{ex,ME}}^{\mathrm{OH}}(P_{XY}, R) > E_{\mathrm{rc}}(P_{XY}, R), \qquad R > H(X^{(\rho)})|_{\rho=1}.$$

D. Comparison

It is easy to verify that

$$\begin{aligned} &E_{\mathrm{ex,MAP}}^{\mathrm{CK}}(P_{XY}, R) \\ &\quad \geq E_{\mathrm{ex,MAP}}^{C}(P_{XY}, R) \\ &E_{\mathrm{ex,ME}}^{\mathrm{CK}}(P_{XY}, R) \\ &\quad \geq \max\{E_{\mathrm{ex,ME}}^{C}(P_{XY}, R), E_{\mathrm{ex,ME}}^{\mathrm{OH}}(P_{XY}, R)\}. \end{aligned}$$

Therefore, Csiszár and Körner's expurgated exponent is tightest. However, it is worth noting that Oohama and Han's expurgation technique and Csiszár's linear coding argument can be directly applied to the setting where the side information is also encoded; in contrast, it is not clear how to generalize the graph decomposition method in [22] to the aforementioned general setting.

Proposition 1: Assuming that M is a prime number and $E_{\mathrm{ex,map}}^{C}(P_{XY}, \log M) < \infty$ (i.e., $b(\tilde{x}) \neq 0$ for all $\tilde{x} \in \mathcal{X}$), we have

$$E_{\mathrm{ex,map}}^{C}(P_{XY}, \log M) < E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, \log M)$$

unless the following condition is satisfied:

C2) for any given $\tilde{x}$, the value of $\sum_y \sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x}, y)}$ does not depend on x.

Remark: By setting $\tilde{x} = 0$, we can see that C2) implies $P_X(x) = \frac{1}{M}$ for all $x \in \mathcal{X}$.

Proof: Note that

$$\begin{aligned} &E_{\mathrm{ex,map}}^{C}(P_{XY}, \log M) \\ &\quad = -\frac{1}{M}\sum_{\tilde{x}} \log\left[\sum_{x,y} \sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x}, y)}\right] \\ &\quad = -\log M \\ &\qquad - \frac{1}{M}\sum_{\tilde{x}} \log\left[\sum_{x,y} \frac{1}{M}\sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x}, y)}\right] \\ &\quad \leq -\log M \\ &\qquad - \frac{1}{M^2}\sum_{x,\tilde{x}} \log\left[\sum_{y} \sqrt{P_{XY}(x,y)P_{XY}(x -_M \tilde{x}, y)}\right] \\ &\quad = E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, \log M) \end{aligned} \tag{28}$$

where the inequality in (28) is strict unless C2) is satisfied. □

Proposition 2: Assuming that M is a prime number, we have $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, R) = E_{\mathrm{ex,map}}^{C}(P_{XY}, R)$ under the conditions C2) and:

C3) for any $\rho \geq 1$, the maximum of $E_{\mathrm{ex}}(\rho, P, P_{Y|X})$ as a function of P is attained when P is the uniform distribution on $\mathcal{X}$.

Remark: A sufficient condition for C3) to hold is that $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is an equidistant channel. Moreover, conditions C2) and C3) hold simultaneously if P_X is the uniform distribution on $\mathcal{X}$ and $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is an equidistant channel [e.g., $M = 2$ and $P_X(0) = P_X(1)$].

Proof: It suffices to show that $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, R) \leq E_{\mathrm{ex,map}}^{C}(P_{XY}, R)$ under conditions C2) and C3). Since C2) implies $P_X(x) = \frac{1}{M}$ for all $x \in \mathcal{X}$, it follows that

$$\begin{aligned} &E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, R) \\ &\quad = \min_{P_{\tilde{X}}} \left[D(P_{\tilde{X}}||P_X) + E_{\mathrm{ex}}(P_{\tilde{X}}, P_{Y|X}, H(\tilde{X}) - R)\right] \\ &\quad \leq D(P_X||P_X) + E_{\mathrm{ex}}(P_X, P_{Y|X}, H(X) - R) \\ &\quad = E_{\mathrm{ex}}(P_X, P_{Y|X}, \log M - R). \end{aligned}$$

In view of C3), we have

$$\begin{aligned} &E_{\mathrm{ex}}(P_X, P_{Y|X}, \log M - R) \\ &\quad \leq E_{\mathrm{ex}}(P_{Y|X}, \log M - R) \\ &\quad = \sup_{\rho \geq 1}\left[-\rho(\log M - R) + \max_{P} E_{\mathrm{ex}}(\rho, P, P_{Y|X})\right] \\ &\quad = \sup_{\rho \geq 1}\left\{-\rho(\log M - R)\right. \\ &\qquad \left. -\rho \log\left[\frac{1}{M^2}\sum_{x,\tilde{x}}\left(\sum_{y}\sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\tilde{x})}\right)^{\frac{1}{\rho}}\right]\right\}. \end{aligned}$$

Note that under condition C2), we have the equation shown at the bottom of the following page. The proof is complete. □

Since $E_{\mathrm{ex}}(Q_{V|U}, 0) = \lim_{R \downarrow 0} E(Q_{V|U}, R)$, to disprove the conjecture raised at the beginning of this section, it suffices to find a joint distribution P_{XY} and its dual channel $Q_{V|U}$ such that $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, \log M) > E_{\mathrm{ex}}(Q_{V|U}, 0)$. In view of Theorem 6 and Proposition 1, one can easily obtain sufficient conditions for $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, \log M) > E_{\mathrm{ex}}(Q_{V|U}, 0)$ to hold. The above analysis also reveals a subtle difference between the roles of linear codes in Slepian–Wolf coding and channel coding. Specifically, if $E_{\mathrm{ex,map}}^{\mathrm{CK}}(P_{XY}, \log M) > E_{\mathrm{ex,map}}^{C}(P_{XY}, \log M) = E_{\mathrm{ex}}(Q_{V|U}, 0)$, then linear Slepian–Wolf codes are suboptimal at high rates in terms of rate-error tradeoff; in contrast, linear codes are asymptotically optimal at low rates in the dual channel coding problem. An illustrative example is given in the next section.

IV. Example

Consider a joint probability distribution P_{XY} on $\mathbb{Z}_2 \times \mathbb{Z}_2$ with $P_{X|Y}(1|0) = P_{X|Y}(0|1) = p$ and $P_Y(0) = \tau$. We assume $p \in (0, \frac{1}{2})$, $\tau \in (0, \frac{1}{2}]$. It is easy to see that for this distribution P_{XY} and its dual channel $Q_{V|U}$, we have $R(P_{XY}) = H_b(p)$ and $C(Q_{V|U}) = \log 2 - H_b(p)$, where $H_b(\cdot)$ is the binary entropy function. Moreover, it can be verified that conditions C1) and C3) are satisfied while condition C2) is satisfied if and only if $\tau = \frac{1}{2}$.

Given $R \in [0, \log 2]$, let q be the unique number satisfying $H_b(q) = R$ and $q \le \frac{1}{2}$. For $R \in [H_b(p), \log 2]$, we have

$$E_{\rm sp}(P_{XY}, R) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$$

$$E_{\rm rc}(P_{XY}, R) = \begin{cases} q \log \dfrac{q}{p} + (1-q) \log \dfrac{1-q}{1-p}, & R \le R_{\rm cr}(P_{XY}) \\ R - 2\log(\sqrt{p} + \sqrt{1-p}), & \text{otherwise} \end{cases}$$

$$E^C_{\rm ex,map}(P_{XY}, R) = \begin{cases} \dfrac{q}{2} \log \dfrac{1}{4p(1-p)}, & q \ge \dfrac{\sqrt{4p(1-p)}}{1+\sqrt{4p(1-p)}} \\ R - 2\log(\sqrt{p} + \sqrt{1-p}), & \text{otherwise} \end{cases}$$

where $R_{\rm cr}(P_{XY}) = H_b\left(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}\right)$. For $R \in [0, H_b(p)]$, we have

$$E^*(P_{XY}, R) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}.$$

It follows from Theorems 3 and 6 that

$$\begin{aligned} E_{\rm sp}(Q_{V|U}, R) &= E_{\rm sp}(P_{XY}, \log 2 - R) \\ E_{\rm rc}(Q_{V|U}, R) &= E_{\rm rc}(P_{XY}, \log 2 - R) \\ E_{\rm ex}(Q_{V|U}, R) &= E^C_{\rm ex,map}(P_{XY}, \log 2 - R) \\ E^*(Q_{V|U}, R) &= E^*(P_{XY}, \log 2 - R). \end{aligned}$$

Note that none of these exponents depends on τ, which is barely surprising since the dual channel $Q_{V|U}$ is equivalent to the binary symmetric channel with crossover probability p.

For $R \in [H_b(p), \log 2]$, we have

$$E^{\rm OH}_{\rm ex,ME}(P_{XY}, R) = \min_{P_{\tilde{X}\tilde{Y}}: H(\tilde{X}) \ge R} [D(P_{\tilde{X}\tilde{Y}} || P_{XY}) + |R - H(\tilde{X}|\tilde{Y})|^+].$$

In particular

$$E^{\rm OH}_{\rm ex,ME}(P_{XY}, R) = E_{\rm rc}(P_{XY}, R)$$

for

$$R \le H_b\left(\frac{\tau\sqrt{1-p} + (1-\tau)\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}\right).$$

It can be verified that $E^{\rm OH}_{\rm ex,ME}(P_{XY}, R)$ approaches $E_{\rm sp}(P_{XY}, R)$ as $\tau \to 0$, and approaches $E_{\rm rc}(P_{XY}, R)$ as $\tau \to \frac{1}{2}$.

For $R \in [H_b(p), \log 2]$, we have

$$E^{\rm CK}_{\rm ex,map}(P_{XY}, R) = \min \left[-\mathbb{E} \log \sum_y \sqrt{P_{XY}(\hat{X}, y) P_{XY}(\tilde{X}, y)} - H(\hat{X}, \tilde{X}) + R \right]$$

where the minimization is over all joint distributions $P_{\hat{X}\tilde{X}}$ with $P_{\hat{X}} = P_{\tilde{X}}$ and $H(\hat{X}|\tilde{X}) \ge R$. Specifically, $E^{\rm CK}_{\rm ex,map}(P_{XY}, R) = R - 2\log(\sqrt{p} + \sqrt{1-p})$ for

$$R \le \frac{P_X(0) + \sqrt{p(1-p)}}{1 + 2\sqrt{p(1-p)}} H_b\left(\frac{P_X(0)}{P_X(0) + \sqrt{p(1-p)}}\right) + \frac{P_X(1) + \sqrt{p(1-p)}}{1 + 2\sqrt{p(1-p)}} H_b\left(\frac{P_X(1)}{P_X(1) + \sqrt{p(1-p)}}\right).$$

It can be verified that $E^{\rm CK}_{\rm ex,map}(P_{XY}, R)$ converges to $E_{\rm sp}(P_{XY}, R)$ [for $R \ge R_{\rm cr}(P_{XY})$] as $\tau \to 0$, and converges to $E^C_{\rm ex,map}(P_{XY}, R)$ as $\tau \to \frac{1}{2}$.

For the dual channel $Q_{V|U}$, the straight-line exponent $E_{\rm sl}(Q_{V|U}, R)$ is the smallest linear function of R which touches the curve $E_{\rm sp}(Q_{V|U}, R)$ and also satisfies $E_{\rm sl}(Q_{V|U}, 0) = E_{\rm ex}(Q_{V|U}, 0) = -\frac{1}{4}\log[4p(1-p)]$. Let R_1 be the R for which $E_{\rm sl}(Q_{V|U}, R) = E_{\rm sp}(Q_{V|U}, R)$. It is known [13] that $E(Q_{V|U}, R)$ is upper bounded by $E_{\rm sl}(Q_{V|U}, R)$ for $R \in (0, R_1]$. By the linear codebook-level duality, the dual straight-line exponent (see the dashed lines in Figs. 1 and 2) applies to linear Slepian–Wolf codes. However, the dual straight-line exponent does not hold for general Slepian–Wolf

---

$$\begin{aligned} &\sup_{\rho \ge 1} \left\{ -\rho(\log M - R) - \rho \log \left[\frac{1}{M^2} \sum_{x,\tilde{x}} \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \ge 1} \left\{ -\rho(\log M - R) - \rho \log \left[\frac{1}{M^2} \sum_{x,\tilde{x}} \left(M \sum_y \sqrt{P_{XY}(x,y) P_{XY}(\tilde{x},y)} \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \ge 1} \left\{ -\rho(\log M - R) - \rho \log \left[\frac{1}{M^2} \sum_{x,\tilde{x}} \left(M \sum_y \sqrt{P_{XY}(x,y) P_{XY}(x -_M \tilde{x},y)} \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \ge 1} \left\{ -\rho(\log M - R) - \rho \log \left[\frac{1}{M} \sum_{\tilde{x}} \left(\sum_{x,y} \sqrt{P_{XY}(x,y) P_{XY}(x -_M \tilde{x},y)} \right)^{\frac{1}{\rho}} \right] \right\} \\ &= \sup_{\rho \ge 1} \left\{ \rho R - \rho \log \left[\sum_{\tilde{x}} b^{\frac{1}{\rho}}(\tilde{x}) \right] \right\} \\ &= E^C_{\rm ex,map}(P_{XY}, R). \end{aligned}$$

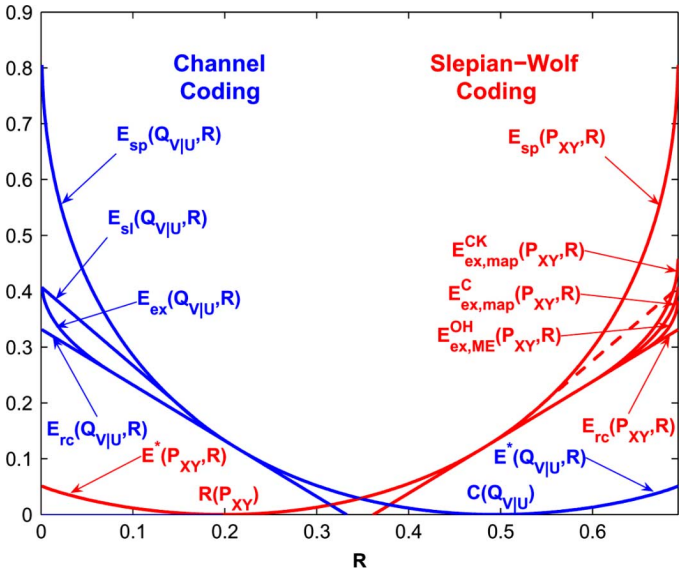

Fig. 2. Mirror symmetry and symmetry breaking: $p = 0.05$, $\tau = 0.28$.

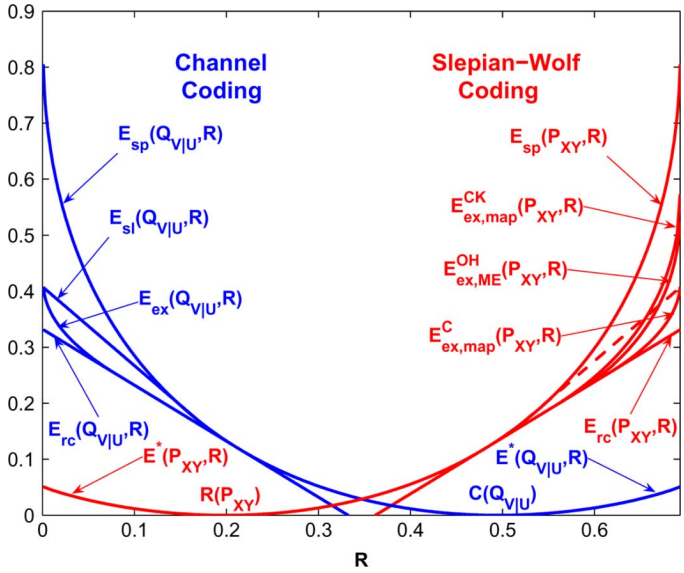

Fig. 3. Mirror symmetry and symmetry breaking: $p = 0.05$, $\tau = 0.12$.

codes since it can be dominated by $E^{\mathrm{CK}}_{\mathrm{ex,map}}(P_{XY}, R)$ at high rates.

It is thus clear that in order to operate above the dual straight-line exponent, one has to use nonlinear Slepian–Wolf codes. Moreover, we can see from Fig. 2 that it is possible for $E^{\mathrm{OH}}_{\mathrm{ex,ME}}(P_{XY}, R)$ to be strictly above the dual straight-line exponent in the high rate regime. This implies that at high rates, nonlinear Slepian–Wolf codes under suboptimal decoding can outperform the best linear Slepian–Wolf codes under MAP decoding.

V. General Sources and Channels

Now we proceed to study Slepian–Wolf coding for general sources and its duality with channel coding. We first quote some definitions of entropy and mutual information of general sources from [24], [25]. The *limsup in probability* of a sequence of random variables $\{Z_t\}_{t=1}^{\infty}$ is defined as the smallest extended real number α such that for all $\epsilon > 0$

$$\lim_{t\to\infty} \Pr\{Z_t \geq \alpha + \epsilon\} = 0.$$

Analogously, the *liminf in probability* is the largest extended real number β such that for all $\epsilon > 0$

$$\lim_{t\to\infty} \Pr\{Z_t \leq \beta - \epsilon\} = 0.$$

A general source $\mathbf{Z}$ with alphabet $\mathcal{Z}$ is a sequence of random variables $\{Z^n\}_{n=1}^{\infty}$ with $Z^n \in \mathcal{Z}^n$. For general sources $\mathbf{X}, \mathbf{Y}$, and $\mathbf{Z}$, $\overline{H}(\mathbf{X})$ and $\overline{H}(\mathbf{X}|\mathbf{Y})$ are defined as the *limsup in probability* of $\{-\frac{1}{n}\log P_{X^n}(X^n)\}_{n=1}^{\infty}$ and $\{-\frac{1}{n}\log P_{X^n|Y^n}(X^n|Y^n)\}_{n=1}^{\infty}$, respectively, while $\underline{I}(\mathbf{X};\mathbf{Y})$ and $\underline{I}(\mathbf{X};\mathbf{Z}|\mathbf{Y})$ are defined as the *liminf in probability* of

$$\left\{\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)}\right\}_{n=1}^{\infty}$$

and

$$\left\{\frac{1}{n}\log\frac{P_{Z^n|X^nY^n}(Z^n|X^n,Y^n)}{P_{Z^n|Y^n}(Z^n|Y^n)}\right\}_{n=1}^{\infty}$$

respectively.

We will denote the Slepian–Wolf limit for general source $(\mathbf{X},\mathbf{Y})$ by $R(P_{\mathbf{XY}})$, where $P_{\mathbf{XY}} = \{P_{X^nY^n}\}_{n=1}^{\infty}$. It is shown in [3] that $R(P_{\mathbf{XY}}) = \overline{H}(\mathbf{X}|\mathbf{Y})$. Note that the minimum achievable rate is still $\overline{H}(\mathbf{X}|\mathbf{Y})$ even if the side information $\mathbf{Y}$ is available at both the encoder and the decoder. Therefore, similarly to the stationary and ergodic case, the lack of encoder side information does not incur any penalty in terms of the minimum achievable rate.

We will show that Slepian–Wolf coding for general sources can be better understood through the linear codebook-level duality. Given general source $(\mathbf{X},\mathbf{Y})$, we construct a dual channel $Q_{\mathbf{V}|\mathbf{U}} = \{Q_{V^n|U^n} : \mathcal{U}^n \to \mathcal{V}^n\}_{n=1}^{\infty}$ with $\mathcal{U}^n = \mathcal{X}^n$, $\mathcal{V}^n = \mathcal{X}^n \times \mathcal{Y}^n$, and $V^n = (U^n +_M X^n, Y^n)$. The input U^n is assumed to be independent of X^n and Y^n. The capacity of the dual channel $Q_{\mathbf{V}|\mathbf{U}}$ is denoted as $C(Q_{\mathbf{V}|\mathbf{U}})$.

It is clear that if $(\mathbf{X},\mathbf{Y})$ is not stationary, then the dual channel $Q_{\mathbf{V}|\mathbf{U}}$ is also not stationary. In general, it is a formidable task to find the capacity-achieving input process for a nonstationary channel. However, due to the cyclic symmetry of the dual channel, the capacity-achieving input process can be easily characterized.

Theorem 7: For the dual channel $Q_{\mathbf{V}|\mathbf{U}}$, we have

$$C(Q_{\mathbf{V}|\mathbf{U}}) = \log M - R(P_{\mathbf{XY}}).$$

Moreover, the capacity is achieved by the stationary and memoryless process with the uniform marginal distribution over $\mathcal{X}$.

Remark: Theorem 2 is a special case of Theorem 7.

Proof: The proof is straightforward and thus omitted. □

It is easy to verify that the linear codebook-level duality [cf., (18) and (19)] continues to hold in the current general setting. To complete the connection, we need to show that linear codes can achieve the Slepian–Wolf limit for general sources.

Theorem 8: The Slepian–Wolf limit $R(P_{\mathbf{XY}})$ is achievable with linear codes over $\mathbb{Z}_M$ when M is a prime number.

Remark: Theorem 8 can be viewed as a partial generalization of Theorem 4.

Proof: This result can be proved using the standard techniques. The details are omitted. □

VI. Conclusion

A linear codebook-level duality between Slepian–Wolf coding and channel coding is established. This duality provides a conceptual tool for studying linear Slepian–Wolf codes. In particular, it reveals that linear Slepian–Wolf codes are in general suboptimal in terms of rate-error tradeoff at high rates. A generalization of this duality to the mismatched decoding scenario can be found in [26].

Appendix I
Proof of Theorem 4

First, we will quote a few basic definitions related to the method of types [15]. For any finite set $\mathcal{Z}$, let $\mathcal{P}(\mathcal{Z})$ denote the set of all probability distributions on $\mathcal{Z}$. The type of a sequence $z^n \in \mathcal{Z}^n$, denoted as P_{z^n}, is the empirical probability distribution of z^n. Define $\mathcal{P}_n(\mathcal{Z}) = \{P_{z^n} : z^n \in \mathcal{Z}^n\}$ and $\mathcal{T}_n(P) = \{z^n \in \mathcal{Z}^n : P_{z^n} = P\}$ for any $P \in \mathcal{P}_n(\mathcal{Z})$. A few elementary results are listed below:

$$|\mathcal{P}_n(\mathcal{Z})| \le (n+1)^{|\mathcal{Z}|}$$

$$\frac{1}{(n+1)^{|\mathcal{Z}|}} e^{nH(P)} \le |\mathcal{T}_n(P)| \le e^{nH(P)}, \qquad P \in \mathcal{P}_n(\mathcal{Z})$$

$$\prod_{i=1}^{n} P(z_i) = e^{-n[D(Q\|P)+H(Q)]}, \qquad z^n \in \mathcal{T}_n(Q),$$

$$Q \in \mathcal{P}_n(\mathcal{Z}), \qquad P \in \mathcal{P}(\mathcal{Z}).$$

The following fundamental lemma regarding linear Slepian–Wolf codes can be found in [5].

Lemma 2: Let $\delta_n = \frac{\log(n+1)}{n}|\mathcal{X}|^2|\mathcal{Y}|$. If M is a prime number, then for arbitrary positive integers n and k, there exists a linear Slepian–Wolf encoding function $f_n : \mathbb{Z}_M^n \to \mathbb{Z}_M^k$ such that for every joint type $P_{\hat{X}\tilde{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{Y})$

$$\frac{1}{|\mathcal{T}_n(P_{\hat{X}\hat{Y}})|} N_{f_n}(P_{\hat{X}\tilde{X}\hat{Y}}) \le 2^{-n\left|R(f_n)-H(\tilde{X}|\hat{X},\hat{Y})-\delta_n\right|^+}$$

if $\hat{X} \ne \tilde{X}$, where $N_{f_n}(P_{\hat{X}\tilde{X}\hat{Y}})$ denotes for each joint type $P_{\hat{X}\tilde{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{Y})$ the number of pairs $(\hat{x}^n, \hat{y}^n) \in \mathcal{T}_n(P_{\hat{X}\hat{Y}})$ such that for some $\tilde{x}^n \ne \hat{x}^n$ with $P_{\hat{x}^n\tilde{x}^n\hat{y}^n} = P_{\hat{X}\tilde{X}\hat{Y}}$ the relation $f_n(\hat{x}^n) = f_n(\tilde{x}^n)$ holds.

Now we are ready to prove Theorem 4. It is shown in [5] that the random coding exponent $E_{\mathrm{rc}}(P_{XY}, R)$ is universally attainable by linear codes under ME decoding. Therefore, it suffices to prove an analogous result for the correct decoding exponent $E^*(P_{XY}, R)$.

For any $R > 0$, let $\{f_n(\cdot)\}$ be a sequence of linear Slepian–Wolf encoders as specified in Lemma 2 with $R(f_n)$ converging to R as $n \to \infty$, and let $\{g_n(\cdot)\}$ be a sequence of ME decoders. Define

$$\Theta_n(P_{\hat{X}\hat{Y}}) = \{P_{\tilde{X}|\hat{X}\hat{Y}} : P_{\hat{X}\tilde{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}), \hat{X} \ne \tilde{X}, H(\tilde{X},\hat{Y}) \le H(\hat{X},\hat{Y})\}, \qquad P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}).$$

Since $H(\tilde{X},\hat{Y}) \le H(\hat{X},\hat{Y})$ implies $H(\tilde{X}|\hat{Y}) \le H(\hat{X}|\hat{Y})$, we have

$$\begin{aligned}
&\sum_{P_{\tilde{X}|\hat{X}\hat{Y}} \in \Theta_n(P_{\hat{X}\hat{Y}})} \frac{1}{|\mathcal{T}_n(P_{\hat{X}\hat{Y}})|} N_{f_n}(P_{\hat{X}\tilde{X}\hat{Y}}) \\
&\quad \le \sum_{P_{\tilde{X}|\hat{X}\hat{Y}} \in \Theta_n(P_{\hat{X}\hat{Y}})} 2^{-n\left|R(f_n)-H(\tilde{X}|\hat{X},\hat{Y})-\delta_n\right|^+} \\
&\quad \le \sum_{P_{\tilde{X}|\hat{X}\hat{Y}} \in \Theta_n(P_{\hat{X}\hat{Y}})} 2^{-n\left|R(f_n)-H(\tilde{X}|\hat{Y})-\delta_n\right|^+} \\
&\quad \le \sum_{P_{\tilde{X}|\hat{X}\hat{Y}} \in \Theta_n(P_{\hat{X}\hat{Y}})} 2^{-n\left|R(f_n)-H(\hat{X}|\hat{Y})-\delta_n\right|^+} \\
&\quad \le (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} 2^{-n\left|R(f_n)-H(\hat{X}|\hat{Y})-\delta_n\right|^+}.
\end{aligned}$$

Let $\mathcal{E}_n(\epsilon) = \{P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}) : H(\hat{X}|\hat{Y}) \le R - \epsilon\}$ with $\epsilon > 0$. Note the equation at the bottom of the page. Since $R(f_n) - H(\hat{X}|\hat{Y}) - \delta_n$ can be uniformly bounded away from zero for $P_{\hat{X}\hat{Y}} \in \mathcal{E}_n(\epsilon)$ when n is sufficiently large, it follows that

$$\limsup_{n\to\infty} -\frac{1}{n}\log[1 - P_e(f_n, g_n, P_{XY})] \le \limsup_{n\to\infty} \min_{P_{\hat{X}\hat{Y}} \in \mathcal{E}_n(\epsilon)} D(P_{\hat{X}\hat{Y}}\|P_{XY}).$$

In view of the fact that $\epsilon > 0$ is arbitrary, one can readily complete the proof by a simple continuity argument.

Appendix II
Proof of Theorem 5

Since $R_{\mathrm{cr}}(P_{XY}) > R(P_{XY})$, it follows that

$$\lim_{r\downarrow 0} \frac{E(P_{XY}, R(P_{XY})+r)}{r^2} = \lim_{r\downarrow 0} \frac{E_{\mathrm{sp}}(P_{XY}, R(P_{XY})+r)}{r^2}.$$

$$\begin{aligned}
1 - P_e(f_n, g_n, P_{XY}) &\ge \max_{P_{\hat{X}\hat{Y}} \in \mathcal{E}_n(\epsilon)} \left[1 - \sum_{P_{\tilde{X}|\hat{X}\hat{Y}} \in \Theta_n(P_{\hat{X}\hat{Y}})} \frac{1}{|\mathcal{T}_n(P_{\hat{X}\hat{Y}})|} N_{f_n}(P_{\hat{X}\tilde{X}\hat{Y}})\right] \Pr\{X^n \in \mathcal{T}_n(P_{\hat{X}\hat{Y}})\} \\
&\ge \max_{P_{\hat{X}\hat{Y}} \in \mathcal{E}_n(\epsilon)} \left[1 - (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} 2^{-n\left|R(f_n)-H(\hat{X}|\hat{Y})-\delta_n\right|^+}\right] \Pr\{X^n \in \mathcal{T}_n(P_{\hat{X}\hat{Y}})\} \\
&\ge \max_{P_{\hat{X}\hat{Y}} \in \mathcal{E}_n(\epsilon)} \left[1 - (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} 2^{-n\left|R(f_n)-H(\hat{X}|\hat{Y})-\delta_n\right|^+}\right] \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} e^{-nD(P_{\hat{X}\hat{Y}}\|P_{XY})}.
\end{aligned}$$

In view of Theorem 3, it suffices to show

$$\lim_{r\downarrow 0}\frac{E_{\rm sp}(P_{XY},R(P_{XY})+r)}{r^2}=\lim_{r\downarrow 0}\frac{E^*(P_{XY},R(P_{XY})-r)}{r^2}=\kappa$$

which boils down to verifying

$$\lim_{r\to 0}\min_{P_{\tilde{X}\tilde{Y}}:H(\tilde{X}|\tilde{Y})=H(X|Y)+r}\frac{D(P_{\tilde{X}\tilde{Y}}\|P_{XY})}{r^2}=\kappa.$$

Let $\Delta_y=P_{\tilde{Y}}(y)-P_Y(y)$ and $\Delta_{xy}=P_{\tilde{X}|\tilde{Y}}(x|y)-P_{X|Y}(x,y)$ for $x\in\mathcal{X}$, $y\in\mathcal{Y}$. The Taylor expansion yields

$$\begin{aligned}
&H(\tilde{X}|\tilde{Y})\\
&\quad=\sum_y P_{\tilde{Y}}(y)H(\tilde{X}|\tilde{Y}=y)\\
&\quad=\sum_y(P_Y(y)+\Delta_y)H(X|Y=y)-\sum_{x,y}(P_Y(y)+\Delta_y)\\
&\qquad\times[(\log P_{X|Y}(x|y)+1)\Delta_{xy}+o(\Delta_{xy})]\\
&D(P_{\tilde{X}\tilde{Y}}\|P_{XY})\\
&\quad=\sum_{x,y}\left\{\frac{1}{2}P_{XY}^{-1}(x,y)(P_{\tilde{X}\tilde{Y}}(x,y)-P_{XY}(x,y))^2\right.\\
&\qquad\left.+o[(P_{\tilde{X}\tilde{Y}}(x,y)-P_{XY}(x,y))^2]\right\}\\
&\quad=\sum_{x,y}\left\{\frac{1}{2}P_{XY}^{-1}(x,y)(\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y)+\Delta_y\Delta_{xy})^2\right.\\
&\qquad\left.+o[(P_{\tilde{X}\tilde{Y}}(x,y)-P_{XY}(x,y))^2]\right\}
\end{aligned}$$

where $f(z)=o(z)$ means $\lim_{z\to 0}\frac{f(z)}{z}=0$. Therefore, by ignoring the high-order terms which do not affect the limit, we get

$$\begin{aligned}
&\lim_{r\to 0}\min_{P_{\tilde{X}\tilde{Y}}:H(\tilde{X}|\tilde{Y})=H(X|Y)+r}\frac{D(P_{\tilde{X}\tilde{Y}}\|P_{XY})}{r^2}\\
&\qquad=\lim_{r\to 0}\frac{1}{r^2}\min\sum_{x,y}\frac{(\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y))^2}{2P_{XY}(x,y)}
\end{aligned}$$

where the minimization is over $\Delta_y,\Delta_{xy}(x\in\mathcal{X},y\in\mathcal{Y})$ subject to the constraints:

1) $\sum_y\Delta_yH(X|Y=y)-\sum_{x,y}\Delta_{xy}P_Y(y)(\log P_{X|Y}(x|y)+1)=r$;
2) $\sum_y\Delta_y=0$;
3) $\sum_x\Delta_{xy}=0$ for all $y\in\mathcal{Y}$.

Introduce the Lagrange multipliers $\alpha,\beta,\gamma_y(y\in\mathcal{Y})$ for these constraints, and define

$$\begin{aligned}
G=&\sum_{x,y}P_{XY}^{-1}(x,y)(\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y))^2\\
&-\beta\sum_y\Delta_y-\sum_{x,y}\gamma_y\Delta_{xy}-\alpha\sum_y\Delta_yH(X|Y=y)\\
&+\alpha\sum_{x,y}\Delta_{xy}P_Y(y)(\log P_{X|Y}(x|y)+1).
\end{aligned}$$

The Karush–Kuhn–Tucker conditions yield

$$\begin{aligned}
\frac{\partial G}{\partial\Delta_y}&=\sum_x\frac{2P_{X|Y}(x|y)}{P_{XY}(x,y)}(\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y))\\
&\quad-\alpha H(X|Y=y)-\beta\\
&=0 \qquad (29)\\
\frac{\partial G}{\partial\Delta_{xy}}&=\frac{2P_Y(y)}{P_{XY}(x,y)}(\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y))+\alpha P_Y(y)\\
&\quad\times(\log P_{X|Y}(x|y)+1)-\gamma_y\\
&=0. \qquad (30)
\end{aligned}$$

By (29) and constraint 3), we get

$$\Delta_y=\frac{P_Y(y)}{2}(\alpha H(X|Y=y)+\beta) \qquad (31)$$

which, together with (30), implies that

$$\begin{aligned}
\Delta_{xy}=\frac{P_{X|Y}(x|y)}{2P_Y(y)}[&\gamma_y-\alpha P_Y(y)(\log P_{X|Y}(x|y)+1)\\
&-\alpha P_Y(y)H(X|Y=y)-\beta P_Y(y)].
\end{aligned} \qquad (32)$$

Substituting (31) into constraint 2), we obtain

$$\frac{\alpha}{2}H(X|Y)+\frac{\beta}{2}=0. \qquad (33)$$

Similarly, it follows by (32) and constraint 3) that

$$\gamma_y=(\alpha+\beta)P_Y(y),\qquad y\in\mathcal{Y}. \qquad (34)$$

In view of (32) and (34), we have

$$\Delta_{xy}=-\frac{\alpha P_{X|Y}(x|y)}{2}[\log P_{X|Y}(x|y)+H(X|Y=y)]. \qquad (35)$$

It can be seen by substituting (31) and (35) into constraint 1) that

$$\frac{\beta}{2}H(X|Y)+\frac{\alpha}{2}\sum_{x,y}P_{XY}(x,y)\log^2P_{X|Y}(x|y)=r$$

which, together with (33), yields

$$-\frac{\alpha}{2}H^2(X|Y)+\frac{\alpha}{2}\sum_{x,y}P_{XY}(x,y)\log^2P_{X|Y}(x|y)=r. \qquad (36)$$

By (31), (33), and (35), it can be easily verified that

$$\begin{aligned}
&\Delta_yP_{X|Y}(x|y)+\Delta_{xy}P_Y(y)\\
&\quad=\frac{P_{XY}(x,y)}{2}(\alpha H(X|Y=y)+\beta)\\
&\qquad-\frac{\alpha P_{XY}(x,y)}{2}[\log P_{X|Y}(x|y)+H(X|Y=y)]\\
&\quad=-\frac{\alpha P_{XY}(x,y)}{2}[H(X|Y)+\log P_{X|Y}(x|y)]. \qquad (37)
\end{aligned}$$

The proof is complete by combining (36) and (37).

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[2] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.
[3] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. Fundamentals*, vol. E78-A, no. 9, pp. 1063–1070, Sep. 1995.
[4] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, pp. 430–443, May 1982.
[5] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 4, pp. 585–592, Jul. 1982.
[6] J. K. Wolf, "Tribute to David Slepian," *IEEE Inf. Theory Soc. Newslett.*, vol. 58, pp. 3–4, Mar. 2008.
[7] A. D. Wyner, "Recent results in Shannon theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 2–10, Jan. 1974.
[8] T. C. Ancheta, Jr., "Syndrome-source-coding and its universal generalizations," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 4, pp. 432–436, Jul. 1976.
[9] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 2, pp. 155–165, Mar. 1980.
[10] R. G. Gallager, "Source coding with side information and universal coding," MIT LIDS, Cambridge, MA, Tech. Rep. LIDS-P-937, 1976.
[11] F. Jelinek, *Probabilistic Information Theory*. New York: McGraw-Hill, 1968.
[12] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.
[13] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[14] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on $\mathbb{Z}_m$ and binary LDPC codes with belief propagation decoders," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 56–81, Jan. 2007.
[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[16] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 82–85, Jan. 1979.
[17] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 3, pp. 357–359, May 1973.
[18] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. IT-2, no. 5, pp. 8–19, Sep. 1956.
[19] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1675–1682, Nov. 1991.
[20] G. Caire and E. Biglieri, "Linear block codes over cyclic groups," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1246–1256, Sep. 1995.
[21] D.-K. He, L. Lastras-Montaño, E.-H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian-Wolf coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5607–5627, Dec. 2009.
[22] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 5–12, Jan. 1981.
[23] F. Jelinek, "Evaluation of expurgated bound exponents," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 501–505, May 1968.
[24] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
[25] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[26] J. Chen, D.-K. He, and A. Jagmohan, "On the duality between Slepian-Wolf coding and channel coding under mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4006–4018, Sep. 2009.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2003 and 2006, respectively.

He was a Postdoctoral Research Associate at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, from 2005 to 2006, and a Josef Raviv Memorial Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. Currently, he is an Assistant Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.

Da-ke He received the B.S. and M.S. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, Hubei, China, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2003.

From 1996 to 1998, he worked at Apple Technology China, Zhuhai, China, as a Software Engineer. From 2003 to 2004, he worked at the Department of Electrical and Computer Engineering, University of Waterloo as a Postdoctoral Research Fellow at the Leitch-University of Waterloo Multimedia Communications Lab. From 2005 to 2008, he was a research staff member at the Department of Multimedia Technologies, IBM T. J. Watson Research Center in Yorktown Heights, NY. Since 2008, he has been a Technical Manager in Slipstream Data, a subsidiary of Research In Motion, Waterloo, ON, Canada. His research interests are in source coding theory and algorithm design, multimedia data compression and transmission, multiterminal source coding theory and algorithms, and digital communications.

Ashish Jagmohan received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 1999 and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2002 and 2004, respectively.

Since 2004, he has worked at the Departments of Multimedia Technologies and Memory Systems, IBM T. J. Watson Research Center, Yorktown Heights, NY. His research interests include memory system technologies, video compression, multimedia communication, signal processing, and information theory.

Luis A. Lastras-Montaño (M'96–SM'06) received the Licenciatura in electronics and digital systems from the Universidad Autónoma de San Luis Potosí, México, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY, in 1998 and 2000, respectively.

He has been a Research Staff Member with the IBM T. J. Watson Research Center since 2000, where he is currently a member of the Memory Systems Department. His academic research interests lie in the areas of information, coding theory, and statistical signal processing; his work for IBM is focused on the reliability, performance, and architecture of computer memory systems with a special interest on nonvolatile memory technologies.

En-hui Yang (M'97–SM'00–F'08) was born in Jiangxi, China, on December 26, 1966. He received the B.S. degree in applied mathematics from HuaQiao University, Qianzhou, China, in 1986 and the Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 1991.

Since June 1997, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor and Canada Research Chair in Information Theory and Multimedia Compression. He held a Visiting Professor position at the Chinese University of Hong Kong, Hong Kong, from September 2003 to June 2004; positions of Research Associate and Visiting Scientist at the University of Minnesota, Minneapolis-St. Paul, the University of Bielefeld, Bielefeld, Germany, and the University of Southern California, Los Angeles, from January 1993 to May 1997; and a faculty position (first as an Assistant Professor and then an Associate Professor) at Nankai University, Tianjin, China, from 1991 to 1992. He is the founding Director of the Leitch-University of Waterloo Multimedia Communications Laboratory, and a Co-Founder of SlipStream Data, Inc. (now a subsidiary of Research In Motion). His current research interests are multimedia compression, multimedia watermarking, multimedia transmission, digital communications, information theory, source and channel coding including distributed source coding and space-time coding, Kolmogorov complexity theory, and applied probability theory and statistics.

Dr. Yang is a recipient of several research awards including the 1992 Tianjin Science and Technology Promotion Award for Young Investigators; the 1992 third Science and Technology Promotion Award of Chinese National Education Committee; the 2000 Ontario Premier's Research Excellence Award, Canada; the 2000 Marsland Award for Research Excellence, University of Waterloo; the 2002 Ontario Distinguished Researcher Award; the prestigious Inaugural (2007)

Premier's Catalyst Award for the Innovator of the Year; and the 2007 Ernest C. Manning Award of Distinction, one of the Canada's most prestigious innovation prizes. Products based on his inventions and commercialized by SlipStream received the 2006 Ontario Global Traders Provincial Award and were deployed by over 2200 Service Providers in more than 50 countries, servicing millions of home and wireless subscribers worldwide every day. He served, among many other roles, as a General Co-Chair of the 2008 IEEE International Symposium on Information Theory, a Technical Program Vice-Chair of the 2006 IEEE International Conference on Multimedia & Expo (ICME), the Chair of the award committee for the 2004 Canadian Award in Telecommunications, a Co-Editor of the 2004 Special Issue of the IEEE TRANSACTIONS ON INFORMATION THEORY, a Co-Chair of the 2003 U.S. National Science Foundation (NSF) workshop on the interface of Information Theory and Computer Science, and a Co-Chair of the 2003 Canadian Workshop on Information Theory. He currently also serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. He is a Fellow of the Canadian Academy of Engineering and a Fellow of the Royal Society of Canada: the Academies of Arts, Humanities and Sciences of Canada.