Polar Coding for the Slepian-Wolf Problem Based on Monotone Chain Rules

Erdal Arıkan

Bilkent University, Ankara, Turkey

Abstract—**We give a polar coding scheme that achieves the full admissible rate region in the Slepian-Wolf problem without time-sharing. The method is based on a source polarization result using monotone chain rule expansions.**

Index Terms—**Monotone chain rules, polar codes, Slepian-Wolf problem, source polarization.**

I. INTRODUCTION

Consider a memoryless source with generic variables $(X, Y) \sim P_{X,Y}$ where $P_{X,Y}$ is a fixed but arbitrary probability distribution on $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Let (X^N, Y^N) denote N successive outputs of this source, $X^N = (X_1, \ldots, X_N)$, $Y^N = (Y_1, \ldots, Y_N)$. This paper considers the Slepian-Wolf problem for this source. As usual, the coding system consists of two encoders and one decoder. For a specified rate pair (R_1, R_2), encoder 1 observes X^N and encodes it into a codeword of length $\lfloor NR_1 \rfloor$ bits; encoder 2 observes Y^N and encodes it into a codeword of length $\lfloor NR_2 \rfloor$ bits. The decoder observes the two codewords and is expected to recover (X^N, Y^N) with small probability of error. The Slepian-Wolf result [1] states that this is possible if (R_1, R_2) falls strictly inside the Slepian-Wolf rate region defined as $\mathcal{R}_{SW} = \{(R_x, R_y) : R_x \geq H(X|Y), R_y \geq H(Y|X), R_x + R_y \geq H(X, Y)\}$. The subset of $\mathcal{R}_{SW}$ consisting of points for which $R_x + R_y = H(X, Y)$ is referred to as the *dominant face* (of the rate region); and the points $(R_x, R_y) = (H(X), H(Y|X))$ and $(R_x, R_y) = (H(X|Y), H(Y))$ are referred to as the *corner* points.

Polar coding for the above Slepian-Wolf problem was first considered by Hussami et al [2] (see also Korada [3]) who showed that the corner points of $\mathcal{R}_{SW}$ could be achieved by polar codes for the special case where P_X and P_Y are uniform on $\{0, 1\}$. In [4], this result was proved without any restrictions on P_X and P_Y. These results showed that polar codes could achieve the entire region $\mathcal{R}_{SW}$ by time-sharing between two codes designed for the corner points.

This paper is concerned with the question of whether polar codes can achieve $\mathcal{R}_{SW}$ without aid from time-sharing. This question is motivated by the fact that there are random-coding methods, such as Cover's "binning" method [5], that do not require time-sharing to achieve $\mathcal{R}_{SW}$. Thus, the question is important for understanding the power of polar coding relative to other coding methods both as a proof method and also for practical applications.

In fact, such questions on the relative power of polar coding first arose in the context of the multiple access channel (MAC), which is the dual of the Slepian-Wolf problem. In [6], Şaşoğlu

et al described a polar coding scheme for the MAC that did not use time-sharing and yet was able to achieve some interior (*i.e.*, non-corner) points of the dominant face of the MAC capacity region. The method in [6] was based on "joint polarization" for the MAC and it produced a multitude of extreme channels, revealing a novel aspect of polarization in the multi-terminal case. Abbe and Telatar refined and extended the joint polarization approach in [7], [8]. Meanwhile, on the Slepian-Wolf front, the joint polarization approach was formulated in [9]. In [10], Abbe gave a unified treatment of joint polarization for the MAC and Slepian-Wolf problems using "matrix polarization." The question of whether joint polarization *alone* could achieve the entire achievable rate regions for the MAC or the Slepian-Wolf problems remained unsolved until recently when Şaşoğlu [11] answered the question in the negative by giving counter-examples. This was a set-back for the polarization approach.

In this paper, we consider polarization in a broader setting and give a polar coding method that achieves $\mathcal{R}_{SW}$ without time-sharing. In this broader setting joint polarization appears as a special instance of a general approach. The main idea of our approach is described in the next section.

II. CHAIN RULES AND POLAR CODES

Consider a source block (X^N, Y^N) as above. Suppose $N = 2^n$ for some $n \geq 1$ and define

$$U^N = X^N G_N, \qquad V^N = Y^N G_N \qquad (1)$$

where G_N is the polar transform defined as

$$G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} B_N \qquad (2)$$

where the exponent denotes the nth Kronecker power and B_N is the "bit-reversal" permutation (see [12]). Since $(X^N, Y^N) \to (U^N, V^N)$ is a one-to-one mapping, we have

$$H(U^N, V^N) = H(X^N, Y^N) = NH(X, Y), \qquad (3)$$

which states that entropy is conserved. Polar codes can be obtained from (3) by various chain rule expansions of $H(U^N, V^N)$. To construct a polar code that achieves a corner point of $\mathcal{R}_{SW}$, one expands $H(U^N, V^N)$ as

$$\sum_{i=1}^{N} H(U_i|U^{i-1}) + \sum_{j=1}^{N} H(V_j|V^{j-1}, U^N) \qquad (4)$$

and shows that the entropy terms polarize to 0 or 1 as N increases.

In the joint polarization approach mentioned above, one uses the expansion

$$\sum_{i=1}^{N} H(U_i, V_i | U^{i-1}, V^{i-1}), \qquad (5)$$

and proves that the entropy terms in (5) polarize to 0, 1, or 2. Actually, to construct a specific polar code, one needs to expand (5) further, for example, as

$$\sum_{i=1}^{N} \left[H(U_i | U^{i-1}, V^{i-1}) + H(V_i | U^i, V^{i-1}) \right], \qquad (6)$$

and show that the entropy terms in (6) converge to 0 or 1. By using the degrees of freedom in expanding (5) into an expansion of type (6), one obtains polar codes achieving various rates on the dominant face of $\mathcal{R}_{SW}$ directly (without time-sharing). However, as shown in [11], this approach cannot achieve the entire dominant face in general.

It is clear that there are many other ways in which the total entropy $H(U^N, V^N)$ can be expanded into a sum of individual entropy terms, suggesting that there may exist many more polar codes, again raising the hope that the entire dominant face may be achievable by polar coding. This is the idea pursued in this paper.

III. MONOTONE CHAIN RULES

We call a chain rule expansion of $U^N V^N$ *monotone* w.r.t. U^N if the expansion is of the form

$$\sum_{i=1}^{2N} H(S_i | S^{i-1}) \qquad (7)$$

where $S^{2N} = (S_1, \ldots, S_{2N})$ is a permutation of $U^N V^N$ such that the permutation preserves the relative order of the elements of U^N. We define the monotonicity of a chain rule w.r.t. V^N similarly. A chain rule for $U^N V^N$ is said to be monotone if it is monotone w.r.t. both U^N and V^N. The expansions (4) and (6) are examples of monotone chain rules. The expansion $H(U_2) + H(U_1 | U_2) + H(V_1 | U_1, U_2) + H(V_2 | U_1, U_2, V_1)$ is monotone in V^2 but not in U^2.

We use diagrams, as in Fig. 1, to represent monotone chain rules, and refer to them briefly as "chain rule diagrams." Each directed path in Fig. 1, from $\emptyset$ to $U^4 V^4$, corresponds to a monotone chain rule on $H(U^4, V^4)$. For example, the "corner-point" path that goes from $\emptyset$ horizontally to U^4 and then vertically down to $U^4 V^4$ corresponds to the expansion (4). The "staircase" path $(\emptyset, U^1, U^1 V^1, U^2 V^1, U^2 V^2, U^3 V^2, U^3 V^3, U^4 V^3, U^4 V^4)$ corresponds to (6).

A label $U^i V^j$ attached to a node in a chain rule diagram designates the known variables when, and if, a chain rule visits that node; the entropy $H(U^i, V^j)$ is used to measure the amount of that knowledge. The edge connecting node $U^{i-1} V^j$ to node $U^i V^j$ is associated with the variable U_i and carries $H(U_i | U^{i-1}, V^j)$ units of incremental knowledge. Likewise, the edge connecting two vertically adjacent nodes
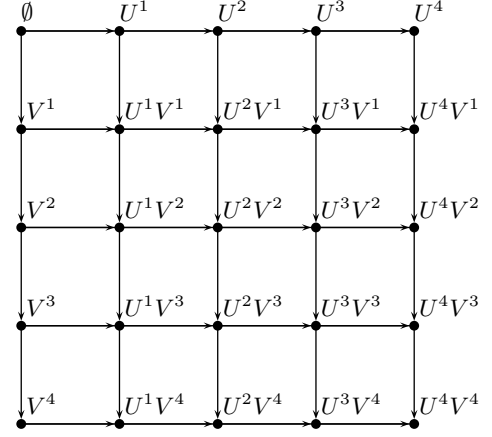

Fig. 1. Diagram for representing monotone chain rules on $H(U^4, V^4)$.

$U^i V^{j-1}$ and $U^i V^j$ is associated with V_j and carries an incremental knowledge of $H(V_j | U^i, V^{j-1})$ units. There is a path-independence property associated with states of knowledge in chain rule diagrams in the sense that the accumulated knowledge $H(U^i, V^j)$ at a node $U^i V^j$ is the sum of the conditional entropy terms along *any* path from $\emptyset$ to $U^i V^j$. In this sense, the entropy values assigned to the nodes form a *potential* function. Investigation of the properties of this potential function is left for future work. Here, we just note an elementary monotonicity property that may be useful for such studies.

Proposition 1. *The conditional entropy terms associated with vertical edges in the chain rule diagram for $U^N V^N$ are monotone in the sense that, for any fixed $1 \leq j \leq N$, $H(V_j | U^{i-1}, V^{j-1}) \geq H(V_j | U^i, V^{j-1})$ for all $1 \leq i \leq N$. Likewise, for any fixed $1 \leq i \leq N$, $H(U_i | U^{i-1}, V^{j-1}) \geq H(U_i | U^{i-1}, V^j)$ for all $1 \leq j \leq N$.*

A. Paths, rates

The chain rule diagram for $U^N V^N$ contains $\binom{2N}{N}$ paths from the initial node $\emptyset$ to the final node $U^N V^N$. We identify each path in the diagram by a string $b^{2N} = b_1 b_2 \cdots b_{2N}$ where b_i is 0 if the ith move along the path is in the horizontal direction and 1 otherwise. For instance, the corner-point path in Fig. 1 that goes from $\emptyset$ to U^4 then to $U^4 V^4$ has the label 00001111. The label 01010101 designates the staircase path of expansion (6).

Let $S^{2N} = (S_1, \ldots, S_{2N})$ denote the *edge variables* along a given path b^{2N}. For example, for $b^8 = 01010101$, the edge variables are $S^8 = (U_1, V_1, U_2, V_2, U_3, V_3, U_4, V_4)$.

For any given path b^{2N} with edge variables S^{2N}, we define a pair of rates

$$R_1 = \frac{1}{N} \sum_{i:b_i=0} H(S_i | S^{i-1}) \qquad (8)$$

and

$$R_2 = \frac{1}{N} \sum_{i:b_i=1} H(S_i | S^{i-1}). \qquad (9)$$

The rate R_1 (R_2) is the sum of the conditional entropy terms on the horizontal (vertical) edges in the path, normalized by N. For $b^8 = 01010101$, the rate R_1 is given by

$$R_1 = \frac{1}{4}\left[H(U_1) + H(U_2|U^1, V^1) + \right. $$
$$\left. H(U_3|U^2, V^2) + H(U_4|U^3, V^3)\right]. \quad (10)$$

Clearly, for any path for $U^N V^N$ the rate pair (R_1, R_2) satisfies

$$R_1 \geq \frac{1}{N}H(U^N|V^N), \quad R_2 \geq \frac{1}{N}H(V^N|U^N),$$
$$R_1 + R_2 = \frac{1}{N}H(U^N, V^N).$$

Stated in terms of the original source variables, these inequalities take the following form.

Proposition 2. *Let $U^N V^N$ be obtained from a memoryless source $X^N Y^N$ by (1). Then, the rate pair (R_1, R_2) for any monotone chain rule expansion of $U^N V^N$ satisfies*

$$R_1 \geq H(X|Y), \quad R_2 \geq H(Y|X), \quad R_1 + R_2 = H(X, Y).$$

The first inequality is satisfied with equality for the path $1^N 0^N$, and the second inequality is satisfied with equality for $0^N 1^N$.

This follows easily from the fact that the transform (1) is one-to-one. Thus, the rate pairs (R_1, R_2) over the class of monotone chain rules lie on the dominant face of the region $\mathcal{R}_{SW}$, spanning its two end-points. The next question we address is whether the rate pairs from this class form a dense subset of the dominant face.

B. Continuity of rates and approximations

Let b^{2N} and $\tilde{b}^{2N}$ be any two paths in the chain rule diagram for $U^N V^N$, with rate pairs (R_1, R_2) and $(\tilde{R}_1, \tilde{R}_2)$, respectively. We define the distance between b^{2N} and $\tilde{b}^{2N}$ as

$$d(b^{2N}, \tilde{b}^{2N}) = |R_1 - \tilde{R}_1|. \quad (11)$$

Note that since $R_1 + R_2 = \tilde{R}_1 + \tilde{R}_2 = H(X, Y)$, this distance is also given by $|R_2 - \tilde{R}_2|$.

We now seek a combinatorial notion of neighborhood among paths that is consistent with the above notion of distance. It is tempting to define two paths as neighbors if they differ by a transposition; however, this does not quite work here. For example, the path $b^8 = 01010011$ has a rate R_1 given by (10) while the path $\tilde{b}^8 = 11010010$, which differs from b^8 by a single transposition, has a rate given by

$$\tilde{R}_1 = \frac{1}{4}\left[H(U_1|V^2) + H(U_2|U^1 V^3) + \right.$$
$$\left. H(U_3|U^2 V^3) + H(U_4|U^3 V^4)\right].$$

It is not clear if $|R_1 - \tilde{R}_1|$ is small. If we restrict the class of transpositions as follows, we obtain a notion of neighborhood which serves our purposes.

Let two paths $\tilde{b}^{2N}$ and b^{2N} be *neighbors* if $\tilde{b}^{2N}$ can be obtained from b^{2N} by transposing b_i with b_j for some $i < j$ such that (i) $b_i \neq b_j$ and (ii) the substring $b_{i+1}b_{i+2}\cdots b_{j-1}$ bracketed by the transposed elements is either a string of all 0s

or all 1s. For instance, 10000111 and 00001111 are neighbors but 01001011 and 00001111 are not. Note that a path cannot be a neighbor of itself according to this definition.

Proposition 3. *For paths b^{2N} and $\tilde{b}^{2N}$ that are neighbors,*

$$d(b^{2N}, \tilde{b}^{2N}) \leq 1/N. \quad (12)$$

Proof: Let b^{2N} be a path with edge variables S^{2N} and let $\tilde{b}^{2N}$ differ from b^{2N} by a transposition in the coordinates $i < j$. Assume that $b_i = 0$, $b_j = 1$, and that the bracketed string $b_{i+1}\cdots b_{j-1}$ is all 1s. Then, observe that $R_1 - \tilde{R}_1 = (1/N)[H(S_i|S^{i-1}) - H(S_i|S^{i-1}, S_j, S_{i+1}^{j-1})]$. It is clear that $R_1 - \tilde{R}_1 \geq 0$. Moreover, $R_1 - \tilde{R}_1 \leq (1/N)H(S_i|S^{i-1}) \leq 1/N$. Thus, $|R_1 - \tilde{R}_1| \leq 1/N$. This covers the case of b_i^j being equal to 01^{j-i}. There are three other possibilities for b_i^j, namely, $0^{j-i}1$, $1^{j-i}0$, and 10^{j-i}. These other cases can be treated similarly to the first by exchanging the roles of b^{2N} and $\tilde{b}^{2N}$ or by considering $R_2 - \tilde{R}_2$ or both. ∎

We now turn our attention to rate approximations. For this, we focus on the subset of paths $\mathcal{V}_{2N} \triangleq \{0^i 1^N 0^{N-i} : 0 \leq i \leq N\}$ that have only one vertical segment.

Theorem 1. *Let (R_x, R_y) be a given rate pair on the dominant face of the Slepian-Wolf rate region. For any given $\epsilon > 0$, there exists N and a chain rule b^{2N} on $U^N V^N$ such that b^{2N} belongs to the class $\mathcal{V}_{2N}$ and has a rate pair (R_1, R_2) satisfying*

$$|R_1 - R_x| \leq \epsilon \quad and \quad |R_2 - R_y| \leq \epsilon. \quad (13)$$

Proof: Fix $N > 1/\epsilon$. Let $(R_1(i), R_2(i))$ denote the rate pair for the path $0^i 1^N 0^{N-i}$, for $0 \leq i \leq N$. We have $R_1(0) = H(X|Y)$ and $R_1(N) = H(X)$. Also, $|R_1(i+1) - R_1(i)| \leq 1/N$ by Proposition 3. Thus, for any $R_x \in [H(X|Y), H(X)]$ there exists $0 \leq i \leq N$ such that $|R_1(i) - R_x| \leq 1/N$. For this i, we must also have that $|R_2(i) - R_y| \leq 1/N$. ∎

Theorem 1 shows that we can approximate arbitrary points on the dominant face of $\mathcal{R}_{SW}$ with paths from the class $\{\mathcal{V}_{2N} : N = 2^n, n \geq 1\}$. Clearly, other classes of paths could have been used (some more effectively) for rate approximations. The class $\{\mathcal{V}_{2N}\}$ has the advantage of being simple.

C. Path scaling and polarization

Although we have found a way of approximating rates, the polarization issue has not yet been addressed. Here, we introduce an operation on paths that achieves polarization while keeping the rate approximation intact.

For any path $b^{2N} = b_1 b_2 \cdots b_{2N}$ representing a monotone chain rule for $U^N V^N$ and any integer $k = 2^m$, let kb^{2N} denote

$$\underbrace{b_1 \cdots b_1}_{k} \underbrace{b_2 \cdots b_2}_{k} \cdots \cdots \underbrace{b_{2N} \cdots b_{2N}}_{k},$$

which represents a monotone chain rule for $U^{kN} V^{kN}$. This operation is a geometric scaling operation in the sense that it preserves the "shape" of the original path. In particular, if b^{2N} belongs to the class $\mathcal{V}_{2N}$ then kb^{2N} belongs to $\mathcal{V}_{2kN}$.

Fix a path b^{2N} for $U^N V^N$ and consider the path $2b^{2N}$ for $U^{2N} V^{2N}$. Let S^{2N} and T^{4N} denote the edge variables for

b^{2N} and $2b^{2N}$, respectively. Let $\tilde{S}^{2N}$ be an independent copy of S^{2N}. The transformation (1) may be viewed as a mapping from the pair of random vectors $(S^{2N}, \tilde{S}^{2N})$ to T^{4N} with

$$T_{2i-1} = S_i \oplus \tilde{S}_i, \quad T_{2i} = S_i, \quad i = 1, \ldots, 2N. \quad (14)$$

This gives the following relationship between the entropies

$$
\begin{aligned}
H(T_{2i-1}&|T^{2i-2}) + H(T_{2i}|T^{2i-1}) \\
&= H(T_{2i-1}, T_{2i}|T^{2i-2}) \\
&= H(S_i \oplus \tilde{S}_i, S_i|S^{i-1} \oplus \tilde{S}^{i-1}, S^{i-1}) \\
&= H(\tilde{S}_i, S_i|\tilde{S}^{i-1}, S^{i-1}) \\
&= 2H(S_i|S^{i-1}). \quad (15)
\end{aligned}
$$

This may be interpreted as a local conservation law for conditional entropies under path scaling. As a corollary, the path rates (R_1, R_2) are preserved under path scaling.

Proposition 4. *Let b^{2N} be a fixed path. Let (R_1, R_2) be the rate pair for b^{2N}. Then, for any $m \geq 1$, (R_1, R_2) is also the rate pair for the path $2^m b^{2N}$.*

Another aspect of path scaling is polarization:

$$
\begin{aligned}
H(T_{2i}|T^{2i-1}) &= H(S_i|T^{2i-2}, T_{2i-1}) \\
&\leq H(S_i|T^{2i-2}) \\
&\leq H(S_i \oplus \tilde{S}_i|T^{2i-2}) \\
&= H(T_{2i-1}|T^{2i-2}) \quad (16)
\end{aligned}
$$

where there is equality if and only if $H(T_{2i}|T^{2i-1})$ equals 0 or 1. Thus,

$$H(T_{2i}|T^{2i-1}) \leq H(S_i|S^{i-1}) \leq H(T_{2i-1}|T^{2i-2}) \quad (17)$$

with equality if and only if $H(S_i|S^{i-1})$ equals 0 or 1.

We can keep doubling (scaling by two) the paths to enhance polarization. Asymptotically, we obtain the following result.

Theorem 2. *Let $(X, Y) \sim P_{X,Y}$ be an arbitrary memoryless source over the alphabet $\mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Consider the setting defined by equations (1) and (2). Fix $N_0 = 2^{n_0}$ for some $n_0 \geq 1$. Fix a path b^{2N_0} for $U^{N_0} V^{N_0}$. Let (R_1, R_2) be the rate pair for b^{2N_0}. Let $N = 2^m N_0$ for $m \geq 1$ and let T^{2N} be the edge variables for $2^m b^{2N_0}$. Then, for any given $\delta > 0$, as m goes to infinity, we have*

$$\frac{1}{2N}\left|\{1 \leq i \leq 2N : \delta < H(T_i|T^{i-1}) < 1 - \delta\}\right| \to 0, \quad (18)$$

$$\frac{|A_1(\delta)|}{N} \to R_1 \quad and \quad \frac{|A_2(\delta)|}{N} \to R_2, \quad (19)$$

where $A_j(\delta) = \{1 \leq i \leq 2N : b_i = j, \ H(T_i|T^{i-1}) > \delta\}$ for $j = 0, 1$. Furthermore, this statement remains true even if δ is allowed to go to zero as a function of N as $\delta = O(2^{-N^\beta})$, where β is fixed as any positive number less than 1/2.

We omit the proof of this theorem due to space limitations and also because it follows by standard ideas presented in detail elsewhere. We just note that the first step of the proof is to set up a martingale for the conditional entropy terms using

the conservation law (15). One may then use the approach taken in [4] which uses an auxiliary supermartingale based on the source Bhattacharyya parameters; alternatively, one may use the direct approach by Şaşoğlu [11, Lemma 2.1] in which only the main martingale is used. To prove the exponential convergence claim of the theorem one may use the method presented in [13].

To summarize, this subsection has shown that one can achieve rate-approximation and polarization without leaving the class of paths $\{\mathcal{V}_{2N} : N = 2^n, n \geq 1\}$.

IV. SLEPIAN-WOLF CODING

We now combine the above results to give a polar coding scheme for the Slepian-Wolf problem. The polar codes considered here are defined by two parameters (b^{2N}, δ) where b^{2N} is a monotone chain rule for $U^N V^N$ and $\delta > 0$ is a threshold parameter.

A. Encoding

Given a source realization (x^N, y^N), encoders 1 and 2 compute $u^N = x^N G_N$ and $v^N = y^N G_N$, respectively, as defined by equations (1) and (2). The realizations u^N and v^N define a realization t^{2N} of the edge variables T^{2N} associated with b^{2N}. Encoder 1 possesses $u^N = (t_i : b_i = 0)$ and transmits the variables $(t_i : i \in A_1(\delta))$, while encoder 2 possesses $v^N = (t_i : b_i = 1)$ and transmits $(t_i : i \in A_2(\delta))$. The rates for this scheme are given by $|A_1(\delta)|/N$ for user 1 and $|A_2(\delta)|/N$ for user 2.

B. Decoding

The decoder receives the variables $(t_i : i \in A(\delta))$ where $A(\delta) = A_1(\delta) \cup A_2(\delta)$ and wishes to reconstruct the missing variables $(t_i : i \notin A(\delta))$. For this task, we consider a *successive cancellation* (SC) decoder, as in [12] and [4]. The SC decoder enters the ith step of decoding with the decisions $\hat{t}^{i-1}$ from previous steps and sets the current decision as $\hat{t}_i = 0$ if $\Pr(T_i = 0|T^{i-1} = \hat{t}^{i-1})$ is greater than $\Pr(T_i = 1|T^{i-1} = \hat{t}^{i-1})$ and as $\hat{t}_i = 1$ otherwise. If $i \in A(\delta)$, the decoder overrides this rule by setting $\hat{t}_i = t_i$ since in that case the decoder already knows the correct value of t_i.

Once the estimate $\hat{t}^{2N}$ of t^{2N} is obtained, the decoder sets $\hat{u}_N = (\hat{t}_i : b_i = 0)$ and $\hat{v}^N = (\hat{t}_i : b_i = 1)$, and calculates $\hat{x}^N = \hat{u}^N(G_N)^{-1}$ and $\hat{y}^N = \hat{v}^N(G_N)^{-1}$, to obtain the estimates of x^N and y^N, respectively. Note that for the mapping G_N here, the inverse of G_N is itself, so this final step is just another encoding operation.

C. Performance

The performance of the above coding scheme is measured by the probability of frame error, defined as $P_e \triangleq \Pr[(\hat{X}^N, \hat{Y}^N) \neq (X^N, Y^N)]$. Equivalent expressions for the frame error probability are $P_e = \Pr[(\hat{U}^N, \hat{V}^N) \neq (U^N, V^N)]$ and $P_e = \Pr(\hat{T}^N \neq T^N)$. By the "genie-bound" for SC decoders (see, e.g., [12]), the frame error can be bounded as $P_e \leq \sum_{i \notin A(\delta)} \Pr(\hat{T}_i \neq T_i|\hat{T}^{i-1} = T^{i-1})$. Further, one has $\Pr(\hat{T}_i \neq T_i|\hat{T}^{i-1} = T^{i-1}) \leq Z(T_i|T^{i-1})$ where $Z(T_i|T^{i-1})$

is the source Bhattacharyya parameter defined in [4]. The parameter $Z(T_i|T^{i-1})$ is in turn bounded by $\sqrt{H(T_i|T^{i-1})}$ by Prop. 2 of [4]. Thus, $P_e \leq \sum_{i \notin A(\delta)} \sqrt{H(T_i|T^{i-1})} \leq (N - |A(\delta)|)\sqrt{\delta} \leq N\sqrt{\delta}$.

D. A polar coding theorem

Theorem 3. *Consider an arbitrary memoryless source* $(X, Y) \sim P_{X,Y}$ *over the alphabet* $\mathcal{X} \times \mathcal{Y}$ *with* $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. *Let* (R_x, R_y) *be a target point in the Slepian-Wolf rate region. Given any* $\epsilon > 0$ *and* $\beta < 1/2$*, there exists a polar coding scheme* (b^{2N}, δ) *such that (i) the path* b^{2N} *has the form* $0^i 1^N 0^{N-i}$ *for some* $0 \leq i \leq N$*, (ii) the threshold parameter satisfies* $\delta = O(2^{-N^\beta})$*, (iii) users 1 and 2 transmit at rates* $|A_1(\delta)|/N \leq R_x + \epsilon$ *and* $|A_2(\delta)|/N \leq R_y + \epsilon$*, respectively, and (iv) the probability of error under successive cancellation decoding satisfies* $P_e = O(2^{-\frac{1}{2}N^\beta})$.

Proof: We may assume without loss of generality that the target rate (R_x, R_y) lies on the dominant face of the rate region $\mathcal{R}_{SW}$. Theorem 1 guarantees the existence of a path b^{2N_0} in $\mathcal{V}_{2N_0}$ for which the rate pair (R_1, R_2) satisfies $R_1 \leq R_x + \epsilon/2$ and $R_2 \leq R_y + \epsilon/2$. We fix such a path. Theorem 2 guarantees that there exists a path $b^{2N} = 2^m b^{2N_0}$ for some $m \geq 1$ for which the sets $A_1(\delta)$ and $A_2(\delta)$ satisfy $|A_1(\delta)|/N \leq R_1 + \epsilon/2$ and $|A_2(\delta)|/N \leq R_2 + \epsilon/2$ with $\delta = O(2^{-N^\beta})$. The Slepian-Wolf code defined by the parameters (b^{2N}, δ) achieves the rates $|A_1(\delta)|/N \leq R_x + \epsilon$ and $|A_2(\delta)|/N \leq R_y + \epsilon$, and has a probability of error bounded by $P_e \leq N\sqrt{\delta} = O(2^{-\frac{1}{2}N^\beta})$. ∎

E. Complexity

The encoding operations in the above Slepian-Wolf polar coding scheme are the same as in the single-user case and have complexity $O(N \log N)$ as in that case [12].

It can be shown that the SC decoder here can be implemented in complexity $O(N \log N)$ as in the single user case. At each step of decoding, the SC decoder needs to calculate a probability of the form $P_{2N}(u^i, v^j) \triangleq \Pr(U^i = u^i, V^j = v^j)$, where the subscript $2N$ indicates the length of the code. Depending on whether i and j are odd or even, there is a different recursive formula to carry out this calculation. For example, $P_{2N}(u^{2i-1}, v^{2j-1})$ can be calculated as

$$\sum_{u_{2i}, v_{2j}} P_N(u_o^{2i} + u_e^{2i}, v_o^{2j} + v_e^{2j}) P_N(u_e^{2i}, v_e^{2j})$$

where u_o^{2i} and u_e^{2i} denote the sub-vectors consisting of odd-numbered and even-numbered coordinates of u^{2i}, respectively, and similarly for v_o^{2j} and v_e^{2j}. This reduction is continued until the desired probabilities can be computed from $P_{X,Y}$ directly.

Finally, for code construction, one needs to be able to compute entropy terms of the form $\{H(T_i|T^{i-1}) : 1 \leq i \leq 2N\}$ along a chosen path. This type of computation is necessary both for rate approximations and also for determining the sets $A_1(\delta)$ and $A_2(\delta)$. We conjecture that the density evolution method for ordinary polar coding developed in [14] and [15] can be adapted to this case, too, so as to compute these entropy terms with sufficient precision in complexity $O(N)$.

V. SUMMARY AND REMARKS

We considered polarization in the context of monotone chain rules, which is the largest class of chain rules that respects the natural decoding order defined by polarization. The main coding result has been the derivation of a polar coding scheme that achieves the Slepian-Wolf rate region without time-sharing.

Most of the discussion has been restricted to the subset of monotone chain rules represented by paths of the type $0^i 1^N 0^{N-i}$ for $0 \leq i \leq N$. On closer inspection, the use of such paths reminds one of the "source-splitting" approach to Slepian-Wolf coding developed by Rimoldi and Urbanke [16]. A path of the form $0^i 1^N 0^{N-i}$ has three segments, with each segment corresponding to a virtual source in the rate-splitting argument. In effect, the polar codes that we have constructed appear to operate at a corner point of a Slepian-Wolf rate region for three virtual sources.

Finally, we wish to note that, although not discussed explicitly, the results of this paper have duals in the context of coding for the MAC and yield capacity-achieving polar codes without time-sharing in that context.

REFERENCES

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

[2] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, (Seoul, South Korea), pp. 1488–1492, 28 June - 3 July 2009.

[3] S. B. Korada, *Polar codes for channel and source coding*. PhD thesis, EPFL, Lausanne, 2009.

[4] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, (Austin, TX), pp. 899-903, 13-18 June 2010.

[5] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, no. 2, pp. 226-228, Mar 1975

[6] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user binary-input multiple-access channel," in 2010 IEEE Information Theory Workshop (ITW), 2010, pp. 1-5.

[7] E. Abbe and E. Telatar, "MAC polar codes and matroids," in Information Theory and Applications Workshop (ITA), 2010, pp. 1-8.

[8] E. Abbe and E. Telatar, "Polar Codes for the m-User MAC," arXiv:1002.0777, Feb. 2010.

[9] E. Arıkan, "Source polarization," arXiv:1001.3087v1, Jan. 2010.

[10] E. Abbe, "Extracting randomness and dependencies via a matrix polarization," arXiv:1102.1247, Feb. 2011.

[11] E. Şaşoğlu, "Polar Coding Theorems for Discrete Systems," EPFL Thesis no 5219, Lausanne, Nov. 2011.

[12] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3051–3073, July 2009.

[13] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, (Seoul, South Korea), pp. 1493–1495, 28 June - 3 July 2009.

[14] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, (Seoul, South Korea), pp. 1496-1500, 28 June - 3 July 2009.

[15] I. Tal and A. Vardy, "How to Construct Polar Codes," arXiv:1105.6164v1 [cs.IT] 31 May 2011.

[16] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting," *Proc. 1997 IEEE Int. Symp. Inform. Theory*, (Ulm, Germany), p. 271, 29 Jun - 4 Jul 1997.