The Equivalence Between Slepian-Wolf Coding and Channel Coding Under Density Evolution

Jun Chen, Da-ke He, and Ashish Jagmohan

***Abstract*—We consider Slepian-Wolf code design based on low-density parity-check (LDPC) coset codes. The density evolution formula for Slepian-Wolf coding is derived. An intimate connection between Slepian-Wolf coding and channel coding is then established. Specifically we show that, under density evolution, each Slepian-Wolf coding problem is equivalent to a channel coding problem for a binary-input output-symmetric channel.**

***Index Terms*—Channel coding, linear codes, source coding.**

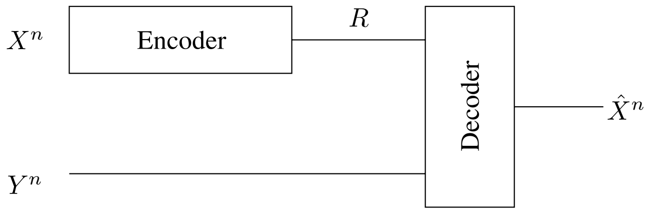

Fig. 1. Slepian-Wolf source coding.

I. Introduction

CONSIDER the problem (see Fig. 1) of encoding the source $\{X_i\}_{i=1}^{\infty}$ with side information $\{Y_i\}_{i=1}^{\infty}$ only at the decoder. Here $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is a stationary and memoryless process with zero-order probability distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. For brevity, we shall identify the memoryless pair $\{(X_i, Y_i)\}$ by using a pair of random variables (X, Y) with distribution P_{XY}. In their landmark paper [1], Slepian and Wolf proved a surprising result that the minimum rate for reconstructing $\{X_i\}_{i=1}^{\infty}$ at the decoder with asymptotically zero error probability is the conditional entropy $H(X|Y)$ of X given Y, which is the same as the case where the side information $\{Y_i\}_{i=1}^{\infty}$ is also available at the encoder. The quantity $H(X|Y)$ is often referred to as the Slepian-Wolf(SW) limit.

Shortly after Slepian and Wolf's seminal work, Wyner [2] pointed out the possibility of using linear codes for SW coding. Consider a simple example[1] in which $\mathcal{X}$ and $\mathcal{Y}$ are both binary, and the channel $P_{X|Y}$ (i.e., the conditional distribution of X given Y induced by P_{XY}) is a binary symmetric channel with parameter $p \in (0, 0.5)$ (BSC(p)), i.e., $X_i = Y_i \oplus Z_i$ for all $i \geq 1$, where $\oplus$ is the modulo-2 addition, and Z_i denotes a binary random variable (independent of Y_i) that takes value 1 with probability p. Note that Y needs not to be uniformly distributed in this example. The SW coding scheme proposed by Wyner works as follows: given the source sequence $X^n = (X_1, X_2, \cdots, X_n)$, 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 observed an intriguing connection between SW coding and channel coding in this example. Let $\mathbf{H}$ be an $n \times k$ parity check matrix of a binary linear channel code $\mathcal{C}$ for which there exists a decoding function $g(\cdot)$ such that Z^n can be decoded from its syndrome $Z^n\mathbf{H}$ with error probability ϵ. Correspondingly in SW coding, upon receiving the syndrome $S^k = X^n\mathbf{H}$, the decoder can calculate

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

and then use the function $g(\cdot)$ to recover Z^n with error probability ϵ. Since $X^n = Y^n \oplus Z^n$, X^n can also be recovered with error probability ϵ. In view of the fact that the capacity of a binary symmetric channel is achievable with linear codes [4], 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 SW coding scheme can be arbitrarily close to $H(X|Y) = H_b(p)$, which is exactly the SW 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 log denotes the logarithm function to base 2.

In the above example, 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 the SW problem. Therefore, Wyner's simple example suggests that there might exist a connection between SW coding and channel coding in the linear coding framework. The significance of this potential connection lies in that it is then possible to convert SW code design to linear channel code design.

In about three decades after [1], SW coding has relatively little impact on practice largely due to the difficulty in designing practical and efficient SW codes. Note that even in Wyner's example, the design of a good channel code for

Paper approved by F. Alajaji, the Editor for Source and Source/Channel Coding of the IEEE Communications Society. Manuscript received January 17, 2008; revised July 17, 2008 and December 3, 2008.

The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Seattle, July 2006.

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. He was with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA when the work was done. He is currently with SlipStream Data, a subsidiary of Research In Motion, 175 Columbia St. West, Waterloo, ON N2L 5Z5, Canada (e-mail: dhe@rim.com).

A. Jagmohan is with the IBM Thomas J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598 USA (e-mail: ashishja@us.ibm.com).

Digital Object Identifier 10.1109/TCOMM.2009.09.080018

[1] It is pointed out in [3] that Slepian and Wolf themselves were the first to notice this example.

0090-6778/09$25.00 © 2009 IEEE

the binary symmetric channel itself was a formidable task. Incidentally, recently we have seen a revolutionary advance in the development of capacity-approaching (linear) channel codes (e.g., Turbo codes and low-density parity-check (LDPC) codes) and practical decoding algorithms (e.g., belief propagation decoding and linear programming decoding). This advance in channel coding theory, coupled with the potential of SW coding for distributed data compression in various networks and multimedia systems, led to a recent surge of interest in practical SW code design. In particular, with a wealth of tools to design linear channel codes available now to exploit, Wyner's linear channel coding approach to SW coding suddenly becomes very appealing.

From the design point of view, the simple example discussed by Wyner [2] implies the following method (hereafter Method A) to design SW codes for a source-side information pair (X, Y).

Step 1: Given the joint probability distribution P_{XY} in SW coding, identify a channel $P_{X|Y}$;
Step 2: Design a good linear code $\mathcal{C}$ for the identified channel;
Step 3: Use the code $\mathcal{C}$ (or equivalently its parity check matrix $\mathbf{H}$) for SW coding (see the example reviewed above).

An alternative design method (hereafter Method B) is to replace the channel in Step 1 by the conditional distribution $P_{Y|X}$ of Y given X.[2] Observe that in both Methods A and B, Step 2 is the standard linear channel code design step. Thus these methods allow us to leverage a large body of channel coding theory directly. Furthermore, from the example reviewed above, we have learned that if $P_{X|Y}$ is a binary symmetric channel, Method A is optimal in the sense that if the linear channel code $\mathcal{C}$ is a capacity-approaching channel code for $P_{X|Y}$, then it can also be used to approach the SW limit for (X, Y). Naturally, it is tempting to generalize that Method A (or Method B) is also optimal for arbitrary P_{XY}. Indeed, almost all the existing practical SW codes [5]–[14] are designed by either dealing with restricted cases (say, Wyner's setup) for which Method A (or Method B) is known to be optimal or following one of the above two methods without questioning its optimality. In the following, we shall argue that contrary to this common assumption, neither Method A nor Method B is optimal for SW coding in general.

Without loss of generality[3], we shall assume $\mathcal{X} = \{0, 1\}$ throughout this paper. According to Step 3 in both Methods A and B, the rate of SW 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 SW 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 equivalent channel is $P_{Y|X}$ as in Method B, 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 $P_X(0) \in (0, 0.5)$ and $P_{Y|X}$ is a BSC(p) with $p \in (0, 0.5)$. It is easy to show 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 SW 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 SW 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 using a linear channel code for channel $P_{Y|X}$ with rate close to the capacity $C(P_{Y|X})$, the resulting SW code rate would beat the fundamental limit $H(X|Y)$. Obviously, this leads to a contradiction. One may 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

$$H(X|Y) = H(X) - I(X;Y) = 1 - I(P_{Y|X}),$$

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[4] 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

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

[2] This channel $P_{Y|X}$ is often considered to be natural since the role of X in SW coding is similar to channel input in channel coding while the role of Y is similar to channel output. Indeed, there is a theoretical justification for this assumption. It was shown in [15] that a good SW 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, SW codes constructed in this way are intrinsically nonlinear, and therefore, do not fit into Wyner's linear coding approach. Unfortunately, this subtle difference seems to have often been overlooked in practice.

[3] The general case can be reduced to this special case via multilevel coding [16]. Specifically, for any finite-alphabet random variable X, we can write it in a binary vector form $(X_1', X_2', \cdots, X_L')$, where X_i', $i = 1, 2, \cdots, L$, are all binary. By the chain rule, we can write $H(X|Y) = H(X_1'|Y) + H(X_2'|X_1', Y) + \cdots + H(X_L'|X_1', \cdots, X_{L-1}', Y)$. Note that the conditional entropy $H(X_k'|X_1', \cdots, X_{k-1}', Y)$, $k = 1, \cdots, L$, can be interpreted as the SW limit for compressing source X_k' with $(X_1', \cdots, X_{k-1}', Y)$ as the side information at the decoder. Therefore, we can decompose the original SW coding problem into a sequence of SW coding problems for binary sources.

[4] 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 $P_{Y|X}(0|1) = 0.3$.

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 equivalent channel is fundamentally flawed.

It should be noted that in Wyner's example, Method A is applied where $P_{X|Y}$, rather than $P_{Y|X}$, is identified as the channel in Step 1. Nevertheless, except for Wyner's example, there is also no justification for considering $P_{X|Y}$ as a candidate for the equivalent 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.

In this paper we address the problem of designing SW codes based on LDPC coset codes. Our strategy is to convert the SW code design problem to an equivalent channel code design problem. The main result in this paper is an explicit construction of the equivalent channel $P_{W|U}$ given the joint distribution P_{XY}. Specifically, for any memoryless pair (X, Y) with joint distribution P_{XY}, our method (hereafter Method C) of constructing (linear) SW codes works as follows.

Step 1: Given the joint probability distribution P_{XY} in SW coding, identify an equivalent channel $P_{W|U}$;

Step 2: Design a good linear code $\mathcal{C}$ for the identified channel;

Step 3: Use the code $\mathcal{C}$ (or equivalently its parity check matrix $\mathbf{H}$) for SW coding.

Contrasting our method with Methods A and B, we see that the key difference lies in Step 1. It turns out that our equivalent channel $P_{W|U}$ is neither $P_{Y|X}$ nor $P_{X|Y}$ in general. Moreover, $P_{W|U}$ is always output-symmetric regardless whether the joint distribution P_{XY} possesses any symmetric structure or not. Though these properties of $P_{W|U}$ seem counter-intuitive at first sight, they are in fact natural in retrospect: It is well known that the SW limit is achievable with linear codes [18]. But it is also known that linear codes cannot be used directly to achieve the capacity of asymmetric channel whose capacity-achieving input distribution is not uniform. Therefore, it is not hard to imagine that if the equivalent channel does exist, it must be a certain symmetric channel. By using the equivalent channel $P_{W|U}$, our method, in contrast to Methods A and B, is always optimal in the sense that if $\mathcal{C}$ achieves the channel capacity of the channel $P_{W|U}$, it can also be used to achieve SW limit of (X, Y).

The rest of this paper is organized as follows. In Section II, we develop the belief-propagation algorithm for SW coding and derive the associated density evolution formula. An intimate connection between SW coding and channel coding under density evolution is established in Section III. Specifically we show that, under density evolution, each SW coding problem is equivalent to a channel coding problem for a binary-input output-symmetric channel. We conclude the paper in Section IV.

II. Belief-Propagation Algorithm and Density Evolution

The literature on LDPC codes is vast (see, for example, [19]–[23] for various definitions). Through this paper, $\mathcal{C}^n(d_v, d_c)$ denotes the ensemble of (d_v, d_c)-regular LDPC codes of length n with variable node degree d_v and check node degree d_c. More generally, we use $\mathcal{C}^n(\lambda, \rho)$ to denote the ensemble of irregular LDPC codes of length n with finite order edge degree distribution polynomials (λ, ρ).

The belief-propagation algorithm [24] is an iterative message-passing algorithm, which can be used to decode LDPC codes. Let $P_{W|U}$ be the channel transition probability. Let $m_{vc}^{(l)}$ denote the message sent from variable node v to its incident check node c in the lth iteration and $m_{cv}^{(l)}$ denote the message sent from check node c to its incident variable node v in the lth iteration. The update equations for the messages under belief propagation are described below [25]:

$$m_{vc}^{(l)} = \begin{cases} m_0, & \text{if } l = 0 \\ m_0 + \sum\limits_{c' \in C_v \setminus \{c\}} m_{c'v}^{(l)}, & \text{if } l \geq 1 \end{cases}$$

$$m_{cv}^{(l)} = \gamma^{-1} \left(\sum_{v' \in V_c \setminus \{v\}} \gamma \left(m_{v'c}^{(l-1)} \right) \right)$$

where C_v is the set of check nodes incident to variable node v, V_c is the set of variable nodes incident to check node c, and $m_0 \triangleq \ln \frac{P_{W|U}(W_i|0)}{P_{W|U}(W_i|1)}$ is the initial message associated with the variable node v. The expression of $\gamma(\cdot)$ can be found in [25].

Given a parity check matrix $\mathbf{H}$, the set of all n-length vectors x^n satisfying $x^n \mathbf{H} = s^k$ for some general syndrome $s^k \in \{0, 1\}^{n-k}$ is called a coset $\mathcal{C}_{s^k}$. Since the Tanner graph is completely determined by the parity check matrix $\mathbf{H}$, all the cosets C_{s^k} are associated with the same Tanner graph. In order to distinguish different cosets by their Tanner graphs, we can label check nodes by their corresponding syndrome values. The belief-propagation algorithm for decoding a coset code is similar to that for decoding a linear code. The only difference lies in the operation at check nodes, which becomes

$$m_{cv}^{(l)} = (-1)^s \gamma^{-1} \left(\sum_{v' \in V_c \setminus \{v\}} \gamma \left(m_{v'c}^{(l-1)} \right) \right),$$

where s is the syndrome value associated with check node c.

The performance of LDPC codes under the belief-propagation algorithm is relatively well-understood for binary-input output-symmetric (BIOS) channels.

Definition 1 ([23], Definition 1): A binary-input channel with transition probability function $P_{W|U}$ from $\mathcal{U}$ to $\mathcal{W}$ with $\mathcal{U} = \{0, 1\}$ is output-symmetric if we have (possibly after relabelling $\mathcal{W}$) $P_{W|U}(w|0) = P_{W|U}(-w|1)$ for all $w \in \mathcal{W}$.

An important property of BIOS channel is that under the belief-propagation algorithm, the decoding error probability is independent of transmitted codeword. So without loss of generality, we can assume the all-zero codeword is transmitted.

In order to analyze the asymptotic (in codeword length) performance of LDPC code ensemble $\mathcal{C}^n(\lambda, \rho)$, a powerful technique called density evolution is developed in [23], [25]. The iterative density evolution formula for BIOS channel is given in the following theorem.

Theorem 1 ([25], Theorem 2): For a given BIOS memoryless channel, let $P^{(0)}$ denote the initial message density of log-likelihood ratios, assuming that the all-zero codeword was transmitted. If, for a fixed degree distribution pair (λ, ρ), $P^{(l)}$ denotes the density of the messages passed from the the variable nodes to the check nodes at the lth iteration of belief propagation then, under the independence assumption

$$P^{(l)} = P^{(0)} \otimes \lambda(\Gamma^{-1}(\rho(\Gamma(P^{(l-1)})))) \qquad (1)$$

where $\otimes$ is the convolution operation, and Γ is the density transformation operator induced by γ.

Remark: Let $p_e^{(l)}$ be the expected number of incorrect messages passed along an edge with a tree-like directed neighborhood of depth at least $2l$ at the lth iteration. We have

$$p_e^{(l)} = \int_{-\infty}^{0^-} P^{(l)}(dm) + \frac{1}{2}\int_{0^-}^{0^+} P^{(l)}(dm).$$

We shall derive a similar density evolution formula for SW coding and establish an intimate connection between SW coding and channel coding under density evolution.

It should be noted that in contrast to channel coding where codewords are assumed be to equally probable, in SW coding X is not necessarily uniformly distributed over $\{0,1\}$ for a general joint distribution P_{XY}. Therefore, we need to incorporate the prior distribution P_X into the belief propagation algorithm for SW decoding. The update equations for the messages in SW decoding are described below:

$$m_{vc}^{(l)} = \begin{cases} m_0, & \text{if } l = 0 \\ m_0 + \sum\limits_{c' \in C_v \setminus \{c\}} m_{c'v}^{(l)}, & \text{if } l \geq 1 \end{cases}$$

$$m_{cv}^{(l)} = (-1)^s \gamma^{-1}\left(\sum_{v' \in V_c \setminus \{v\}} \gamma\left(m_{v'c}^{(l-1)}\right)\right). \qquad (2)$$

where $m_0 \triangleq \ln \frac{P_{X|Y}(0|Y_i)}{P_{X|Y}(1|Y_i)} = \ln \frac{P_{XY}(0,Y_i)}{P_{XY}(1,Y_i)}$. It can be verified that this algorithm produces the exact symbol-by-symbol *a posteriori* estimation of X^n given Y^n when the underlying Tanner graph is a tree. We can see that the only difference from the channel decoding case is the definition of initial message m_0. It will be clear that this small modification has significant consequences on SW code design.

Now we proceed to develop the density evolution formula for this belief-propagation algorithm. We use the standard tree assumption. Let $P^{(l)}(x)$ $(x = 0, 1)$ be the distribution of message from a variable node to a check node at the lth iteration conditioned on that the variable value is x. Similarly, let $Q^{(l)}(x)$ $(x = 0, 1)$ be the distribution of message from a check node to a variable node at the lth iteration conditioned on that the target variable value is x. Assume $P_X(0) = p$. Let $\langle P^{(l)} \rangle = pP^{(l)}(0) + (1-p)P^{(l)}(1) \circ I^{-1}$, where $I(m) \triangleq -m$ is a parity reversing function, and $\circ$ is the composition operation. We have

$$\langle p_e^{(l)} \rangle = \int_{-\infty}^{0^-} \langle P^{(l)} \rangle (dm) + \frac{1}{2}\int_{0^-}^{0^+} \langle P^{(l)} \rangle (dm), \qquad (3)$$

where $\langle p_e^{(l)} \rangle$ is the expected number of incorrect messages sent from a variable node at the lth iteration.

We shall derive a density-evolution formula for $\langle P^{(l)} \rangle$. By the tree assumption, the iterative equations at variable node are given by

$$P^{(l)}(x) = P^{(0)}(x) \otimes (Q^{l-1}(x))^{\otimes(d_v-1)}, \quad x = 0, 1.$$

Let $\mathcal{E} = \{v : 0 \leq v \leq d_c - 1, v \text{ is even}\}$, and $\mathcal{O} = \{v : 0 \leq v \leq d_c - 1, v \text{ is odd}\}$. The iterative equations at check node are

$$\begin{aligned} Q^{(l-1)}(0) \;=\; & \sum_{v \in \mathcal{E}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v} + \\ & \sum_{v \in \mathcal{O}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v} \circ I^{-1}, \end{aligned} \qquad (4)$$

$$\begin{aligned} Q^{(l-1)}(1) \;=\; & \sum_{v \in \mathcal{E}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v} \circ I^{-1} \\ & + \sum_{v \in \mathcal{O}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v}. \end{aligned} \qquad (5)$$

Note that the check node operation depends on its syndrome value s (see (2)). If $s = 1$, then the check node negates the message, which results in the flip of the message distribution. This is the reason why the operator I^{-1} comes into equations (4) and (5).

Comparing (4) and (5), we immediately get

$$Q^{(l-1)}(0) = Q^{(l-1)}(1) \circ I^{-1}. \qquad (6)$$

The expression of $Q^{(l-1)}(0)$ can be simplified as follows:

$$\begin{aligned} Q^{(l-1)}(0) \;=\; & \sum_{v \in \mathcal{E}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v} + \\ & \sum_{v \in \mathcal{O}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1)))^{\otimes v} \circ I^{-1} \\ =\; & \sum_{v \in \mathcal{E}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1) \circ I^{-1}))^{\otimes v} + \\ & \sum_{v \in \mathcal{O}} \binom{d_c - 1}{v} p^{d_c-1-v}(1-p)^v \times \\ & \Gamma^{-1}(\Gamma(P^{(l-1)}(0)))^{\otimes(d_c-1-v)} \otimes \\ & (\Gamma(P^{(l-1)}(1) \circ I^{-1}))^{\otimes v} \\ =\; & \Gamma^{-1}(\Gamma(pP^{(l-1)}(0) + \\ & (1-p)P^{(l-1)}(1) \circ I^{-1}))^{\otimes(d_c-1)} \\ =\; & \Gamma^{-1}(\Gamma(\langle P^{(l-1)} \rangle))^{\otimes(d_c-1)}. \end{aligned} \qquad (7)$$

By (6) and (7), we have

$$\begin{aligned}
\langle P^{(l)}\rangle &= pP^{(l)}(0)+(1-p)P^{(l)}(1)\circ I^{-1}\\
&= pP^{(0)}(0)\otimes\left(Q^{(l-1)}(0)\right)^{\otimes(d_v-1)}+\\
&\quad\left((1-p)P^{(0)}(1)\otimes\left(Q^{(l-1)}(1)\right)^{\otimes(d_v-1)}\right)\circ I^{-1}\\
&= pP^{(0)}(0)\otimes\left(Q^{(l-1)}(0)\right)^{\otimes(d_v-1)}+\\
&\quad\left((1-p)P^{(0)}(1)\circ I^{-1}\right)\otimes\\
&\quad\left(Q^{(l-1)}(1)\circ I^{-1}\right)^{\otimes(d_v-1)}\\
&= pP^{(0)}(0)\otimes\left(Q^{(l-1)}(0)\right)^{\otimes(d_v-1)}+\\
&\quad\left((1-p)P^{(0)}(1)\circ I^{-1}\right)\otimes\left(Q^{(l-1)}(0)\right)^{\otimes(d_v-1)}\\
&= \langle P^{(0)}\rangle\otimes\left(Q^{(l-1)}(0)\right)^{\otimes(d_v-1)}\\
&= \langle P^{(0)}\rangle\otimes\left(\Gamma^{-1}(\Gamma(\langle P^{(l-1)}\rangle))^{\otimes(d_c-1)}\right)^{\otimes(d_v-1)}.
\end{aligned}$$

The above formula can be easily generalized to the irregular code ensembles $\mathcal{C}^n(\lambda,\rho)$:

$$\langle P^{(l)}\rangle=\langle P^{(0)}\rangle\otimes\lambda(\Gamma^{-1}(\rho(\Gamma(\langle P^{(l-1)}\rangle)))). \tag{8}$$

Remark: When $p=0.5$, the above formula becomes density evolution for decoding coset codes over channel $P_{Y|X}$ [26], which further reduces to (1) when $P_{Y|X}$ is output-symmetric.

III. Source-Channel Equivalence

It is easy to see that the density evolution formula (1) in channel coding and density evolution formula (8) in SW coding are almost identical; the only difference lies in the definitions of $P^{(l)}$ and $\langle P^{(l)}\rangle$. So virtually all results regarding channel coding density evolution (e.g., the stability condition) also hold in the SW coding case. But in this paper instead of restating results from channel coding theory in the SW setting, we shall mainly use the similarity in two density evolution formulas to establish connections between SW coding and channel coding.

In both density evolution formulas, the source or channel statistics come in only through the initial message distribution; all the remaining operations depend only on the degree distribution. So for a fixed degree distribution pair (λ,ρ), if $P^{(0)}=\langle P^{(0)}\rangle$, then two density evolutions are completely identical, i.e., we have $P^{(l)}=\langle P^{(l)}\rangle$ for all l. So a natural question is: For a given SW initial message distribution $\langle P^{(0)}\rangle$, does there exist a BIOS channel whose initial message distribution $P^{(0)}$ is the same as $\langle P^{(0)}\rangle$? Clearly, such a BIOS channel, if exists, is the equivalent channel for the joint distribution P_{XY}.

We now proceed to answer this question.

Definition 2 ([25], Definition 1): We call a distribution Q symmetric if

$$\int h(m)Q(dm)=\int e^{-m}h(-m)Q(dm)$$

for any function $h(\cdot)$ for which the integral exists.

The concept of symmetric distribution was originated in the context of channel coding. Specifically, the initial message distribution of a BIOS channel is always symmetric. The following lemma says the converse is also true.

Lemma 1: For any symmetric distribution Q, there exists a BIOS channel $P_{W|U}$ whose initial message distribution $P^{(0)}$ is equal to Q. Furthermore, the mapping between the set of symmetric distributions and the set of BIOS channels is bijective.

Proof: Suppose Q is a symmetric distribution with r probability mass points. By Definition 2, Q must be of the form

$$Q\left(\ln\frac{a_i}{a_{r-1-i}}\right)=a_i,\quad i=0,1,\cdots,r-1,$$

where $a_i\in(0,1]$ and $\sum_{i=0}^{r-1}a_i=1$. The corresponding BIOS channel $P_{W|U}$ is given by

$$P_{W|U}(i|0)=P_{W|U}(r-1-i|1)=a_i,\quad i=0,1,\cdots,r-1.$$

Clearly, each symmetric distribution with r probability mass points is associated with a unique BIOS channel with output alphabet size r up to different ways of labelling. If the initial message distribution of a binary-input r'-output channel is symmetric with r $(r<r')$ probability mass points, it implies that there exist w' and w'' such that $\frac{P_{W|U}(w'|0)}{P_{W|U}(w'|1)}=\frac{P_{W|U}(w''|0)}{P_{W|U}(w''|1)}$, i.e, w' and w'' can be combined to a single output symbol. So this channel can be eventually reduced to a BIOS channel with output alphabet size r. We shall view the original r'-output channel and resulting r-output channel as the same channel. In this sense, the mapping between the set of symmetric distributions and the set of BIOS channels is bijective. ■

Lemma 2: $\langle P^{(0)}\rangle$ is symmetric.

Proof: Note that $P^{(0)}(0)$, $P^{(0)}(1)$, and $\langle P^{(0)}\rangle$ all act on the random variable m given by

$$m=m_0=\ln\frac{P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}.$$

By a change of measure,

$$\begin{aligned}
&\int h(m)\langle P^{(0)}\rangle(dm)\\
=&\int h(m)P_X(0)P^{(0)}(0)(dm)+\\
&\int h(m)P_X(1)P^{(0)}(1)\circ I^{-1}(dm)\\
=&\ \mathbb{E}_{P_{Y|X}(Y|0)}\left[P_X(0)h\left(\ln\frac{P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}\right)\right]+\\
&\ \mathbb{E}_{P_{Y|X}(Y|1)}\left[P_X(1)h\left(\ln\frac{P_{X|Y}(1|Y)}{P_{X|Y}(0|Y)}\right)\right]\\
=&\ \mathbb{E}_{P_{Y|X}(Y|1)}\left[\frac{P_X(0)P_{Y|X}(Y|0)}{P_{Y|X}(Y|1)}h\left(\ln\frac{P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}\right)\right]+\\
&\ \mathbb{E}_{P_{Y|X}(Y|0)}\left[\frac{P_X(1)P_{Y|X}(Y|1)}{P_{Y|X}(Y|0)}h\left(\ln\frac{P_{X|Y}(1|Y)}{P_{X|Y}(0|Y)}\right)\right]\\
=&\ \mathbb{E}_{P_{Y|X}(Y|1)}\left[\frac{P_X(1)P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}h\left(\ln\frac{P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}\right)\right]+\\
&\ \mathbb{E}_{P_{Y|X}(Y|0)}\left[\frac{P_X(0)P_{X|Y}(1|Y)}{P_{X|Y}(0|Y)}h\left(\ln\frac{P_{X|Y}(1|Y)}{P_{X|Y}(0|Y)}\right)\right]
\end{aligned}$$

$$\begin{aligned}
&= \int e^{-m}h(-m)P_X(1)P^{(0)}(1)\circ I^{-1}(dm) + \\
&\quad \int e^{-m}h(-m)P_X(0)P^{(0)}(0)(dm) \\
&= \int e^{-m}h(-m)\langle P^{(0)}\rangle(dm).
\end{aligned}$$

This completes the proof. ■

Remark: The above argument can be easily generalized to show that $\langle P^{(l)}\rangle$ is symmetric for all l. The reason why $\langle P^{(l)}\rangle$ is symmetric even when there is no symmetry in the joint distribution P_{XY} is that the cosets used in SW coding have the symmetrizing effect. This should be contrasted with the case of using linear codes over asymmetric channels [27].

Theorem 2: For any joint distribution P_{XY} on $\mathcal{X}\times\mathcal{Y}$ ($\mathcal{X}=\{0,1\}$) with conditional entropy $H(X|Y)$, there exists a unique BIOS channel $P_{W|U}$ with capacity $C(P_{W|U})$ such that its initial message distribution $P^{(0)}$ is the same as the initial message distribution $\langle P^{(0)}\rangle$ induced by P_{XY}. Furthermore, we have $H(X|Y)+C(P_{W|U})=1$.

Proof: Since $m_0=\ln\frac{P_{X|Y}(0|Y)}{P_{X|Y}(1|Y)}$, we have

$$\begin{aligned}
P_{X|Y}(0|Y) &= \frac{e^{m_0}}{1+e^{m_0}}, \\
P_{X|Y}(1|Y) &= \frac{1}{1+e^{m_0}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}}\left[P_{Y|X}(y|0)H(X|Y=y)\right] \\
&= \mathbb{E}_{P^{(0)}(0)}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right],
\end{aligned}$$

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}}\left[P_{Y|X}(y|1)H(X|Y=y)\right] \\
&= \mathbb{E}_{P^{(0)}(1)}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right] \\
&= \mathbb{E}_{P^{(0)}(1)\circ I^{-1}}\left[H_b\left(\frac{e^{-m_0}}{1+e^{-m_0}}\right)\right] \\
&= \mathbb{E}_{P^{(0)}(1)\circ I^{-1}}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right], \qquad (9)
\end{aligned}$$

where (9) follows from the fact that $H_b(p)=H_b(1-p)$ for $p\in[0,1]$. Therefore,

$$\begin{aligned}
H(X|Y) &= P_X(0)\sum_{y\in\mathcal{Y}}\left[P_{Y|X}(y|0)H(X|Y=y)\right] + \\
&\quad P_X(1)\sum_{y\in\mathcal{Y}}\left[P_{Y|X}(y|1)H(X|Y=y)\right] \\
&= P_X(0)\mathbb{E}_{P^{(0)}(0)}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right] + \\
&\quad P_X(1)\mathbb{E}_{P^{(0)}(1)\circ I^{-1}}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right] \\
&= \mathbb{E}_{\langle P^{(0)}\rangle}\left[H_b\left(\frac{e^{m_0}}{1+e^{m_0}}\right)\right]
\end{aligned}$$

So if two distributions P_{XY} and $P_{X'Y'}$ induce the same initial message distribution $\langle P^{(0)}\rangle$, then we must have $H(X|Y)=H(X'|Y')$.

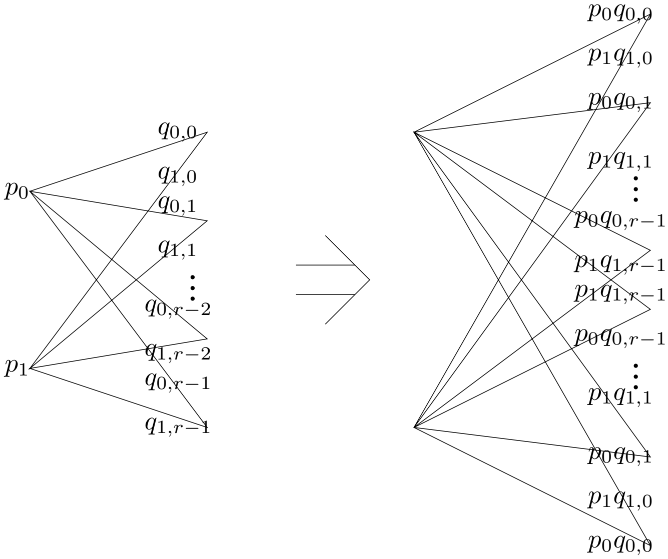

Fig. 2. Source-to-channel conversion.

For any joint distribution P_{XY}, by Lemmas 1 and 2, there exists a unique BIOS channel $P_{W|U}$ whose initial message distribution is the same as the one induced by P_{XY}. We can also view the initial message distribution associated with channel $P_{W|U}$ as the one induced by joint distribution P_{UW} where $P_U(0)=P_U(1)=0.5$. By the argument in the previous paragraph, we must have $H(X|Y)=H(U|W)$. Note that $C(P_{W|U})=1-H(U|W)$. Therefore, the proof is complete. ■

Given the initial message distribution $\langle P^{(0)}\rangle$ induced by the joint distribution P_{XY}, one can write down the corresponding BIOS channel $P_{W|U}$ explicitly (cf. the proof of Lemma 1). Actually it is possible to directly convert P_{XY} to $P_{W|U}$ without computing $\langle P^{(0)}\rangle$. This conversion is given in Fig. 2. It should be noted that although a joint distribution P_{XY} with $|\mathcal{X}|=2$ and $|\mathcal{Y}|=r$ can always be converted into a BIOS channel with output alphabet size $2r$, some output symbols of that channel might be equivalent[5] and thus can be combined. With the equivalent output symbols all combined, the output alphabet size of the resulting BIOS channel should be equal to the number of probability mass points of $\langle P^{(0)}\rangle$. In particular, if $P_X(0)=P_X(1)=0.5$ and $P_{Y|X}$ is output-symmetric, then $P_{W|U}$ degenerates to $P_{Y|X}$.

It is clear now that each joint distribution P_{XY} is associated with a unique initial message distribution, and thus a unique BIOS channel, which is denoted by $\mathrm{Ch}(P_{XY})$. As it turns out, the mapping $\mathrm{Ch}(\cdot)$ is not invertible. This leads to the following definition.

Definition 3 (Equivalence): Two joint distributions, P_{XY} and $P_{X'Y'}$, are equivalent if they induce the same initial message distribution $\langle P^{(0)}\rangle$ (i.e., if $\mathrm{Ch}(P_{XY})=\mathrm{Ch}(P_{X'Y'})$).

[5] For a BIOS channel $P_{W|U}$, we say two channel output symbols w' and w'' are equivalent if $\frac{P_{W|U}(w'|0)}{P_{W|U}(w'|1)}=\frac{P_{W|U}(w''|0)}{P_{W|U}(w''|1)}$. Generally, for a joint distribution P_{XY}, we say y' and y'' are equivalent if $\frac{P_{X|Y}(0|y')}{P_{X|Y}(1|y')}=\frac{P_{X|Y}(0|y'')}{P_{X|Y}(1|y'')}$. If $P_X(0)=P_X(1)=0.5$, this definition reduces to that in the channel case. Here "equivalent" means that the *a posteriori* distributions of X given $Y=y'$ and $Y=y''$ are the same.

Remark: Equivalent joint distributions are not distinguishable under density evolution.

IV. Conclusion

We have studied the problem of designing SW codes by leveraging its connection to designing LDPC channel codes. Specifically we have shown that, under density evolution, each SW coding problem is equivalent to a channel coding problem for a binary-input output-symmetric channel. Note that this channel is often different from the channel between the source and the side information in the original SW coding problem. This is in sharp contrast to the practice in the existing works where the two channels are assumed the same. It should be emphasized that the connection between SW coding and channel coding depends critically on the type of codes and decoding methods, and therefore, should be used with great caution.

Acknowledgment

The authors would like to acknowledge the anonymous reviewers and the associate editor, Dr. Alajaji, for their constructive comments that helped improve the presentation of the paper.

References

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

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

[3] J. K. Wolf, "Tribute to David Slepian," *IEEE Inform. Theory Society Newsletter*, vol. 58, Mar. 2008.

[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 Commun. Lett.*, vol. 5, pp. 417-419, Oct. 2001.

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

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

[10] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, 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," in *Proc. IEEE Data Compression Conf. (DCC)*, Mar. 2004.

[12] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," *IEEE Trans. Signal Processing*, vol. 53, pp. 3269-3281, Aug. 2005.

[13] 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.

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

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

[16] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 23, pp. 371-377, 1977.

[17] J. Li, Z. Tu, and R. S. Blum, "Slepian-Wolf coding for nonuniform sources using Turbo codes," in *Proc. IEEE Data Compression Conf. (DCC)*, Mar. 2004, pp. 312-321.

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

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

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

[21] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Practical loss-resilient codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569-584, Feb. 2001.

[22] D. MacKay, S. Wilson, and M. Davey, "Comparison of constructions of irregular Gallager codes," *IEEE Trans. Commun.*, vol. 47, pp. 1449-1454, Oct. 1999.

[23] 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.

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

[25] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619-637, Feb. 2001.

[26] 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, July 2003.

[27] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4216-4236, Dec. 2005.