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

Jun Chen Da-ke He
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598, USA
Email: {junchen, dakehe}@us.ibm.com

En-hui Yang
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Email: E.Yang@ece.uwaterloo.ca

***Abstract*— A codebook-level duality between Slepian-Wolf coding and channel coding is established. Specifically, it is shown that using linear codes over $\mathbb{Z}_M$ (the ring of integers mod M), each Slepian-Wolf coding problem is equivalent to a channel coding problem for a semi-symmetric additive channel under optimal decoding, belief propagation decoding, and minimum entropy decoding. Various notions of symmetric channels are discussed and their connections with semi-symmetric additive channels are clarified.**

I. Introduction

Consider the problem (see Fig. 1) of encoding $\{X_i\}_{i=1}^{\infty}$ with side information $\{Y_i\}_{i=1}^{\infty}$ at the decoder. Here $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is a memoryless process with joint probability distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. Throughout this paper, $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be finite with $\mathcal{X} = \mathbb{Z}_M$ and $\mathcal{Y} = \mathbb{Z}_N$ unless specified otherwise; for any positive integer K, $+_K$ and $-_K$ denote modulo-K addition and subtraction, respectively, while $a =_K b$ means $a -_K b = 0$.

Slepian and Wolf [1] proved a surprising result[1] that the minimum rate for reconstructing $\{X_i\}_{i=1}^{\infty}$ at the decoder with asymptotically zero error probability is $H(X|Y)$, which is the same as the case where the side information $\{Y_i\}_{i=1}^{\infty}$ is available at the decoder.

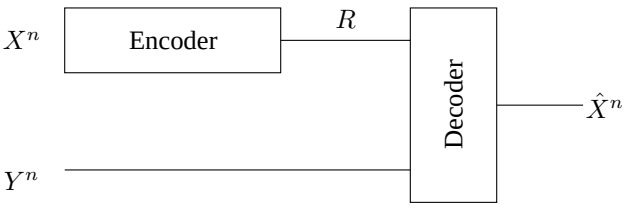

Fig. 1. Slepian-Wolf coding

Shortly after Slepian and Wolf's seminal work, Wyner [3] pointed out the possibility of using linear codes for Slepian-Wolf coding. The scheme works as follows: given the source sequences X^n, the encoder sends $X^n\mathbf{H}$ to the decoder, where $\mathbf{H}$ is the parity check matrix of a linear code $\mathcal{C}$; the decoder then tries to recover X^n from $X^n\mathbf{H}$ given the side information Y^n. Wyner also noticed an intriguing connection between Slepian-Wolf coding and channel coding in a simple example. Suppose that both $\mathcal{X}$ and $\mathcal{Y}$ are binary, and the correlation between X and Y can be modelled by a binary-symmetric channel with parameter $p \in [0, 0.5)$ (BSC(p)), i.e., $X_i = Y_i +_2 Z_i$ for all $i \geq 1$, where Z_i denotes a binary random variable (independent of Y_i) that takes value 1 with probability p. Let $\mathbf{H}$ be a $n \times k$ parity check matrix of a binary linear channel code $\mathcal{C}$ for which there exists a decoding function $g(\cdot)$ such that $Z^n = (Z_1, Z_2, \cdots, Z_n)$ can be decoded from its syndrome $Z^n\mathbf{H}$ with error probability ϵ. Now in the Slepian-Wolf problem, upon receiving the syndrome $S^k = X^n\mathbf{H}$, the decoder calculates

$$S^k +_2 Y^n\mathbf{H} = (X^n +_2 Y^n)\mathbf{H} = Z^n\mathbf{H}$$

and then uses $g(\cdot)$ to recover Z^n with error probability ϵ. Since $X^n = Y^n +_2 Z^n$, X^n can also be recovered with error probability ϵ. It is well-known [4] that the capacity of binary symmetric channel is achievable with linear codes, therefore, we can let the rate $\frac{n-k}{n}$ of channel code $\mathcal{C}$ be arbitrarily close to the channel capacity $1 - H_b(p)$ while maintaining any prescribed error probability $\epsilon > 0$. Hence, the compression rate $\frac{k}{n}$ of Wyner's coding scheme can be arbitrarily close to $H(X|Y) = H_b(p)$, which is exactly the Slepian-Wolf limit. Throughout this paper, $H_b(\cdot)$ stands for the binary entropy function, i.e., $H_b(p) = -p\log p - (1-p)\log(1-p)$, and the logarithm function is to base 2.

If we view $g(\cdot)$ as the *maximum likelihood* (ML) decoding function for BSC(p), then it is not hard to verify that the decoding in the aforementioned example is exactly the *maximum a posteriori* (MAP) decoding for Slepian-Wolf coding. Therefore, Wyner's simple example suggests that a general linear codebook-level duality may exist between Slepian-Wolf coding and channel coding under optimal decoding and Slepian-Wolf code design might be reduced to channel code design. Unfortunately, designing practical capacity-approaching channel codes was still a formidable task at that time. As a result, Wyner's observation had relatively little impact on the design of practical Slepian-Wolf codes.

Inspired by the potential applications of distributed data compression in various networks and multimedia systems, Slepian-Wolf code design has received much attention in recent years. Moveover, due to the revolutionary advance in the development of capacity-approaching channel codes (e.g.,

[1]The original problem considered by Slepian and Wolf is more general. But it can be shown that the general problem can be reduced to this special case via time-sharing or source-splitting [2].

Turbo codes and low-density parity-check (LDPC) codes) and practical decoding algorithms (e.g., belief propagation decoding and linear programming decoding), Wyner's channel coding approach to Slepian-Wolf coding suddenly becomes feasible. Indeed, almost all the existing practical Slepian-Wolf codes [5]–[13] are designed using linear channel codes.

However, although Wyner's idea has been extremely influential, some of its subtleties have been largely neglected. From the design point of view, the central problem is to construct the parity check matrix $\mathbf{H}$. If we view the linear code $\mathcal{C}$ as a channel code, then we have to know for which channel the linear code $\mathcal{C}$ should be designed. Therefore, the following two steps are crucial for Slepian-Wolf code design: 1. Given any distribution P_{XY} in Slepian-Wolf coding, identify a dual channel $P_{V|U}$ such that a good linear code for channel $P_{V|U}$ can be used as a good Slepian-Wolf code for distribution P_{XY}; 2. Design a good linear code for channel $P_{V|U}$. It can be argued that the first step is more important since the second one has been extensively studied in channel coding theory. Surprisingly, little attention has be paid to the first step in the literature. This phenomenon might be explained by the fact that the dual channel $P_{V|U}$ is often assumed (either explicitly or implicitly) to be equal to $P_{Y|X}$, where $P_{Y|X}$ is the conditional distribution of Y given X induced by the joint distribution P_{XY}. This assumption is considered to be natural since the role of X in Slepian-Wolf coding is similar to channel input in channel coding while the role of Y is similar to channel output. Actually there is even a theoretical justification for this assumption. For example, it was shown in [14] that a good Slepian-Wolf code for distribution P_{XY} can be obtained by partitioning the typical sequences (with respect to P_X) in $\mathcal{X}^n$ into roughly $2^{nH(X|Y)}$ channel codes (for channel $P_{Y|X}$), each of rate approximately $I(X;Y)$. However, Slepian-Wolf codes constructed in this way are nonlinear, and therefore, do not fit into Wyner's linear coding approach. Unfortunately, this result has been misinterpreted by many practitioners in the area of Slepian-Wolf code design to justify $P_{Y|X}$ as the right dual channel. We will show that choosing $P_{Y|X}$ as the dual channel leads to a wrong design metric in Wyner's framework.

To give a simple explanation, we temporarily assume $\mathcal{X}$ is binary. In Wyner's framework, the rate of Slepian-Wolf code is equal to the rate of syndrome, so we have the following equation

$$R_{SW} = 1 - R_{CH}$$

where R_{SW} is the rate of Slepian-Wolf code, and R_{CH} is the rate of linear channel code $\mathcal{C}$. It is clear that minimizing R_{SW} is equivalent to maximizing R_{CH}. If the dual channel is $P_{Y|X}$, then the maximum achievable R_{CH} is the capacity of channel $P_{Y|X}$, which is denoted by $C(P_{Y|X})$. Now consider any distribution P_{XY} with the property that P_X is non-uniform, and channel $P_{Y|X}$ is output-symmetric in the sense of [15]. It was shown [16] that in this case

$$H(X|Y) < 1 - C(P_{Y|X}).$$

That is, even if we can design a linear channel code that achieves the capacity $C(P_{Y|X})$, the resulting Slepian-Wolf code rate $R_{SW} = 1 - C(P_{Y|X})$ is still bounded away from the fundamental limit $H(X|Y)$. This phenomenon was also observed in [17] and led to the claim that in this case the Slepian-Wolf limit is not achievable with linear channel codes. Now consider another example. Let P_{XY} be a joint distribution satisfying the property that P_X is uniform, but the capacity-achieving input distribution for channel $P_{Y|X}$ is non-uniform. In this case, we have

$$\begin{aligned} H(X|Y) &= H(X) - I(X;Y) \\ &= 1 - I(X;Y) \\ &> 1 - C(P_{Y|X}). \end{aligned}$$

This implies that if one design a linear channel code with rate close to the capacity $C(P_{Y|X})$, then the resulting Slepian-Wolf code rate would beat the fundamental limit $H(Y|X)$. Obviously, this leads to a contradiction. But one can argue that the maximum rate achievable with linear codes is not $C(P_{Y|X})$, but the mutual information across the channel $P_{Y|X}$ with the uniform input, which is denoted by $I(P_{Y|X})$. Since P_X is uniform in the current example, we have

$$\begin{aligned} H(X|Y) &= H(X) - I(X;Y) \\ &= 1 - I(P_{Y|X}), \end{aligned}$$

which seemingly resolves the contradiction.

However, the above example can be slightly modified to make the contradiction unresolvable. We fix a conditional probability distribution $P_{Y|X}$ and assume[2] that $H(X|Y)$ is maximized by a non-uniform P_X. Let P_{XY} be the joint distribution induced by $P_{Y|X}$ and the maximizer P_X. For the conditional entropy $H(X|Y)$ associated with this joint distribution, we have

$$\begin{aligned} H(X|Y) &> H(\tilde{X}|\tilde{Y}) \\ &= H(\tilde{X}) - I(\tilde{X};\tilde{Y}) \\ &= 1 - I(P_{Y|X}) \end{aligned}$$

where $\tilde{X}$ is a binary random variable with the uniform distribution and $\tilde{Y}$ is a random variable generated by $\tilde{X}$ through channel $P_{Y|X}$. This example shows that in Wyner's linear coding framework, adopting $P_{Y|X}$ as the dual channel is fundamentally flawed.

It should be noted that in Wyner's simple example, it is $P_{X|Y}$, rather than $P_{Y|X}$, that is used as the dual channel. Nevertheless, except for Wyner's example, there is also no justification for considering $P_{X|Y}$ as a candidate for the dual channel; especially when the size of $\mathcal{X}$ and $\mathcal{Y}$ are different, linear channel codes designed for $P_{X|Y}$ can not be directly used to encode X^n.

Intuitively, the dual channel $P_{V|U}$ should contain all the essential information in P_{XY}. In this sense, neither $P_{Y|X}$ nor

[2]Such a conditional probability distribution $P_{Y|X}$ can be easily constructed. Consider the case where $\mathcal{X} = \mathcal{Y} = \{0,1\}$. Let $P_{Y|X}(1|0) = 0.2$, and $P_{Y|X}(0|1) = 0.3$.

$P_{X|Y}$ preserves enough information about P_{XY}, and therefore, they cannot be the right dual channel in general.

Probably the first general result on this problem appeared in [16] (see also [18]), where $\mathcal{X}$ is assumed to be binary. It is shown that using binary LDPC codes, each Slepian-Wolf coding problem is a dual to a channel coding problem for a binary-input output-symmetric channel. Different from Wyner's approach, belief propagation decoding, instead of optimal decoding, is used, and the duality is established under density evolution.

A result of the same nature [19] has been established under optimal decoding for the case where $|\mathcal{X}|$ is a prime number p. Specifically, a class of semi-symmetric additive channels is introduced in [19]; and it is shown that using linear codes over GF(p), each slepian-wolf coding problem is equivalent to a channel coding problem for a semi-symmetric additive channel under optimal decoding. This result, when specialized to the case where $\mathcal{X}$ is binary, is consistent with the result in [16], albeit they are derived under different decoding algorithms.

A surprising conclusion one can draw from [16] and [19] is that no matter whether the joint distribution P_{XY} possesses any symmetric structure or not, the dual channel $P_{V|U}$ is always symmetric. On the other hand, this result is very natural in retrospect. It is well-known that the Slepian-Wolf limit is achievable with linear codes [20]. But it is also known that linear codes cannot be used directly to achieve the capacity of asymmetric channel whose optimal input distribution is not uniform. Therefore, it is not hard to imagine that if the dual channel does exist, it must be a certain symmetric channel.

In this paper, we will generalize and strengthen the results in [16] and [19]. Specifically, the duality between Slepian-Wolf coding and channel is established in the general finite alphabet case. To do so, we use linear codes over $\mathbb{Z}_M$ (i.e., the ring of integers mod M) instead of linear codes over finite fields. Since the duality in [16] was established for belief propagation decoding under density evolution, it is an asymptotic result (in codeword length) concerning an ensemble of codes. By utilizing the property of semi-symmetric additive channel, we will prove a much stronger result — the duality under belief propagation decoding actually holds for every individual linear code of arbitary length. In certain sense, our result can be viewed as a completion of Wyner's framework.

The rest of this paper is organized as follows. In Section II, we review the definition of a semi-symmetric additive channel, and establish the linear codebook-level duality between Slepian-Wolf coding for an arbitrarily correlated source-side information pair and channel coding for a particular semi-symmetric additive channel under optimal decoding, belief propagation decoding as well as minimum entropy decoding. In Section III, the relation between semi-symmetric additive channels and other well-established notions of symmetric channels in the literature is clarified, which sheds further light on the duality between Slepian-Wolf coding and channel coding. We conclude the paper in Section IV.

II. Duality

A. Dual Channel

We first review the definition of a semi-symmetric additive channel.

Definition 1 ([19]): A discrete memoryless channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$ with input alphabet $\mathcal{U} = \mathbb{Z}_{|\mathcal{U}|}$ and output alphabet $\mathcal{V} = \mathbb{Z}_{|\mathcal{V}|}$ is defined to be additive and semi-symmetric if there exists a positive integer N such that $|\mathcal{V}| = N|\mathcal{U}|$, and for any input U_i, the output is given by

$$V_i = NU_i +_{|\mathcal{V}|} W_i, \quad \text{for all } i \geq 1 \tag{1}$$

where $\{W_i\}_{i=1}^{\infty}$ is an i.i.d. process (with marginal distribution P_W) that is independent of the channel input and takes values in $\mathcal{V}$.

For any nonnegative integer k and positive integer j, let $Q(k,j) \triangleq \lfloor \frac{k}{j} \rfloor$ and $R(k,j) \triangleq k - jQ(k,j)$. It is easy to verify that

$$Q(V_i, N) = U_i +_{|\mathcal{U}|} Q(W_i, N), \tag{2}$$

$$R(V_i, N) = R(W_i, N). \tag{3}$$

Since V_i can be uniquely recovered from $Q(V_i, N)$ and $R(V_i, N)$, we can view (2) and (3) as an alternative representation of channel model (1).

Theorem 1: The capacity-achieving input distribution for channel model (1) (also, (2) and (3)) is the uniform distribution over $\mathcal{U}$. Furthermore, the channel capacity is given by

$$C(P_{V|U}) = \log |\mathcal{U}| - H(Q(W,N)|R(W,N))$$

where W is a generic random variable with probability distribution P_W.

Remark: This theorem was proved in [19]. See Section III for a more general result.

It turns out that this class of channels is intrinsically related to Slepian-Wolf coding. Given an i.i.d. random process $\{(X_i, Y_i)\}_{i=1}^{\infty}$ with marginal distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X} = \mathbb{Z}_M$ and $\mathcal{Y} = \mathbb{Z}_N$, one can define a semi-symmetric additve channel $P_{V|U}$ with $\mathcal{U} = \mathcal{X}$ and $\mathcal{V} = \mathbb{Z}_{MN}$ by specifying $\{W_i\}_{i=1}^{\infty}$ with the following equations

$$Q(W_i, N) = X_i, \quad R(W_i, N) = Y_i, \quad \text{for all } i \geq 1. \tag{4}$$

Through this construction, Slepian-Wolf coding and channel coding are put in the same probability space, which enables us to relate the error events of these two problems. The main result of this section is to show that the constructed channel $P_{V|U}$ is the right dual channel. The following empirical evidences are immediate:

1) By Theorem 1 and Eqn. (4), we have
$$H(X|Y) = \log M - C(P_{V|U}). \tag{5}$$
This is desirable in Wyner's framework (at least when $M = 2$) in view of Eqn. (1).
2) Using the alternative representation given in (2) and (3), we can write the dual channel $P_{V|U}$ in the form
$$Q(V_i, N) = U_i +_M X_i, \tag{6}$$
$$R(V_i, N) = Y_i. \tag{7}$$

It can be seen from (7) that the decoder has access to $\{Y_i\}_{i=1}^{\infty}$ just as in the Slepian-Wolf coding problem. Furthermore, since $\{U_i +_M X_i\}_{i=1}^{\infty}$ is available at the decoder, recovering $\{U_i\}_{i=1}^{\infty}$ is equivalent to recovering $\{X_i\}_{i=1}^{\infty}$. This illustrates the intimate connection between Slepian-Wolf coding and channel coding as well as the intuition behind the construction of the dual channel.

B. Linear Codes over $\mathbb{Z}_M$

To have a precise formulation of the duality between Slepian-Wolf coding and channel coding, we need to introduce linear codes over $\mathbb{Z}_M$ (i.e., the ring of integers mod M). There are several reasons for using linear codes over $\mathbb{Z}_M$ instead of more standard linear codes over GF(q). Firstly, this allows us to establish the duality result for general finite alphabet size $|\mathcal{X}|$. But more importantly, linear codes over $\mathbb{Z}_M$ match the cyclic symmetry of the dual channel $P_{V|U}$, while linear codes over GF(q) do not (unless q is a prime number, in which case they coincide with $\mathbb{Z}_M$ codes). As it will be seen later, to establish the duality between Slepian-Wolf coding and channel coding using linear codes over GF(q), the definition of a semi-symmetric additive channel has to be slightly modified.

Linear codes over $\mathbb{Z}_M$ have been well-studied, especially in the context of coded modulation [21]–[23].

Definition 2 ([22]): A linear block code $\mathcal{C}$ of length n over $\mathbb{Z}_M$ is a subgroup of $\mathbb{Z}_M^n$ (we write $\mathcal{C} <_s \mathbb{Z}_M^n$), where $\mathbb{Z}_M^n$ is the group of n-tuples of elements of $\mathbb{Z}_M$ with componentwise addition.

The subgroup $\mathcal{C}$ partitions the group $\mathbb{Z}_M^n$ into $\frac{M^n}{|\mathcal{C}|}$ disjoint cosets, each of size $|\mathcal{C}|$. We can label each coset with a single element (a coset representative) drawn from that coset. A set of representatives for the cosets of C in $\mathbb{Z}_M^n$ (one representative for each coset) is denoted $[\mathbb{Z}_M^n/\mathcal{C}]$.

We can see that only the group property is needed in order to define the cosets of $\mathcal{C}$. This turns out to be sufficient for the purpose of establishing the duality between Slepian-Wolf coding and channel coding under optimal decoding. However, for belief-propagation decoding, we have to define the parity check matrix of $\mathcal{C}$. Fortunately, this is possible due to the ring structure of $\mathbb{Z}_M$. Here we collect some basic facts about $\mathbb{Z}_M$ codes from [23].

Theorem 2 ([23]): Let $\mathcal{C} \subset \mathbb{Z}_M^n$. The following statements are equivalent:

1) $\mathcal{C}$ is a subgroup of $\mathbb{Z}_M^n$, i.e., $\mathcal{C} <_s \mathbb{Z}_M^n$.
2) There exist an integer r $(0 \le r \le n)$, a set of linearly independent vectors $\{x_1^n, x_2^n, \cdots, x_r^n\} \subset \mathbb{Z}_M^n$, and a set of nested ideals of $\mathbb{Z}_M$ (not necessarily distinct)

   $$\{0\} <_s a_r\mathbb{Z}_M <_s \cdots <_s a_2\mathbb{Z}_M <_s a_1\mathbb{Z}_M <_s \mathbb{Z}_M$$

   such that $\mathcal{C}$ can be written as the direct sum

   $$\mathcal{C} = \bigoplus_{i=1}^{r} a_i\mathbb{Z}_M x_i^n. \tag{8}$$

   Moveover, the ideals and r are uniquely determined by $\mathcal{C}$ and M.
3) There exists a unique lattice Λ, $M\mathbb{Z}^n <_s \Lambda <_s \mathbb{Z}^n$, such that $\mathcal{C}$ is isomorphic to $\Lambda/M\mathbb{Z}^n$ (we write $\mathcal{C} \simeq \Lambda/M\mathbb{Z}^n$). Given any set $[\Lambda/M\mathbb{Z}^n]$ of coset representatives, $\mathcal{C}$ can be written as

   $$\mathcal{C} = [\Lambda/M\mathbb{Z}^n] \bmod M.$$

Statement 2) allows us to define the generator matrix of $\mathcal{C}$. Since $a_i\mathbb{Z}_M \simeq \mathbb{Z}_{M/a_i}$, we have [23]

$$\mathcal{C} = \bigoplus_{i=1}^{r} a_i\mathbb{Z}_M x_i^n \simeq \mathbb{Z}_{M/a_1} \times \mathbb{Z}_{M/a_2} \times \cdots \times \mathbb{Z}_{M/a_r}.$$

Define the information group $\mathcal{J}$ of $\mathcal{C}$ as

$$\mathcal{J} = \{z^n = (z_1, z_2 \cdots, z_r, 0, \cdots, 0) : z_i \in \mathbb{Z}_{M/a_i}, i = 1, 2 \cdots, r\}.$$

Write $\mathcal{C} = \mathcal{J}\mathbf{G}$, where $\mathbf{G}$ is an $n \times n$ matrix representing the isomorphism between $\mathcal{J}$ and $\mathcal{C}$ (expressed in a given basis). It follows from (8) that $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} a_1x_1^n \\ a_2x_2^n \\ \vdots \\ a_rx_r^n \\ 0^n \\ \vdots \\ 0^n \end{pmatrix}.$$

$\mathbf{G}$ is called the generator matrix of $\mathcal{C}$.

Let Λ be the lattice given in Statement 3). If Λ^* denotes the dual lattice of Λ, then $M\Lambda^*$ can be written as $M\Lambda^* = M\mathbb{Z}^n + \mathcal{C}^{\perp}$, where

$$\mathcal{C}^{\perp} = [M\Lambda^*/M\mathbb{Z}^n] \bmod M$$

is the dual code of $\mathcal{C}$ [23]. Similarly, we can use Statement 2) to construct the generator matrix $\mathbf{G}^{\perp}$ of $\mathcal{C}^{\perp}$. The parity check matrix of $\mathcal{C}$ is defined as

$$\mathbf{H} = (\mathbf{G}^{\perp})^T.$$

It can be shown that $c^n\mathbf{H} = 0^n \iff c^n \in \mathcal{C}$.

Define the syndrome group $\mathcal{S} = \mathrm{Im}(\mathbf{H})$ using the surjective homomorphism $\mathbf{H} : \mathbb{Z}_M^n \to \mathcal{S}$. It can be shown [23] that

$$\mathbb{Z}_M^n/\mathcal{C} \simeq \mathcal{S} \simeq \mathbb{Z}_{a_1} \times \mathbb{Z}_{a_2} \times \cdots \times \mathbb{Z}_{a_r} \times \mathbb{Z}_M^{n-r}.$$

Therefore, the syndrome group $\mathcal{S}$ can be used to label the cosets of $\mathcal{C}$. Specifically, a coset of $\mathcal{C}$ is denoted by $\mathcal{C}_{s^n}$ ($s^n \in \mathcal{S}$) if $c^n\mathbf{H} = s^n$ for all c^n in this coset. For example, we have $\mathcal{C} = \mathcal{C}_{0^n}$.

C. Optimal Decoding

Now we are ready to establish the duality between the Slepian-Wolf coding problem for the source distribution P_{XY} and the channel coding problem for the dual channel $P_{V|U}$ under optimal decoding. Let $\mathcal{C}$ be a linear block code over $\mathbb{Z}_M$ and $\{\mathcal{C}_{s^n}\}_{s^n \in S}$ be the sets of cosets of $\mathcal{C}$ in $\mathbb{Z}_M^n$.

In channel coding, each codeword $c^n \in \mathcal{C}$ is transmitted with probability $\frac{1}{|\mathcal{C}|}$. The channel code rate is

$$R_{CH} = \frac{1}{n} \log |\mathcal{C}|.$$

In Slepian-Wolf coding, given the source sequence $\tilde{x}^n$, the encoder computes $\tilde{s}^n = \tilde{x}^n \mathbf{H}$ and sends the syndrome $\tilde{s}^n$ to the decoder. The Slepian-Wolf code rate is given by

$$R_{SW} = \frac{1}{n} \log |\mathcal{S}| = \frac{1}{n} \log \frac{M^n}{|\mathcal{C}|}.$$

Therefore, we have

$$R_{SW} = \log M - R_{CH}. \tag{9}$$

Let $\tilde{y}^n$ be the realization of Y^n, and $\tilde{c}^n$ be the transmitted codeword. Temporarily ignoring the tie-breaking issue, the optimal channel decoding rule (ML decoding) can be written as

$$\hat{c}^n = \arg\max_{c^n \in \mathcal{C}} Pr(U^n +_M X^n = \tilde{c}^n +_M \tilde{x}^n, Y^n = \tilde{y}^n \, | U^n = c^n)$$

while the optimal Slepian-Wolf decoding rule (MAP decoding) is given by

$$\hat{x}^n = \arg\max_{x^n \in \mathcal{C}_{\tilde{s}^n}} Pr(X^n = x^n | Y^n = \tilde{y}^n). \tag{10}$$

The optimal channel decoding rule can be rewritten as

$$\begin{aligned} \hat{c}^n &= \arg\max_{c^n \in \mathcal{C}} Pr(Y^n = \tilde{y}^n) Pr(U^n +_M X^n = \tilde{c}^n +_M \tilde{x}^n \, | U^n = c^n, Y^n = \tilde{y}^n) \\ &= \arg\max_{c^n \in \mathcal{C}} Pr(U^n +_M X^n = \tilde{c}^n +_M \tilde{x}^n | U^n = c^n, \, Y^n = \tilde{y}^n) \\ &= \arg\max_{c^n \in \mathcal{C}} Pr(X^n = \tilde{c}^n -_M c^n +_M \tilde{x}^n | Y^n = \tilde{y}^n). \end{aligned}$$

Since $(\tilde{c}^n -_M c^n +_M \tilde{x}^n)\mathbf{H} = \tilde{x}^n \mathbf{H} = \tilde{s}^n$ for all $c^n \in \mathcal{C}$, it follows that

$$\hat{c}^n = \tilde{c}^n +_M \tilde{x}^n -_M \arg\max_{x^n \in \mathcal{C}_{\tilde{s}^n}} Pr(X^n = x^n | Y^n = \tilde{y}^n). \tag{11}$$

It can be seen that the optimal decoding rule for the dual channel $P_{V|U}$ can be converted to the optimal Slepian-Wolf decoding rule for the source distribution P_{XY}. Therefore, the decoding error probabilities in these two problems must be the same. We can also see that the channel decoding error is independent of the transmitted codeword and only depends on the realization of X^n and Y^n. Moreover, any realization of X^n and Y^n that leads to a channel decoding error also causes a Slepian-Wolf decoding error, and vice versa. In this sense, the duality exists not only at the level of average decoding error probability, but also at the more fundamental individual sequence level. This observation is crucial for analyzing the duality under belief propagation decoding because it is hard to analyze the average decoding error probability directly in that setting.

Now consider the case where $Pr(X^n = x^n | Y^n = \tilde{y}^n)$ is maximized by more than one $x^n \in \mathcal{C}_{\tilde{s}^n}$. It is clear that even if different tie-breaking rules are adopted in Slepian-Wolf coding and channel coding, the decoding error probabilities, conditioned on any realization of Y^n, are still the same. However, to preserve the strong duality at the individual sequence level, the tie-breaking rules used in Slepian-Wolf coding and channel coding must be consistent. It can be seen that every tie-breaking rule for (10) leads to a tie-breaking rule for (11). Specifically, we can rewrite (10) and (11) in the following form:

$$\begin{aligned} \hat{x}^n &= f(\theta, \tilde{s}^n, \tilde{y}^n) \\ \hat{c}^n &= \tilde{c}^n +_M \tilde{x}^n -_M f(\theta, \tilde{s}^n, \tilde{y}^n) \end{aligned}$$

where θ is the realization of a tie-breaking random variable Θ, which, condtioned on $X^n\mathbf{H}$ and Y^n, is independent of X^n and the transmitted codeword. It is obvious that any realization of (X^n, Y^n, Θ) that leads to a Slepian-Wolf decoding error also causes a channel decoding error, and vice versa. More generally, we can use two different tie-breaking random variables Θ_1 and Θ_2 in Slepian-Wolf decoding and channel decoding, respectively. As long as the conditional probability distribution $P_{\Theta_1|X^n\mathbf{H},Y^n}$ is identical with $P_{\Theta_2|X^n\mathbf{H},Y^n}$, the decoding error probabilities in Slepian-Wolf coding and channel coding, conditioned on any realization of X^n and Y^n, are still the same. Moveover, all the results continue to hold even if we replace the linear code $\mathcal{C}$ by any of its cosets in the channel coding part.

It can be shown by the standard arguments [24] that when M is a prime number, linear codes over $\mathbb{Z}_M$ can achieve the capacity of semi-symmetric additive channels under optimal decoding, and hence, can achieve the Slepian-Wolf limit in view of (5) and (9). Therefore, the duality between Slepian-Wolf coding and channel coding not only provides a framework for Slepian-Wolf code design, but also leads to a new proof of the Slepian-Wolf theorem[3].

More generally, for any process $\{X_i, Y_i\}_{i=1}^{\infty}$ with $\mathcal{X} = \mathbb{Z}_M$ and $\mathcal{Y} = \mathbb{Z}_N$, we can construct a dual channel using (6) and (7). Here $\{X_i, Y_i\}_{i=1}^{\infty}$ does not need to be memoryless or even stationary. It is easy to verify that, even in such a general setting, the duality between Slepian-Wolf coding and channel coding at the individual sequence level continues to hold under optimal decoding. One can use this fact to prove the Slepian-Wolf theorem for general sources using Verdú and Han's channel capacity formula [25].

D. Belief Propagation Decoding

The literature on the belief propagation (BP) algorithm and its application to channel coding is vast (see, for example, [15], [26], [27]). A detailed description of Slepian-Wolf decoding using belief propagation algorithm can be found in [16].

It should be emphasized that although the cosets are uniquely determined by the linear code $\mathcal{C}$, there can be many different parity check matrices associated with the same linear

[3]Requiring M to be a prime number is not a real restriction for Slepian-Wolf coding since we can always extend $\mathcal{X}$ by adding symbols with zero probability.

code. Under optimal decoding, the performance is completely determined by the linear code $\mathcal{C}$ since different parity check matrices just give different labelling of the cosets. However, it is known that the performance under belief propagation decoding depends not only on the linear code $\mathcal{C}$ but also on the code representation (i.e., the choice of the parity check matrix $\mathbf{H}$). It will be seen that the duality established under belief propagation decoding is, in certain sense, weaker than that established under optimal decoding because it requires that not only the same linear code but also the same parity check matrix should be used in Slepian-Wolf coding and channel coding. In practice, due to the complexity constraint, belief propagation algorithm is mostly used to decode LDPC codes. However, since here we are only concerned with the duality between Slepian-Wolf coding and channel coding, we do not restrict the linear code $\mathcal{C}$ to be an LDPC code (i.e, $\mathbf{H}$ does not need to be a sparse matrix).

There is no change in the encoding procedure, therefore, we shall only focus on the decoding part. It is well-known that the parity check matrix $\mathbf{H}$ can be represented by a Tanner graph [28]. Let $\mathcal{C}_i$ be the set of check nodes that are connected to variable node i. Let $\mathcal{V}_j$ be the set of variable nodes that are connected to check node j. Let h_{ij}, the (i,j) entry of the parity check matrix $\mathbf{H}$, be the label on the edge connecting variable node i and check node j. As in the case of optimal coding, we let $\tilde{s}^n = \tilde{x}^n\mathbf{H}$, where $\tilde{x}^n$ is the realization of X^n. Also, let $\tilde{c}^n$ be the transmitted codeword, and $\tilde{y}^n$ be the realization of Y^n.

In certain sense, the duality between Slepian-Wolf coding for the source distribution P_{XY} and channel coding for the dual channel $P_{V|U}$ is barely surprising: since the channel decoder has access to $(\tilde{c}^n +_M \tilde{x}^n)\mathbf{H} = \tilde{x}^n\mathbf{H}$ and $\tilde{y}^n$, it can first do Slepian-Wolf decoding to recover $\tilde{x}^n$, which in turn can be used to recover $\tilde{c}^n$ from $\tilde{c}^n + \tilde{x}^n$. Therefore, any Slepian-Wolf decoder can be converted to a channel decoder in this way. We have seen that the channel decoder converted from the optimal Slepian-Wolf decoder (i.e., MAP decoder) is equivalent to the optimal channel decoder (i.e., ML decoder). However, it is less transparent whether the channel decoder converted from the BP Slepian-Wolf decoder is equivalent to the BP channel decoder. We will show that the answer is affirmative.

Now we proceed to establish the duality between Slepian-Wolf coding and channel coding under belief propagation decoding. In Slepian-Wolf coding, the initial message[4] at variable node i is

$$M_{v=i}^{(0)} = \left[m_{v=i}^{(0)}(0), m_{v=i}^{(0)}(1), \cdots, m_{v=i}^{(0)}(M-1)\right]$$

where

$$m_{v=i}^{(0)}(k) = Pr(X_i = k, Y_i = \tilde{y}_i), \quad k \in \mathbb{Z}_M.$$

In channel coding, the initial message at variable node i is

$$\overline{M}_{v=i}^{(0)} = \left[\overline{m}_{v=i}^{(0)}(0), \overline{m}_{v=i}^{(0)}(1), \cdots, \overline{m}_{v=i}^{(0)}(M-1)\right]$$

[4]There are many different ways to represent a message, but they are all equivalent. So here we just choose the most basic one.

where $\overline{m}_{v=i}^{(0)}(k)$, $k \in \mathbb{Z}_M$, is given by

$$\overline{m}_{v=i}^{(0)}(k) = Pr(U_i +_M X_i = \tilde{c}_i +_M \tilde{x}_i, Y_i = \tilde{y}_i | U_i = k).$$

Note that for any $k \in \mathbb{Z}_M$, we have

$$\begin{aligned}
&\overline{m}_{v=i}^{(0)}(k) \\
&= Pr(Y_i = \tilde{y}_i) Pr(U_i +_M X_i = \tilde{c}_i +_M \tilde{x}_i | U_i = k, Y_i = \tilde{y}_i) \\
&= Pr(Y_i = \tilde{y}_i) Pr(X_i = \tilde{c}_i +_M \tilde{x}_i -_M k | Y_i = \tilde{y}_i) \\
&= Pr(X_i = \tilde{c}_i +_M \tilde{x}_i -_M k, Y_i = \tilde{y}_i) \\
&= m_{v=i}^{(0)}(\tilde{c}_i +_M \tilde{x}_i -_M k).
\end{aligned}$$

Now consider the message from check node j to variable node i in the first iteration. In Slepian-Wolf coding, the message is

$$M_{c=j,v=i}^{(1)} = \left[m_{c=j,v=i}^{(1)}(0), \cdots, m_{c=j,v=i}^{(1)}(M-1)\right]$$

where $m_{c=j,v=i}^{(1)}(k)$, $k \in \mathbb{Z}_M$, is given by

$$m_{c=j,v=i}^{(1)}(k) = \sum_{\mathcal{A}_k} \prod_{l \in \mathcal{C}_j \setminus \{i\}} m_{v=l}^{(0)}(x_l)$$

and

$$\mathcal{A}_k = \left\{ (x_l)_{l \in \mathcal{C}_j \setminus \{i\}} : h_{ij}k + \sum_{l \in \mathcal{C}_j \setminus \{i\}} h_{lj}x_l =_M \tilde{s}_j \right\}.$$

In channel coding, the message is

$$\overline{M}_{c=j,v=i}^{(1)} = \left[\overline{m}_{c=j,v=i}^{(1)}(0), \cdots, \overline{m}_{c=j,v=i}^{(1)}(M-1)\right]$$

where $\overline{m}_{c=j,v=i}^{(1)}(k)$, $k \in \mathbb{Z}_M$, is given by

$$\overline{m}_{c=j,v=i}^{(1)}(k) = \sum_{\overline{\mathcal{A}}_k} \prod_{l \in \mathcal{C}_j \setminus \{i\}} \overline{m}_{v=l}^{(0)}(u_l)$$

and

$$\overline{\mathcal{A}}_k = \left\{ (u_l)_{l \in \mathcal{C}_j \setminus \{i\}} : h_{ij}k + \sum_{l \in \mathcal{C}_j \setminus \{i\}} h_{lj}u_l =_M 0 \right\}.$$

We have

$$\begin{aligned}
\overline{m}_{c=j,v=i}^{(1)}(k) &= \sum_{\overline{\mathcal{A}}_k} \prod_{l \in \mathcal{C}_j \setminus \{i\}} \overline{m}_{v=l}^{(0)}(u_l) \\
&= \sum_{\overline{\mathcal{A}}_k} \prod_{l \in \mathcal{C}_j \setminus \{i\}} m_{v=l}^{(0)}(\tilde{c}_l +_M \tilde{x}_l -_M u_l) \\
&= \sum_{\mathcal{A}_{\tilde{c}_i +_M \tilde{x}_i -_M k}} \prod_{l \in \mathcal{C}_j \setminus \{i\}} m_{v=l}^{(0)}(x_l) \qquad (12) \\
&= m_{c=j,v=i}^{(1)}(\tilde{c}_i +_M \tilde{x}_i -_M k), \quad k \in \mathbb{Z}_M
\end{aligned}$$

where (12) follows from the fact that $(u_l)_{l \in \mathcal{C}_j \setminus \{i\}}$ is in $\overline{\mathcal{A}}_k$ if and only if $(\tilde{c}_l +_M \tilde{x}_l -_M u_l)_{l \in \mathcal{C}_j \setminus \{i\}}$ is in $\mathcal{A}_{\tilde{c}_i +_M \tilde{x}_i -_M k}$.

Then consider the message from variable node i to check node j in the first iteration. In Slepian-Wolf coding, the message is

$$M_{v=i,c=j}^{(1)} = \left[m_{v=i,c=j}^{(1)}(0), \cdots, m_{v=i,c=j}^{(1)}(M-1)\right]$$

where $m^{(1)}_{v=i,c=j}(k)$, $k \in \mathbb{Z}_M$, is given by

$$m^{(1)}_{v=i,c=j}(k) = m^{(0)}_{v=i}(k) \prod_{l \in \mathcal{V}_i \setminus \{j\}} m^{(1)}_{c=l,v=i}(k).$$

In channel coding, this message is

$$\overline{M}^{(1)}_{v=i,c=j} = \left[\overline{m}^{(1)}_{v=i,c=j}(0), \cdots, \overline{m}^{(1)}_{v=i,c=j}(M-1)\right]$$

where $\overline{m}^{(1)}_{v=i,c=j}(k)$, $k \in \mathbb{Z}_M$, is given by

$$\overline{m}^{(1)}_{v=i,c=j}(k) = \overline{m}^{(0)}_{v=i}(k) \prod_{l \in \mathcal{V}_i \setminus \{j\}} \overline{m}^{(1)}_{c=l,v=i}(k).$$

Note that

$$\begin{aligned} & \overline{m}^{(1)}_{v=i,c=j}(k) \\ = \;& m^{(0)}_{v=i}(\tilde{c}_i + \tilde{x}_i - k) \prod_{l \in \mathcal{V}_i \setminus \{j\}} m^{(1)}_{c=l,v=i}(\tilde{c}_i +_M \tilde{x}_i -_M k) \\ = \;& m^{(1)}_{v=i,c=j}(\tilde{c}_i +_M \tilde{x}_i -_M k), \quad k \in \mathbb{Z}_M. \end{aligned}$$

By induction, for any iteration number t, any variable node i and any check node j, we have

$$\begin{aligned} \overline{m}^{(t)}_{v=i,c=j}(k) &= m^{(t)}_{v=i,c=j}(\tilde{c}_i +_M \tilde{x}_i -_M k), \\ \overline{m}^{(t)}_{c=j,v=i}(k) &= m^{(t)}_{c=j,v=i}(\tilde{c}_i +_M \tilde{x}_i -_M k). \end{aligned}$$

When a decision is to be made at variable node i at the tth iteration, variable node i will form a decision vector. In Slepian-Wolf coding, the decision vector is

$$D^{(t)}_i = [d^{(t)}_i(0), d^{(t)}_i(1), \cdots, d^{(t)}_i(M-1)]$$

where

$$d^{(t)}_i(k) = m^{(0)}_{v=i}(k) \prod_{l \in \mathcal{V}_i} m^{(t)}_{c=l,v=i}(k), \quad k \in \mathbb{Z}_M.$$

In channel coding, the decision vector is

$$\overline{D}^{(t)}_i = [\overline{d}^{(t)}_i(0), \overline{d}^{(t)}_i(1), \cdots, \overline{d}^{(t)}_i(M-1)]$$

where

$$\overline{d}^{(t)}_i(k) = \overline{m}^{(0)}_{v=i}(k) \prod_{l \in \mathcal{V}_i} \overline{m}^{(t)}_{c=l,v=i}(k), \quad k \in \mathbb{Z}_M.$$

Clearly, we have

$$\begin{aligned} & \overline{d}^{(t)}_i(k) \\ = \;& m^{(0)}_{v=i}(\tilde{c}_i +_M \tilde{x}_i -_M k) \prod_{l \in \mathcal{V}_i} m^{(t)}_{c=l,v=i}(\tilde{c}_i +_M \tilde{x}_i -_M k) \\ = \;& d^{(t)}_i(\tilde{c}_i +_M \tilde{x}_i -_M k), \quad k \in \mathbb{Z}_M. \end{aligned}$$

A few comments are ready:

1) $\overline{d}^{(t)}_i(k)$ is maximized at $k = \tilde{c}_i$, then $d^{(t)}_i(k')$ is maximized at $k' = \tilde{x}_i$, and vice versa. Therefore, given any realization of (X^n, Y^n), a correct decision is made in Slepian-Wolf decoding if and only if a correct decision is made in channel decoding.
2) The decoding error is independent of the transmitted codeword.
3) It can be verified that all the results still hold even if in the channel coding part, we replace the linear code $\mathcal{C}$ by any of its cosets.
4) The tie-breaking issue is similar to the optimal decoding case, and therefore, is omitted.
5) Similar to the optimal decoding case, the duality under belief propagation decoding holds at the individual sequence level. Therefore, it is unnecessary to have the assumption that the process $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is memoryless. However, it should be noted that the belief propagation algorithm used here is developed for memoryless processes. Its performance might not be good if used directly for processes with memory.

E. Minimum Entropy Decoding

The duality between Slepian-Wolf coding and channel coding can also be established under universal decoding. We shall focus on minimum entropy decoding although it will be clear that any universal Slepian-Wolf decoding rule for the source distribution P_{XY} can be converted to a universal channel decoding rule for the dual channel $P_{V|U}$.

The encoding procedure is the same as before, and is omitted. In Slepian-Wolf coding, the minimum entropy decoder selects the reconstruction sequence $\hat{x}^n$ in $\mathcal{C}_{\tilde{s}^n}$ with the property that the entropy of the joint empirical distribution of $\hat{x}^n$ and $\tilde{y}^n$ is minimized. Csiszár [20] proved a surprising result that the Slepian-Wolf limit is achievable universally using linear codes in conjunction with minimum entropy decoding. For the dual channel $P_{V|U}$, we can first use the minimum entropy Slepian-Wolf decoder to recover $\tilde{x}^n$, which in turn can be used to recover $\tilde{c}^n$ from $\tilde{c}^n + \tilde{x}^n$. This is clearly a universal channel decoder. Therefore, Csiszar's result directly implies a universal channel coding theorem for semi-symmetric additive channels. However, strictly speaking, the channel decoder converted from the minimum entropy Slepian-Wolf decoder is not equivalent to the standard minimum entropy channel decoder. Nevertheless, we can view it as a variant of the standard minimum entropy channel decoder.

The duality between Slepian-Wolf coding and channel coding also demystifies Csiszár's surprising result. The universal Slepian-Wolf coding theorem and the sufficiency of linear codes can be viewed as manifestation of the compound channel coding theorem [29] and the fact that the uniform distribution is capacity-achieving for all semi-symmetric additive channels.

F. Linear Codes over GF(q)

So far we have focused on linear codes over $\mathbb{Z}_M$. To establish the duality between Slepian-Wolf coding and channel coding using linear codes over GF(q), we can define the dual channel in the form $V = (U \oplus X, Y)$, where $\oplus$ is the addition operation in GF(q). To guarantee that the dual channel is well-defined, we need to assume $\mathcal{U} = \mathcal{X} =$ GF(q). The proof of duality between Slepian-Wolf coding and channel coding using linear codes over GF(q) follows almost verbatim from that using linear codes over $\mathbb{Z}_M$. The only change is that

modulo-M operations should be replaced by the corresponding operations in GF(q).

It should be mentioned that Csiszar's result [20] was derived using linear codes over GF(q). However, since linear codes over $\mathbb{Z}_M$ and linear codes over GF(q) coincide when $M = q = p$, where p is a prime number, Csiszar's result also holds for linear codes over $\mathbb{Z}_p$. Moreover, since we are only interested in the duality between Slepian-Wolf coding and channel coding while the optimality of codes is not our main concern, using linear codes over $\mathbb{Z}_M$ provides us the freedom to treat arbitrary finite alphabet size without adding zero-probability symbols.

III. Symmetry, Equivalence and Ordering

It should be clear from the analysis in the previous section that the symmetry of the dual channel $P_{V|U}$ plays an important role. Since there are many different definitions of symmetric channels in the literature, it is instructive to clarify the relations and differences among them. It will be seen that such a comparison further illuminates the connection between Slepian-Wolf coding and channel coding. For simplicity, we shall focus on finite-input finite-output discrete memoryless channels with input alphabet $\mathcal{U} = \mathbb{Z}_M$ and output alphabet $\mathcal{V}$ although most of the results hold in a much more general setting.

Definition 3 ([30]): For any function $T : \mathcal{V} \to \mathcal{V}$, let T^k denote the corresponding k-times self-composition of T. An M-ary-input channel $P_{V|U} : \mathbb{Z}_M \to \mathcal{V}$ is cyclic-symmetric if there exists a bijective transform $T : \mathcal{V} \to \mathcal{V}$ such that $T^M(v) = v$ for all $v \in \mathcal{V}$ and

$$P_{V|U}(v|0) = P_{V|U}(T^u(v)|u)$$

for all $u \in \mathbb{Z}_M, v \in \mathcal{V}$.

It can be verified that every semi-symmetric additive channel defined by (1) in Section II is a cyclic-symmetric channel by choosing the bijective function T to be $T(v) = v +_{|\mathcal{V}|} N$ for all $v \in \mathcal{V}$.

Now let $P_{V|U}$ be a cyclic-symmetric channel with bijective transform T. Under the action of T, the channel output alphabet $\mathcal{V}$ can be partitioned into N equivalence class $\mathcal{V}_0, \mathcal{V}_1, \cdots, \mathcal{V}_{N-1}$ for some positive integer N such that two elements v_1 and v_2 belong to the same class $\mathcal{V}_j$ if and only if $v_1 = T^u(v_2)$ for some $u \in \mathbb{Z}_M$. Equivalence class is closed under transform T, i.e., $T(\mathcal{V}_j) = \mathcal{V}_j$ for all $j \in \mathbb{Z}_N$. This implies that the conditional probability $P_{V|U}(\mathcal{V}_j|u)$ does not depend on u, and we have

$$Pr(V \in \mathcal{V}_j) = P_{V|U}(\mathcal{V}_j|u) \tag{13}$$

for all $u \in \mathbb{Z}_M, j \in \mathbb{Z}_N$. We can represent the channel output V in the vector form $(\overline{V}_1, \overline{V}_2)$, where $\overline{V}_1$ denotes the equivalence class in which V resides while $\overline{V}_2$ specifies which element in that equivalence class is V. It follows from (13) that $\overline{V}_1$ is independent of the channel input and we can view it as the channel state information at the decoder. Since $T^M(v) = v$ for all $v \in \mathcal{V}$, it follows that $|\mathcal{V}_j|$ must divide exactly into M for all $j \in \mathbb{Z}_N$. We can construct a joint distribution P_{XY} with $\mathcal{X} = \mathbb{Z}_M$ and $\mathcal{Y} = \mathbb{Z}_N$ such that

$$\begin{aligned} P_Y(j) &= Pr(V \in \mathcal{V}_j), \\ P_{X|Y}(k|j) &= \frac{|\mathcal{V}_j|}{M} P_{V|U}(T^k(v_j)|0) \end{aligned}$$

for all $k \in \mathbb{Z}_M, j \in \mathbb{Z}_N$, where v_j is an arbitrary representative element of $\mathcal{V}_j$. Let $P_{\tilde{V}|\tilde{U}}$ be the dual semi-symmetric additive channel of P_{XY} (see (6) and (7)). We can see that Y and $\tilde{U} +_M X$ play the roles of $\overline{V}_1$ and $\overline{V}_2$, respectively. It is easy to verify that actually $P_{V|U}$ and $P_{\tilde{V}|\tilde{U}}$ are identical after merging the indistinguishable output symbols[5] and relabelling. Therefore, in this sense, these two definitions of symmetric channels are equivalent. Henceforth, only the name "cyclic-symmetric channel"[6] will be used.

It is well-known [30], [31] that every (asymmetric) channel can be converted to a cyclic-symmetric channel through the standard argument based on the coset code ensemble. But it should be emphasized that in channel coding, such a symmetrization argument is introduced to simplify the analysis, and is not information-lossless. However, in Slepian-Wolf coding, the cosets are intrinsic since they play the role of bins [32]. Therefore, in Slepian-Wolf coding, symmetrization is a consequence of the nature of the problem itself. This explains why we can convert every Slepian-Wolf coding problem (no matter symmetric or not) to a channel coding problem for a cyclic-symmetric channel without loss of optimality.

Since using linear codes over $\mathbb{Z}_M$, every Slepian-Wolf coding problem is equivalent to a channel coding problem for a cyclic-symmetric channel, we can say two source distributions P_{XY} and $P_{\tilde{X}\tilde{Y}}$ in Slepian-Wolf coding are equivalent if their corresponding dual channels are identical (after merging indistinguishable output symbols and relabelling). More generally, through the duality between Slepian-Wolf coding and channel coding, every partial order on the domain of channel transition probabilities induces a partial order on the domain of source distributions in Slepian-Wolf coding. Such kind of ordering is often useful for establishing monotonicity results under various decoding algorithms.

Now we turn to another important class of symmetric channels defined by Gallager [24].

Definition 4: A discrete memoryless channel $P_{V|U} : \mathbb{Z}_M \to \mathcal{V}$ is defined to be symmetric if $\mathcal{V}$ can be partitioned into subsets in such a way that for each subset of the matrix of transition probabilities (using input as rows and outputs of

[5]For any discrete memoryless channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$, two output symbols v_1 and v_2 are called indistinguishable if $P_{V|U}(v_1|u) = 0 \Leftrightarrow P_{V|U}(v_2|u) = 0$ and the value of $\frac{P_{V|U}(v_1|u)}{P_{V|U}(v_2|u)}$ does not depend on u for $u \in \{u \in \mathcal{U} : P_{V|U}(v_1|u) > 0\}$.

[6]As pointed out in [30], the definition of cyclic-symmetric channels is similar to the definition of signal sets matched to groups in [22]. Interestingly, signal sets matched to groups are equivalent to Slepian signal sets (i.e., Slepian's "group codes for the Gaussian channel"). So in certain sense, Slepian not only invented Slepian-Wolf coding (with Wolf), but also helped to develop the right notion of symmetry that is crucial for establishing the duality between Slepian-Wolf coding and channel coding. However, Slepian himself might not realize this point.

the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.

It is easy to verify that every cyclic-symmetric channel is also G-symmetric (i.e., symmetric in the sense of Gallager). But the reverse is not true. The reason is simple: the G-symmetry is preserved under both column permutation and row permutation of the channel transition probability matrix; however, the cyclic-symmetry is preserved under column permutation but not row permutation. One may argue that such a difference is caused by the fact the input symbols are already ordered in the definition of a cyclic-symmetric channel, which is clearly an unnecessary restriction. However, we will show that there exist transition probability matrices satisfying G-symmetry but not convertible to any transition probability matrix satisfying cyclic-symmetry even if row permutation is allowed. For example, consider the type of transition probability matrices given by the following Latin square [33]

$$\begin{pmatrix} a & b & c & d & e \\ b & a & d & e & c \\ c & e & a & b & d \\ d & c & e & a & b \\ e & d & b & c & a \end{pmatrix}$$

It can be verified that there does not exist a bijective transform T satisfying the conditions in Definition 3 for this kind of matrices even if row permutation is allowed. On the other hand, it is obvious that any channel with a transition probability matrix of this type is G-symmetric.

The following definition of symmetric channels was used in [34] by Cover and Thomas.

Definition 5: A channel $P_{V|U} : \mathbb{Z}_M \rightarrow \mathcal{V}$ is said to be symmetric if every row of the transition matrix is a permutation of every other row, and all the column sums are equal.

It can be verified that the binary erasure channel is not CT-symmetric (i.e., symmetric in the sense of Cover and Thomas). But the binary erasure channel is cyclic-symmetric (and therefore, G-symmetric). However, there exist channels that are CT-symmetric but not G-symmetric (and therefore, not cyclic-symmetric). Consider the following channel transition probability matrix

$$\begin{pmatrix} 0.1 & 0.3 & 0.35 & 0.05 & 0.2 \\ 0.3 & 0.35 & 0.05 & 0.2 & 0.1 \\ 0.35 & 0.05 & 0.1 & 0.2 & 0.3 \\ 0.05 & 0.1 & 0.3 & 0.35 & 0.2 \end{pmatrix}.$$

It is easy to verify that the channel with this transition probability matrix is CT-symmetric but not G-symmetric.

The motivation behind both G-symmetry and CT-symmetry is to guarantee that the capacity-achieving input distribution is the uniform distribution. We shall define a notion of symmetry which includes both G-symmetry and CT-symmetry as special cases, but preserves all their essential features.

Definition 6: A discrete memoryless channel $P_{V|U} : \mathbb{Z}_M \rightarrow \mathcal{V}$ is said to be GCT-symmetric if $\mathcal{V}$ can be partitioned into subsets in such a way that for each subset of the matrix of transition probabilities (using input as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and the column sums are equal.

It should be obvious from the definition that GCT-symmetry includes both G-symmetry and CT-symmetry as special cases.

Although GCT-symmetry is much more general than cyclic-symmetry, they share many important features. Let $(\mathcal{V}_0, \mathcal{V}_1, \cdots, \mathcal{V}_{N-1})$ be a partition of $\mathcal{V}$ as specified in Definition 6. The role of $(\mathcal{V}_0, \mathcal{V}_1, \cdots, \mathcal{V}_{N-1})$ is similar to that of equivalence classes in the cyclic-symmetric channel case. We can represent the output V in the vector form (S, W), where S denotes the subset in which V resides while W specifies which element in that subset is V. More precisely, let $S = s$ if $V \in \mathcal{V}_s$, $s \in \mathbb{Z}_N$. Since each subset of the transition matrix has the property that each row is a permutation of each other row, it implies that the conditional probability $P_{S|U}(s|u)$ does not depend on u, i.e., S is independent of the channel input. We shall view S as the channel state information at the decoder. Clearly, S corresponds to $\overline{V}_1$ in the cyclic-symmetric channel case, which in turn corresponds to Y in Slepian-Wolf coding.

For any input distribution P_U, we have

$$I(U;V) = I(U;S,W) = I(U;W|S)$$

and the channel capacity is given by

$$\begin{aligned} C(P_{V|U}) &= \max_{P_U} I(U;W|S) \\ &= \max_{P_U} \sum_{s=0}^{N-1} P_S(s) I(U;W|S=s). \end{aligned}$$

If the state information S is available at the encoder, then the input distribution can be optimized for each state realization, and the channel capacity with the state information at both the encoder and decoder is given by

$$\begin{aligned} & C_E(P_{V|U}) \\ = & \sum_{s=0}^{N-1} P_S(s) \max_{P_{U|S=s}} I(U;W|S=s) \\ = & \sum_{s=0}^{N-1} P_S(s) \max_{P_{U|S=s}} [H(W|S=s) - H(W|U,S=s)]. \end{aligned}$$

Again, since each subset of the transition matrix has the property that each row is a permutation of each other row, the conditional entropy $H(W|U,S=s)$ does not depend on $P_{U|S=s}$. Therefore, we can write

$$H(W|U,S=s) = H(W|U=0,S=s).$$

Since the column sums are equal in each subset of the transition matrix, it implies that if $P_{U|S=s}$ is the uniform distribution over $\mathbb{Z}_M$, then the resulting $P_{W|S=s}$ is the uniform distribution over $\mathcal{V}_s$. Clearly, the maximum value of

$H(W|S=s)$ is $\log|\mathcal{V}_s|$ and is achieved when $P_{W|S=s}$ is the uniform distribution. Therefore, we have

$$C_E(P_{V|U}) = \sum_{j=0}^{N-1} P_S(s)[\log|\mathcal{V}_s| - H(W|U=0,S=s)].$$

Since the optimal input distribution $P_{U|S=j}$ is always the uniform distribution over $\mathbb{Z}_M$ no matter what the realization of the state information S is, we have

$$\begin{aligned} & C(P_{V|U}) \\ = & \max_{P_U} \sum_{s=0}^{N-1} P_S(s) I(U;W|S=s) \\ = & \sum_{s=0}^{N-1} P_S(s) \max_{P_{U|S=s}} I(U;W|S=s) \\ = & C_E(P_{V|U}) \\ = & \sum_{s=0}^{N-1} P_S(s)[\log|\mathcal{V}_s| - H(W|U=0,S=s)], \quad (14) \end{aligned}$$

i.e., the state information at the encoder does not help to increase the channel capacity, which resembles the phenomenon in Slepian-Wolf coding that the fundamental limit is unaffected no matter the side information is available at the encoder or not. We can also see that Theorem 1 is a special case of (14).

IV. Conclusion

We have established a duality between Slepian-Wolf coding and channel coding. It should be pointed out that this duality holds at the level of each individual linear codebook, and therefore, is considerably stronger than the formula-level duality results based on the random coding argument. Indeed, this codebook-level duality, though technically simple to prove, is surprisingly powerful. Besides its implication on the practical Slepian-Wolf code design, it also provides a link for translating many difficult and profound results of channel coding directly to those of Slepian-Wolf coding. In view of the fact that channel coding is the most extensively studied area in information theory, it is natural to expect that Slepian-Wolf coding will benefit from channel coding through this link.

References

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

[2] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting," in *IEEE International Symposium on Information Theory*, Ulm, Germany, June 29 - July 4 1997, p. 271.

[3] A. D. Wyner, "Recent results in Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2-10, Jan. 1974.

[4] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, vol. 3, pp. 37-46, Mar. 1955.

[5] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inform. Theory*, Mar. 2003.

[6] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250-1276, June 2002.

[7] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes" *IEEE Communications Letters*, 5:417-419, Oct. 2001.

[8] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," In *IEEE GLOBECOM*, pp. 1400-1404, Nov. 2001.

[9] A. Aaron and B. Girod, "Compression with side information using turbo codes," In *Proceeding of IEEE Data Compression Conference (DCC)*, pp. 252-261, Apr. 2002.

[10] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440-442, 2002.

[11] D. Schongberg, K. Ramchandran, and S. S. Pradhan, "Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources," *Proceeding of IEEE Data Compression Conference (DCC)*, Snowbird, UT, Mar. 2004.

[12] V. Stankovic, A. Liveris, Z. Xiong, and C. Georghiades, "On code design for the general Slepian-Wolf problem and for lossless multiterminal communication networks," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1495-1507, Apr. 2006.

[13] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 3546-3561, Aug. 2006.

[14] I. Csiszár and J. Körner,. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[15] T. Richardson and R. Urbanke, "The capacity of low-density paritycheck codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599-618, Feb. 2001.

[16] J. Chen, D.-k. He, and A. Jagmohan "Approaching the Slepian-Wolf limit with LDPC coset codes," *IEEE Trans. Inform. Theory*, submitted for publication.

[17] J. Li, Z. Tu, and R. S. Blum, "Slepian-Wolf coding for nonuniform sources using Turbo codes," *Proceeding of IEEE Data Compression Conference (DCC)*, pp. 312-321, Snowbird, UT, March 2004.

[18] J. Chen, D.-k. He, and A. Jagmohan, "Slepian-Wolf code design via source-channel correspondence," *IEEE International Symposium on Information Theory*, Seattle, WA, Jul. 9-Jul. 14, 2006.

[19] D.-k. He and E.-h. Yang, "On the duality between Slepian-Wolf coding and channel coding," *IEEE International Symposium on Information Theory*, Seattle, WA, Jul. 9-Jul. 14, 2006.

[20] I. Csiszár, "Linear codes for sources and source networks: error exponents, universal coding," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 585-592, Jul. 1982.

[21] F. R. Kschischang, P. G. De Buda, and S. Pasupathy, "Block coset codes for M-ary phase shift keying," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 900-913, Aug. 1989.

[22] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1675-1682, Nov. 1991.

[23] G. Caire and E. Biglieri, "Linear block codes over cyclic groups," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1246-1256, Sept. 1995.

[24] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[25] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147-1157, Jul. 1994.

[26] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[27] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.

[28] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533-547, Sept. 1981.

[29] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2148-2177, Oct. 1998.

[30] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on Z_m and binary LDPC codes with belief propagation decoders," *IEEE Trans. Inform. Theory*, to appear.

[31] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution and code performance bound," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1636-1652, Jul. 2003.

[32] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, pp. 226-228, Mar. 1975.

[33] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.