Polar Codes and Polar Lattices for the Heegard-Berger Problem

Jinwen Shi, Student Member, IEEE, Ling Liu, Deniz Gündüz, Senior Member, IEEE, and Cong Ling, Member, IEEE

Abstract—Explicit coding schemes are proposed to achieve the rate-distortion function of the Heegard-Berger problem using polar codes. Specifically, a nested polar code construction is employed to achieve the rate-distortion function for the doubly-symmetric binary sources when the side information may be absent. The nested structure contains two optimal polar codes for lossy source coding and channel coding, respectively. Moreover, a similar nested polar lattice construction is employed when the source and the side information are jointly Gaussian. The proposed polar lattice is constructed by nesting a quantization polar lattice and a capacity-achieving polar lattice for the additive white Gaussian noise channel.

Index Terms—Heegard-Berger Problem, source coding, lattices.

I. INTRODUCTION

The well-known Wyner-Ziv problem is a lossy source coding problem in which a source sequence is to be reconstructed in the presence of correlated side information at the decoder [1]. An interesting question is whether reconstruction with a non-trivial distortion quality can still be obtained in the absence of the side information. The equivalent coding system contains two decoders, one with the side information, and the other without, as shown by Fig. 1.

In 1985, Heegard and Berger [2] characterized the rate-distortion function $R_{HB}(D_1, D_2)$ for this scenario, where $D_1$ is the distortion achieved without side information, and $D_2$ is the distortion achieved with it, and $R_{HB}(D_1, D_2)$ denotes the minimum rate required to achieve the distortion pair $(D_1, D_2)$. They also gave an explicit expression for the quadratic Gaussian case. Kerpez [3] provided upper and lower bounds on the Heegard-Berger rate-distortion function (HBRDF) for the binary case. Later, the explicit expression for $R_{HB}(D_1, D_2)$ in the binary case was derived in [4] together with the corresponding optimal test channel. Our goal in this paper is to propose explicit coding schemes that can achieve the HBRDF for binary and Gaussian distributions.

The Heegard-Berger problem is a generalization of the standard Wyner-Ziv problem, in which a source sequence is to be reproduced at the decoder within a certain distortion target, and the side information available at the decoder is not available at the encoder. A nested construction of polar codes is presented in [5] to achieve the binary Wyner-Ziv rate-distortion function. For Gaussian sources, a polar lattice to achieve both the standard and Wyner-Ziv rate-distortion functions is proposed in [6]. Different from the solutions of the Wyner-Ziv problem in [5] and [6], we need to consider the requirements for the two decoders jointly. Therefore, we make use of the low-fidelity reconstruction at Decoder 1, and combine it with the original source and the side information to form the nested structure that achieves the optimal distortion for Decoder 2.

The optimality of polar codes for the lossy compression of nonuniform sources is shown in [7]. We employ this scheme as part of our solution, since the optimal forward test channel may be asymmetric in the binary Heegard-Berger problem. Furthermore, it is shown in [8] that polar codes are optimal for general distributed hierarchical source coding problems. The Heegard-Berger problem can also be considered as a successive refinement problem. In this paper, we propose explicit coding schemes using polar codes and polar lattices to achieve the theoretical performance bound in the Heegard-Berger problem. Practical codes for the Gaussian Heegard-Berger problem are also developed in [9] which hybridize trellis and low-density parity-check codes. However, the optimality of this scheme to achieve the HBRDF is not shown in [9].

The contributions of this paper can be summarized as follows:

- We propose a nested construction of polar codes for the non-degenerate region of the binary Heegard-Berger problem, and prove that they achieve the HBRDF for doubly symmetric binary sources (DSBS). We consider the reconstruction of the source sequence at Decoder 1, i.e., the decoder without side information, denoted by $\hat{X}_1^{1:N}$ and the original source sequence $X_1^{1:N}$ as a combined source, and further combine this reconstruction $\hat{X}_1^{1:N}$ with the original side information $Y_1^{1:N}$ to obtain a combined side
information. By this argument, we obtain another nested construction of polar codes, which achieves the HB RDF of the entire non-degenerate region. In addition, we present an explicit coding scheme by using two-level polar codes to achieve the HB RDF whose forward test channel may be asymmetric. Finally, we prove that polar codes achieve an exponentially decaying block error probability, and the distortion spread between the average distortion and the target distortion of Decoder 2 is sub-exponentially decaying with an exponent on the order of $O\left(\sqrt{N}\right)$ for the binary Heegard-Berger problem.

- We then consider the Gaussian Heegard-Berger problem, and propose a polar lattice construction that consists of two nested polar lattices, one of which is additive white Gaussian noise (AWGN) capacity-achieving while the other is Gaussian rate-distortion function achieving. This construction is similar to the one proposed for the Gaussian Wyner-Ziv problem in [6]. However, in the Heegard-Berger problem setting, we need to treat the difference between the original source and its reconstruction at Decoder 1 as a new source, and the difference between the original side information and the reconstruction at Decoder 1 as a new side information. As a result, we can obtain an optimal test channel that connects the new source with the new side information by using additive Gaussian noises. According to this test channel, we can further construct two nested polar lattices that achieve the Gaussian HB RDF of the entire non-degenerate region.

**Organization:** The paper is organized as follows: Section II presents the background on binary and Gaussian Heegard-Berger problems. The construction of polar codes to achieve the HB RDF for DSBSs is investigated in Section III. In Section IV, the polar code construction for the Gaussian Heegard-Berger problem is addressed. The paper is concluded in Section V.

**Notation:** All random variables are denoted by capital letters, while sets are denoted by capital letters in calligraphic font. $P_X$ denotes the probability distribution of a random variable $X$ taking values in set $\mathcal{X}$. For two positive integers $i < j$, $x^{ij}$ denotes the vector $(x_i, \ldots, x_j)$, which represents the realizations of random variables $X^{ij}$. For a set $\mathcal{F}$ of positive integers, $x_{\mathcal{F}}$ denotes the subvector $\{x_i\}_{i \in \mathcal{F}}$. For the Gaussian case in Section IV, we construct polar lattices in multiple levels, in which $X_i$ denotes a random variable at level $i$, and $x_i^j$ its $i$-th realization. Then, $x^{ij}_\mathcal{F}$ denotes the vector $\{x_i^j, \ldots, x_j^j\}$, and $x_{\mathcal{F}}^j$ denotes the subvector $\{x_i^j\}_{i \in \mathcal{F}}$ at the $i$-th level. $\mathcal{F}_c$ and $|\mathcal{F}|$ denote the complement and cardinality of set $\mathcal{F}$, respectively. For a positive integer $N$, we define $\{N\} \triangleq \{1, \ldots, N\}$. $I\{x \in \mathcal{X}\}$ denotes the indicator function, which equals 1 if $x \in \mathcal{X}$ and 0 otherwise. Let $I(X;Y)$ denote the mutual information between $X$ and $Y$. In this paper, all logarithms are base two, and information is measured in bits.

II. PROBLEM STATEMENT

A. Heegard-Berger Problem

Let $(X,Y,P_{XY})$ be discrete memoryless sources (DMSs) characterized by the random variables $X$ and $Y$ with a generic joint distribution $P_{XY}$ over the finite alphabets $\mathcal{X}$ and $\mathcal{Y}$.

Figure 2. Illustration of the HB RDF regions for DSBS, where $d_c$ is the critical distortion.

**Definition 1.** An $(N,M,D_1,D_2)$ Heegard-Berger code for source $X$ with side information $Y$ consists of an encoder $f^{(N)}: \mathcal{X}^N \rightarrow \mathcal{M}$ and two decoders $g^{(N)}_1: \mathcal{M} \rightarrow X^N$; $g^{(N)}_2: \mathcal{M} \times Y^N \rightarrow X^N$, where $\hat{X}_1, \hat{X}_2$ are finite reconstruction alphabets, such that

$$
\mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} d \left( X^i, \hat{X}_i^i \right) \right] \leq D_i, \quad i = 1, 2,
$$

where $\mathbb{E}$ is the expectation operator, and $d(\cdot, \cdot) < \infty$ is a per-letter distortion measure. Specifically, we set $d(\cdot, \cdot)$ to be the Hamming distortion (i.e., $d(0,0) = d(1,1) = 0$, $d(0,1) = 1$) for binary sources, and the squared error distortion (i.e., $d(x,y) = (x-y)^2$) for Gaussian sources.

The rate $R$ is said to be $\{(D_1, D_2) - \text{achievable}\}$, if for every $\epsilon > 0$ and sufficiently large $N$ there exists an $(N,M,D_1+\epsilon,D_2+\epsilon)$ code with $R + \epsilon \geq \frac{1}{N} \log M$. The HB RDF, $R_{HB}(D_1, D_2)$, is defined as the infimum of $(D_1, D_2)$-achievable rates. A single-letter expression for $R_{HB}(D_1, D_2)$ is given in [2, Theorem 1] as

$$
R_{HB}(D_1, D_2) = \min_{(U_1,U_2) \in \mathcal{P}(D_1,D_2)} \{ I(X;U_1) + I(X;U_2|U_1,Y) \},
$$

where $\mathcal{P}(D_1,D_2)$ is the set of all auxiliary random variables $(U_1,U_2) \in \mathcal{U}_1 \times \mathcal{U}_2$ jointly distributed with the generic random variables $(X,Y)$, such that: i) $Y \leftrightarrow X \leftrightarrow (U_1,U_2)$ form a Markov chain; ii) $\mathcal{U}_1$ is independent of $|X|+2$ and $\mathcal{U}_2$ is independent of $\mathcal{X}$; iii) there exist functions $\varphi_1$ and $\varphi_2$ such that $\mathbb{E}[d(X,\varphi_1(U_1))] \leq D_1$ and $\mathbb{E}[d(X,\varphi_2(U_1,U_2,Y))] \leq D_2$.

B. Doubly Symmetric Binary Sources

Let $X$ be a binary DMS, i.e., $X \in \{0,1\}$, with uniform distribution. The binary side information is specified by $Y = X \oplus Z$, where $Z$ is an independent Bernoulli random variable with $P_Z(z=1) = p < 0.5$, and $\oplus$ denotes modulo two addition. The HB RDF for DSBSs can be characterized over four regions [3]. Region I $(0 \leq D_1 < 0.5$ and $0 \leq D_2 < \min(D_1,p))$ is a
non-degenerate region, and \(R_{HB}(D_1, D_2)\) is a function of \(D_1\) and \(D_2\); Region II (\(D_1 \geq 0.5\) and \(0 \leq D_2 \leq p\)) is a degenerate region as the Heegard-Berger problem boils down to the Wyner-Ziv problem for the second decoder; Region III (\(0 \leq D_1 \leq 0.5\) and \(D_2 \geq \min(D_1, p)\)) is also degenerate since the problem boils down to the standard lossy compression problem for the first decoder; Region IV (\(D_1 > 0.5\) and \(D_2 > p\)) can be trivially achieved without coding. These four regions are depicted in Fig. 2. Note that, the HBRDF in the degenerate Regions II and III can be achieved by using polar codes as described in [5]. Here we focus on the non-degenerate Region I.

First, we recall the function \(G(u) \triangleq h(p \ast u) - h(u)\) from [1], defined over the domain \(0 \leq u \leq 1\), where \(h(u)\) is the binary entropy function \(h(u) \triangleq -u \log u - (1 - u) \log (1 - u)\), and \(p \ast u\) is the binary convolution for \(0 \leq p, u \leq 1\), defined as \(p \ast u \triangleq p(1 - u) + u(1 - p)\).

The explicit calculation of HBRDF for DSBS in Region I is given in [4] as follows:

Define the function

\[
S_{D_1}(\alpha, \mu, \theta, \theta_1) \triangleq 1 - h(D_1 \ast p) + (\theta - \theta_1)G(\alpha) + \theta_1G(\mu) + (1 - \theta)G(\gamma),
\]

where

\[
\gamma \triangleq \begin{cases} 
\frac{D_1 - (\theta - \theta_1)(1 - \alpha) - \theta \mu}{1 - \theta} & \text{if } \theta \neq 1 \\
0.5 & \text{if } \theta = 1
\end{cases}
\]

on the domain \(0 \leq \theta_1 \leq \theta \leq 1\), \(0 \leq \alpha, \mu \leq p\), \(p \leq \gamma \leq 1 - p\).

The following theorem characterizes the HBRDF in Region I.

**Theorem 2.** [4, Theorem 2] For \(0 \leq D_1 < 0.5\) and \(0 \leq D_2 < \min(D_1, p)\), we have

\[
R_{HB}(D_1, D_2) = \min S_{D_1}(\alpha, \mu, \theta, \theta_1),
\]

where the minimization is over all \(\theta, \alpha, \mu, \theta_1\) variables that satisfy \(0 \leq \theta_1 \leq \theta \leq 1\), \(0 \leq \alpha, \mu \leq p\) and \((\theta - \theta_1)\alpha + \theta_1\mu + (1 - \theta)p = D_2\).

The corresponding forward test channel structure is also given in [4], reproduced in Table I. It constructs random variables with joint distribution \(P_{X, U_1, U_2}(x, u_1, u_2)\), which satisfy \(I(X; U_1) + I(X; U_2|U_1, Y) = S_{D_1}(\alpha, \mu, \theta, \theta_1)\).

Next, recall the definition of the critical distortion, \(d_c\), in the Wyner-Ziv problem for DSBS [1], for which \(
\frac{G(d_c)}{d_c - p} = G'(d_c)
\).

Then the following corollary from [4] specifies an explicit the characterization of HBRDF for DSBS in Region I-B in Fig. 2 specified by \(D_2 \leq \min(d_c, D_1)\) and \(D_1 \leq 0.5\).

**Corollary 3.** ([4, Corollary 2]) For distortion pairs \((D_1, D_2)\) satisfying \(D_1 \leq 0.5\) and \(D_2 \leq \min(d_c, D_1)\) (i.e., Region I-B in Fig. 2), we have

\[
R_{HB}(D_1, D_2) = 1 - h(D_1 \ast p) + G(D_2).
\]

From [4], the optimal forward test channel for Region I-B is given as a cascade of two binary symmetric channels (BSCs), as depicted in Fig. 3.

In Section III, we first propose a polar code design that achieves the HBRDF in Region I-B for DSBS. We then provide a general polar code construction achieving the HBRDF in the entire Region I.

**C. Gaussian Sources**

Suppose \(Y = X + Z\), where \(X\) and \(Z\) are independent (zero-mean) Gaussian random variables with variances \(\sigma_X^2\) and \(\sigma_Z^2\), respectively, i.e., \(X \sim N(0, \sigma_X^2)\) and \(Z \sim N(0, \sigma_Z^2)\). The explicit expression for \(R_{HB}(D_1, D_2)\) in this case is given in [2]. The optimal test channels are given by \(X = U_1 + Z_1\) and \(X = U_2 + Z_2\), where \(Z_1, Z_2\) are independent zero-mean Gaussian. We have \(Z_i \sim N(0, D_i), \, i = 1, 2\).

For \(D_1 \leq \sigma_X^2\) and \(D_2 \leq \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\), the problem degenerates into a standard lossy compression problem for Decoder 1, and the HBRDF is given by \(R_{HB}(D_1, D_2) = \frac{1}{2} \log \left(\frac{\sigma_X^2}{D_1}\right)\). For \(D_1 > \sigma_X^2\) and \(D_2 \leq \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\), the problem degenerates into a Wyner-Ziv coding problem for Decoder 2, and we have \(R_{HB}(D_1, D_2) = \frac{1}{2} \log \left(\frac{\sigma_Z^2}{D_2}\right)\). The region specified by \(D_1 > \sigma_X^2\) and \(D_2 \geq \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\) requires no coding. Polar lattice codes that meet the standard and Wyner-Ziv rate-distortion functions for Gaussian sources, introduced in [6], can be used to achieve the HBRDF in these degenerate regions. The only non-degenerate distortion region is specified by \(D_1 \leq \sigma_X^2\) and \(D_2 \leq \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\), and the HBRDF in this region is given by [2]:

\[
R_{HB}(D_1, D_2) = \frac{1}{2} \log \left(\frac{\sigma_X^2 \sigma_Z^2}{D_2 (D_1 + \sigma_Z^2)}\right).
\]

We will focus on the construction of polar lattice codes that achieve the HBRDF in (2) in Section IV.

**III. POLAR CODES FOR DSBS**

In this section, we present a construction of polar codes that achieves \(R_{HB}(D_1, D_2)\) for DSBS in Region I. First, we give a brief overview of polar codes.

Let \(G_2 \triangleq \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\), and define \(G_N \triangleq G_2^\otimes n\) as the generator matrix of polar codes with length \(N = 2^n\), where \(\otimes\) denotes
the Kronecker product. A polar code $C_N (\mathcal{F}, u_{\mathcal{F}})$ [5] is a linear code defined by

$$C_N (\mathcal{F}, u_{\mathcal{F}}) = \left\{ v^{1:N} G_N : v_{\mathcal{F}} = u_{\mathcal{F}}, v_{\mathcal{F}^c} \in \{0, 1\}^{|\mathcal{F}^c|} \right\},$$

for any $\mathcal{F} \subseteq [N]$ and $u_{\mathcal{F}} \in \{0, 1\}^{|\mathcal{F}|}$, where $\mathcal{F}$ is referred to as the frozen set. The code $C_N (\mathcal{F}, u_{\mathcal{F}})$ is constructed by fixing $u_{\mathcal{F}}$ and varying the values in $\mathcal{F}^c$. Moreover, the frozen set can be determined by the Bhattacharyya parameter [5]. For a binary memoryless asymmetric channel with input $X \in \{0, 1\}$ and output $Y \in \mathcal{Y}$, the Bhattacharyya parameter $Z$ is defined as

$$Z (X|Y) = \sum_y P_{X,Y} (0, y) P_{X,Y} (1, y).$$

A. Polar Code Construction for Region I-B

We observe from the proof of Theorem 2 in [4] that the auxiliary random variable $U_1$ can be considered as the output of a BSC with crossover probability $D_1$ and input $X$. Therefore, for Region I-B, the minimum rate for Decoder 1 to achieve the target distortion $D_1$ is $R_1 = I (U_1; X) = 1 - h (D_1)$. It is shown in [10, Theorem 3] that polar codes can achieve the rate-distortion function of binary symmetric sources. An explicit code construction is also provided in [10]. Considering the source sequence $X^{1:N}$ as $N$ independent and identically distributed (i.i.d.) copies of $X$, we know from [10, Theorem 3] that Decoder 1 can recover a reconstruction $X_{U_1}^{1:N}$ that is asymptotically close to $U_1^{1:N}$ as $N$ becomes sufficiently large. Therefore, we assume that both Decoder 1 and Decoder 2 can obtain $U_1^{1:N}$ in the following.

Decoder 2 observes the side information $Y^{1:N}$, in addition to $U_1^{1:N}$ that can be reconstructed using the same method as Decoder 1. Note that it is shown in [10, Lemma 11] that the difference between the exact quantization error $X_{U_1}^{1:N} \otimes U_1^{1:N}$ and an i.i.d. $Ber (D_1)$ distribution can be upper bounded, and decays with respect to the blocklength $N$. In other words, the exact quantization error is asymptotically close to an i.i.d. $Ber (D_1)$ distribution when $N$ is sufficiently large. Therefore, $X_{U_1}^{1:N} \otimes Z_{U_1}^{1:N}$ can be considered as the side information, which satisfies the same assumption as in the Wyner-Ziv setting as given by [10], as long as the blocklength $N$ of the polar codes employed by Decoder 1 is sufficiently large. Hence, both $Y^{1:N}$ and $U_1^{1:N}$ can be considered as side information for Decoder 2 to achieve distortion $D_2$. Therefore, the problem of Decoder 2 is similar to Wyner-Ziv coding except that the decoder observes extra side information.

Recall that achieving the Wyner-Ziv rate-distortion function using polar codes is based on the nested code structure proposed in [5]. Consider the Wyner-Ziv problem consisting of compressing a source $X^{1:N}$ in the presence of correlated side information $Y^{1:N}$ using polar codes, where $X$ and $Y$ are DSBS. The code $C_c$ with corresponding frozen set $\mathcal{F}_c$ is designed to be a good source code for distortion $D_2$. Further, the code $C_c$ with corresponding frozen set $\mathcal{F}_c$ is designed to be a good channel code for BSC($D_2 \ast \rho$). It has been shown in [5] that $\mathcal{F}_c \supseteq \mathcal{F}_s$, because the test channel BSC($D_2 \ast \rho$) is degraded with respect to BSC($D_2$). In this case, the encoder transmits to the decoder the sub-vector that belongs to the index set $\mathcal{F}_c \setminus \mathcal{F}_s$. The optimality of this scheme is proven in [5].

Similarly, the optimal rate-distortion performance for Decoder 2 in the Heegard-Berger problem can also be achieved by using nested polar codes. For $(U_1, U_2) \in \mathcal{P} (D_1, D_2)$, we have

$$I (X; U_2 | U_1, Y) = I (U_2; X, U_1, Y) - I (U_2; Y, U_1) = I (U_2; X, U_1) - I (U_2; Y, U_1).$$

The second equality holds since $Y \leftrightarrow X \leftrightarrow (U_1, U_2)$ form a Markov chain. Motivated by (3), the code $C_{c_2}$ with corresponding frozen set $\mathcal{F}_{c_2}$ is designed to be a good source code for the source pair $(X^{1:N}, U_1^{1:N})$ with reconstruction $U_2^{1:N}$. $T_s$ denotes the test channel for this source code. Additionally, the code $C_{c_2}$ with corresponding frozen set $\mathcal{F}_{c_2}$ is designed to be a good channel code for the test channel $T_s$ with input $U_1^{1:N}$ and output $(Y^{1:N}, U_1^{1:N})$.

According to [5, Lemma 4.7], in order to show the nested structure between $C_{c_2}$ and $C_{c_3}$, it is sufficient to show that $T_s$ is stochastically degraded with respect to $T_s$. We then give the definition on the channel degradation [5] as follows. Let $T_1 : \mathcal{U} \rightarrow \mathcal{Y}$ and $T_2 : \mathcal{U} \rightarrow \mathcal{X}$ be two discrete memoryless channels. We say that $T_1$ is stochastically degraded with respect to $T_2$, if there exists a discrete memoryless channel $T : \mathcal{X} \rightarrow \mathcal{Y}$ such that

$$P_{Y|X} (y|x) = \sum_{x \in \mathcal{X}} P_{X,Y} (x,y) P_{Y|X} (y|x).$$

Proposition 4. $T_{c_2} : U_2 \rightarrow (Y, U_1)$ is stochastically degraded with respect to $T_{c_3} : U_2 \rightarrow (X, U_1)$, if the random variables $(X, Y, U_1, U_2)$ agree with the forward test channel as shown in Fig. 3.

Proof: From the test channel structure in Fig. 3, $Y \leftrightarrow X \leftrightarrow U_2 \leftrightarrow U_1$ form a Markov chain. By definition, we have $P_{X,U_1,U_2} (x, u_1|u_2) = P_{X|U_2} (x|u_2) P_{U_1,U_2} (u_1|u_2)$. We also have

$$P_{Y,U_1,U_2} (y, u_1|u_2) = P_{Y|U_2} (y|u_2) P_{U_1,U_2} (u_1|u_2) = \sum_{x} P_{X,Y,U_2} (x,y,u_2) P_{U_1,U_2} (u_1|u_2) = \sum_{x} P_{X,U_1,U_2} (x,u_1,u_2) P_{Y|X} (y|x),$$

completing the proof.

Therefore, we can claim that $\mathcal{F}_{c_2} \supseteq \mathcal{F}_{c_2}$ by [5, Lemma 4.7], and rather than sending the entire vector that belongs to the index set $\mathcal{F}_{c_2}$, the encoder sends only the sub-vector that belongs to $\mathcal{F}_{c_2} \setminus \mathcal{F}_{c_2}$ to Decoder 2, since Decoder 2 can extract some information on $U_2^{1:N}$ from the available side information $(U_1^{1:N}, Y^{1:N})$. As a result, the polar code construction for the Heegard-Berger problem in Region I-B is given as follows:

**Encoding:** The encoder first applies lossy compression to source sequence $X^{1:N}$ with reconstruction $U_1^{1:N}$ and corresponding average distortion $D_1$. We construct the code $C_{s_1} = C_N (\mathcal{F}_{s_1}, \bar{\mathcal{G}}) = \left\{ w^{1:N} G_N : w_{\mathcal{F}_{s_1}} = \bar{\mathcal{G}}, w_{\mathcal{F}_{s_1}} \in \{0, 1\}^{|\mathcal{F}_{s_1}|} \right\}$, and the encoder transmits the compressed sequence $w_{\mathcal{F}_{s_1}}$.
to the decoders. The encoder is also able to recover $U_1^{1:N}$ from $C_s$. Next, the encoder applies lossy compression jointly for sources $(X^{1:N}, U_1^{1:N})$ with reconstruction $U_2^{1:N}$ and target distortion $D_2$ and $d(U_1, U_2) = \eta$, where $d(\cdot, \cdot)$ is the Hamming distortion function. We then construct $C_{s_2} = C_N(\mathcal{F}_{s_2}, \bar{0})$. Finally, the encoder applies channel coding to the symmetric test channel $T_\epsilon$ with input $U_2^{1:N}$ and output $(Y^{1:N}, U_1^{1:N})$. We derive $C_{s_2} = C_N(\mathcal{F}_{s_2}, u_{\mathcal{F}_{s_2}}(\bar{v}))$, where $u_{\mathcal{F}_{s_2}}(\bar{v})$ is defined by $u_{\mathcal{F}_{s_2}} = \bar{0}$ and $u_{\mathcal{F}_{s_2}} = \bar{v}$ for $\bar{v} \in \{0, 1\}^{\mathcal{F}_{s_2} \setminus \mathcal{F}_{s_2}}$. The encoder sends the sub-vector $u_{\mathcal{F}_{s_2}}$ to the decoders.

**Decoding:** Decoder 1 receives $w_{\mathcal{F}_{s_1}}$ and outputs the reconstruction sequence $u_1^{1:N} = w^{1:N}G_{\mathcal{N}}$. Decoder 2 receives $u_{\mathcal{F}_{s_2}}$, and hence, it can derive $u_{\mathcal{F}_{s_2}}$. Moreover, Decoder 2 can also recover $U_1^{1:N}$ from $w_{\mathcal{F}_{s_1}}$. Decoder 2 applies the successive cancellation (SC) decoding algorithm to obtain the codeword $U_1^{1:N}$ from the realizations of $(Y_1^{1:N}, U_1^{1:N})$.

Next we present the rates that can be achieved by the proposed scheme. From the polarization theorem for lossy source coding in [10], we know that reliable decoding at Decoder 1 will be achieved with high probability if

$$I(U_1; X) = 1 - h(D_1).$$

From the polarization theorems for source and channel coding [5], the code required for reliable decoding at Decoder 2 can be derived by

$$\frac{|\mathcal{F}_{s_2}| - |\mathcal{F}_{s_2}|}{N} \xrightarrow{N \to \infty} I(U_2; X, U_1) - I(U_2; Y, U_1)
= I(U_2; X) + I(U_1; X, U_2) - I(U_1; X)
- I(U_2; Y) + I(U_1; Y, U_2) + I(U_1; Y) = G(D_2) - G(D_1).$$

Therefore, the total rate for Region I-B will be asymptotically given by

$$\frac{|\mathcal{F}_{s_1}| + |\mathcal{F}_{s_2}| - |\mathcal{F}_{s_2}|}{N} \xrightarrow{N \to \infty} 1 - h(D_1 * p) + G(D_2).$$

Furthermore, according to [5], [10], the expected distortions asymptotically approach the target values $D_1$ and $D_2$ at Decoders 1 and 2, respectively, as $N$ becomes sufficiently large. The encoding and decoding complexity of this scheme is $O(N \log N)$.

Note that the coding scheme using polar codes for the Heegard-Berger problem is not exactly the same to that for the Wyner-Ziv problem. The differences are stated as following:

1) Since $U_1^{1:N}$ is the reconstruction derived from the codes $C_{s_1}$ to meet the requirements for Decoder 1, both encoder and the two decoders are able to recover it. In the Wyner-Ziv problem, the realization of the side information is available only at the decoder, and the encoder knows only the joint distribution of the source and side information.

2) The encoder is able to recover $U_1^{1:N}$, and then uses $(X^{1:N}, U_1^{1:N})$ as a joint source to construct a good source code $C_{s_2} = C_N(\mathcal{F}_{s_2}, \bar{0})$ using the test channel $T_\epsilon : U_1 \rightarrow (X, U_1)$. Afterwards the encoder constructs a good channel code $C_{s_2} = C_N(\mathcal{F}_{s_2}, u_{\mathcal{F}_{s_2}}(\bar{v}))$ using the test channel $T_\epsilon : U_2 \rightarrow (Y, U_1)$. In the end, the encoder sends the code bits from the set $\mathcal{F}_{s_2} \setminus \mathcal{F}_{s_2}$, since $C_{s_2}$ and $C_{s_2}$ are nested based on the degradation relation between $T_\epsilon$ and $T_\epsilon$ as given by Proposition 4. However, note that the encoder cannot use the realizations of the side information to generate the codewords in the Wyner-Ziv problem.

3) Decoder 2 cannot be treated strictly experiencing the Wyner-Ziv problem, because the two decoders in the Heegard-Berger problem can ‘collaborate’ in the sense that both of them can recover and employ $U_1^{1:N}$ to help meet their distortion requirements. In other words, Decoder 2 is able to recover the source with distortion $D_2$ by using not only its side information $Y$ but also the reconstruction $U_1^{1:N}$ from Decoder 1, since both $Y$ and $U_1^{1:N}$ contain information about the source. However, the coding system in the Wyner-Ziv problem consists only of a single decoder that decodes solely by the help of its own side information without other collaborators.

In our scheme, the performance of Decoder 2 is more challenging than that of Decoder 1. Thus, the simulation is conducted by fixing $D_1 = 0.35$, $p = 0.4$, and varying $D_2 \in (0, \min(d_c, D_1))$. These settings satisfy the requirements for Region I-B. The performance curves are shown in Fig. 4 for $n = 10, 12, 14, 16, 18$, and the number trials is 5000. It shows that the performances achieved by polar codes approach the HBRDF as $n$ increases.

### B. Coding Scheme for Entire Region I

As Theorem 2 defines the $R_{HB}(D_1, D_2)$ for the entire Region I, we now present a coding scheme that can achieve the HBRDF for the entire Region I. Note that $R_{HB}(D_1, D_2)$ of Region I-B can be explicitly calculated by Corollary 3. Therefore, we can also achieve Region I-B straightforwardly as shown in Section III-A.

From the optimal test channel structure shown in Table I, $U_2$ is a ternary random variable, i.e., $U_2 = \{0, 1, 2\}$. Therefore, we express $U_2$ as two binary random variables $U_a$ and $U_b$, where $U_2 = 2U_a + U_b$, i.e., $(U_a, U_b) \in \{00, 10, 01\}$. For Decoder 1, we
can apply the same scheme specified in the previous subsection to achieve $D_1$. Again, $U_1^{1:N}$ and $X^{1:N}$ can be considered as side information for Decoder 2. Then, the rate required to transmit $U_2^{1:N}$ is evaluated as

$$I(X; U_2|U_1, Y) = I(X; U_a, U_b|U_1, Y)$$

$$= I(X; U_a|U_1, Y) + I(X; U_b|U_1, U_a, Y).$$

Accordingly we can design two separate coding schemes to achieve the rates $I(X; U_a|U_1, Y)$ and $I(X; U_b|U_1, U_a, Y)$, respectively.

As for the first level, since $Y \leftrightarrow X \leftrightarrow (U_1, U_a, U_b)$ form a Markov chain, we have

$$I(X; U_a|U_1, Y) = I(U_a; X, U_1) - I(U_a; Y, U_1),$$

and the test channel $T_{Ca} : U_a \rightarrow (Y, U_1)$ is degraded with respect to $T_{Sa} : U_a \rightarrow (X, U_1)$. We can observe from Table 1 that $U_a$ and $U_b$ can be nonuniform.

Let $K^{1:N} = U_1^{1:N}G_N$, and for $0 < \beta < 0.5$, the frozen set $T_{Sa}(I_{Ca})$, the information set $I_{Sa}(I_{Ca})$, and the shaping set $S_{Sa}(S_{Ca})$ can be identified as

$$T_{Sa} = \{ i \in [N] : Z(K^{1:i-1}, X^{1:N}, U_1^{1:N}) \geq 1 - 2^{-N^\beta} \},$$

$$I_{Sa} = \{ i \in [N] : Z(K^{1:i-1}, X^{1:N}, U_1^{1:N}) < 1 - 2^{-N^\beta} \} \cup \{ i \in [N] : Z(K^{1:i-1}) > 2^{-N^\beta} \},$$

$$S_{Sa} = \{ i \in [N] : Z(K^{1:i-1}) \leq 2^{-N^\beta} \},$$

$$T_{Ca} = \{ i \in [N] : Z(K^{1:i-1}, Y^{1:N}, U_1^{1:N}) \geq 1 - 2^{-N^\beta} \},$$

$$I_{Ca} = \{ i \in [N] : Z(K^{1:i-1}, Y^{1:N}, U_1^{1:N}) < 2^{-N^\beta} \} \cup \{ i \in [N] : Z(K^{1:i-1}) \geq 1 - 2^{-N^\beta} \},$$

$$S_{Ca} = \{ i \in [N] : Z(K^{1:i-1}) < 1 - 2^{-N^\beta} \}.\quad (4)$$

By [5, Lemma 4.7] and channel degradation, we have $T_{Sa} \subseteq T_{Ca}, I_{Sa} \subseteq I_{Ca}$ and $S_{Sa} \subseteq S_{Ca}$. In addition, we observe that $S_{Ca}\setminus S_{Sa}$ can be written as

$$S_{Ca}\setminus S_{Sa} = \left\{ i \in [N] : 2^{-N^\beta} < Z(K^{1:i-1}) < 1 - 2^{-N^\beta} \right\} \cup \left\{ i \in [N] : 2^{-N^\beta} < Z(K^{1:i-1}, Y^{1:N}, U_1^{1:N}) < 1 - 2^{-N^\beta} \right\},$$

therefore, the proportion $|S_{Ca}\setminus S_{Sa}| \rightarrow 0$, as $N \rightarrow \infty$.

**Encoding:** The encoder first applies lossy compression to $X^{1:N}$ with target distortion $D_1$ to obtain $U_1^{1:N}$, and treats $(X^{1:N}, U_1^{1:N})$ as a joint source sequence to evaluate $K_{Sa}$ by randomized rounding with respect to $P_{K_{Sa}|U_1^{1:N}, X^{1:N}}$, i.e.,

$$k^i = \begin{cases} 0 & \text{w.p. } P_{K^i|X^{1:N}, U_1^{1:N}}(0 | k^{1:i-1}, X^{1:N}, U_1^{1:N}) \\ 1 & \text{w.p. } P_{K^i|X^{1:N}, U_1^{1:N}}(1 | k^{1:i-1}, X^{1:N}, U_1^{1:N}) \end{cases} \quad (5)$$

if $i \in I_{Sa}$. And

$$k^i = \left\{ \begin{array}{ll} \tilde{k}^i & \text{if } i \in T_{Sa} \\ \arg \max_k P_{K^i|X^{1:N}, U_1^{1:N}}(k | k^{1:i-1}) & \text{if } i \in S_{Ca} \end{array} \right. \quad (6)$$

where ‘w.p.’ is the abbreviation of ‘with probability’, and $\tilde{k}^i$ is chosen uniformly from $\{0, 1\}$ and shared between the encoder and the decoders before lossy compression. Also note that the second formula in (6) is in fact the maximum a posteriori (MAP) decision for $i \in S_{Sa}$. The encoder sends $K_{I_{Ca}\setminus I_{Sa}}$ to the decoders.

**Decoding:** Using the pre-shared $K_{\bar{F}_{Sa}}$ and received $K_{I_{Ca}\setminus I_{Sa}}$, Decoder 2 recovers $K_{I_{Ca}}$ and $K_{S_{Ca}}$ from the side information sequences $Y^{1:N}$ and $U_1^{1:N}$ by SC decoding algorithm and the MAP rule, respectively. Hence we obtain $K^{1:N}$, $K_{I_{Ca}\cup S_{Sa}}$ and $K_{S_{Ca}\setminus S_{Sa}}$ can be recovered with vanishing error probability, since their Bhattacharyya parameters are arbitrarily small when $N \rightarrow \infty$. Therefore, the reconstruction is given by $U_1^{1:N} = K^{1:N}G_N$.

Let $Q^{1:N\times 1:N\times 1:N\times 1:N}$ denote the resulting joint distribution derived from (5) and (6). Let $P_{K^{1:N}\times X^{1:N}\times U_1^{1:N}}$ denote the joint distribution as a result of another encoder that uses (5) for $i \in [N]$. It is shown in [6, Theorem 2] that for any $\beta' < \beta < 0.5$ satisfying (4) and $R_a = \frac{|I_{Sa}\setminus I_{Ca}|}{N} > I(X; U_a|Y, U_1, Y)$, we have

$$\mathbb{V} \left( P_{K^{1:N}\times X^{1:N}\times U_1^{1:N}}, Q_{K^{1:N}\times X^{1:N}\times U_1^{1:N}} \right) = O \left( 2^{-N^\rho} \right), \quad (7)$$

where $\mathbb{V}(P, Q_X) \leq \frac{1}{2} \sum_x |P_X(x) - Q_X(x)|$ denotes the variational distance between distributions $P_X$ and $Q_X$.

Note that, from [7], $K_{S_{Ca}}$ should be covered by a pre-shared random mapping to achieve (7). However, it is shown in [11] that replacing the random mapping with MAP decision for $K_{S_{Ca}}$ preserves the optimality. Thus, we utilize MAP decoder if $i \in S_{Sa}$ in our scheme.

In terms of the second level, the encoder and Decoder 2 first recover $U_1^{1:N}$. Consequently, the encoder treats $(X^{1:N}, U_1^{1:N}, U_2^{1:N})$ as a joint source, and Decoder 2 treats $(X^{1:N}, U_1^{1:N}, U_2^{1:N})$ as a joint side information. Likewise, according to $Y \leftrightarrow X \leftrightarrow (U_1, U_a, U_b)$, we have

$$I(X; U_b|U_1, U_a, Y) = I(U_b; X, U_1, U_a) - I(U_b; Y, U_1, U_a)$$

and the test channel $T_{Cb} : U_b \rightarrow (Y, U_1, U_a)$ is degraded with respect to $T_{Sa} : U_b \rightarrow (X, U_1, U_a)$.

Similar to the first level, let $W^{1:N} = U_1^{1:N}G_N$ and the frozen set $T_{Sa}(I_{Ca})$, the information set $I_{Sa}(I_{Ca})$, and the shaping set $S_{Sa}(S_{Ca})$ can be adopted from (4) by replacing $X$, $(X, U_1)$ and $(Y, U_1)$ with $W$, $(X, U_1, U_a)$ and $(Y, U_1, U_a)$, respectively. As a result, we have $T_{Sa} \subseteq T_{Ca}, I_{Sa} \subseteq I_{Ca}$ and $S_{Sa} \subseteq S_{Ca}$ by [5, Lemma 4.7] and channel degradation. The encoder evaluates $W_{Sa}$ by randomized rounding with respect to $P_{W^{1:N}|X^{1:N}, U_1^{1:N}, U_2^{1:N}}$, $W_{Sb}$ are pre-shared random bits uniformly chosen from $\{0, 1\}$, and $W_{Sb}$ is determined by MAP decoder defined as $\arg \max_{w} P_{W^{1:N}}(w | w^{1:i-1})$. The encoder sends $W_{Sb}$ to $I_{Ca}$ to the decoders. Decoder 2 recovers $W_{Ca}\cup S_{Sa}$ using the pre-shared $W_{Sa}$ and the side information $(Y^{1:N}, U_1^{1:N}, U_2^{1:N})$. Finally, the reconstruction is given by $U_2^{1:N} = W^{1:N}G_N$.

Let $Q^{1:N\times X^{1:N}\times U_1^{1:N}\times U_2^{1:N}}$ denote the joint distribution when the encoder performs compression, according to the coding scheme presented in the above paragraph. Let $P_{W^{1:N}\times X^{1:N}\times U_1^{1:N}\times U_2^{1:N}}$ denote the resulting joint distribution of the encoder using randomized rounding with respect to
Consider a target distortion value of $\alpha$, $\mu$, $\theta$, $\theta'_1$ that achieves an expected distortion $D_Q$ of Decoder 2 satisfies $D_Q \leq D_2 + O\left(2^{-N^p}\right)$.

Proof: See Appendix A.

As for Decoder 1, we know that $U_1$ can always be taken as the output of a BSC with crossover probability $D_1$ and input $X$. Hence, according to [10, Theorem 3] and Theorem 5, this coding scheme can achieve the optimal HBRDF, as long as the optimal parameters $\alpha$, $\mu$, $\theta$, and $\theta_1$ that achieve the minimum value of $S_{D_1}(\alpha, \mu, \theta, \theta_1)$ can be specified. Finally, we state the achievability of the HBRDF for DSBS for the entire Region I in the following theorem.

**Theorem 6.** Consider target distortions $0 \leq D_2 < 0.5$ and $0 \leq D_2 < \min(D_1, p)$ for DSBS $X$ when side information $Y$ is available only at the Decoder 2. For any $0 < \beta' < \beta < 0.5$ and any rate $R > \min S_{D_1}(\alpha, \mu, \theta, \theta_1)$, there exist a two-level polar code with a rate arbitrarily close to $I(X; U_2|U_1, Y)$, such that the expected distortion $D_Q$ of Decoder 2 satisfies $D_Q \leq D_2 + O\left(2^{-N^p}\right)$ at Decoder 1 and $D_2 + O\left(2^{-N^p}\right)$ at Decoder 2, respectively, if $P_{N, U_1, U_2}$ is as given in Table I.

Proof: It has been shown in [10, Theorem 3] that there exists a polar code with a rate arbitrarily close to $I(X; U_1)$ that achieves an expected distortion $D_1 + O\left(2^{-N^p}\right)$. Theorem 5 shows that a two-level polar code with a rate arbitrarily close to $I(X; U_2|U_1, Y)$ achieves $D_2 + O\left(2^{-N^p}\right)$ at Decoder 2. Finally, the total rate $R_1 + R_2 < I(X; U_1) + I(X; U_2|U_1, Y) = \min S_{D_1}(\alpha, \mu, \theta, \theta_1)$. The last equality holds if the joint distribution $P_{X, U_1, U_2}$ is the same as that given in Table I [4].

Finally, we observe that the encoding and decoding complexity of this coding scheme is $O(N\log N)$.

**IV. POLAR LATTICES FOR GAUSSIAN SOURCES**

It is shown in [6] that polar lattices achieve the optimal rate-distortion performance for both the standard and the Wyner-Ziv compression of Gaussian sources under squared-error distortion. The Wyner-Ziv problem for the Gaussian case can be solved by a nested code structure that combines AWGN capacity achieving polar codes [12] and the rate-distortion optimal ones [6]. Here we show that the HBRDF for the non-degenerate region specified in (2) can also be achieved by a similar nested code structure.

**A. Introduction to Polar Lattices**

We start with a basic introduction to polar lattices. An $n$-dimensional lattice is a discrete subgroup of $\mathbb{R}^n$ which can be described by

$$\Lambda = \{\lambda = B\mathbf{z} : \mathbf{z} \in \mathbb{Z}^n\},$$

where $B$ is the full rank generator matrix. The Voronoi region of $\Lambda$ is defined by $V(\Lambda) = \{z : Q_{\Lambda}(z) = 0\}$, where $Q_{\Lambda}(z) = \min_{\lambda \in \Lambda} \|z - \lambda\|$ is the nearest-neighbor quantizer associated with $\Lambda$. The Voronoi region is one example of the fundamental region of a lattice. A measurable set $R(\Lambda) \subset \mathbb{R}^n$ is a fundamental region of $\Lambda$ if $\cup_{\lambda \in \Lambda} (R(\Lambda) + \lambda) = \mathbb{R}^n$ and if $(R(\Lambda) + \lambda) \cap (R(\Lambda) + \lambda')$ has measure 0 for any $\lambda \neq \lambda'$ in $\Lambda$.

For $\sigma > 0$ and $c \in \mathbb{R}^n$, the Gaussian distribution of variance $\sigma^2$ centered at $c$ is defined as

$$f_{c, \sigma}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \|x - c\|^2}, \quad x \in \mathbb{R}^n.$$  

Let $f_{c, 0}(x) = f_{c, \sigma}(x)$ for short. The $\Lambda$-periodic function is defined as

$$f_{\sigma, \Lambda}(x) = \sum_{\lambda \in \Lambda} f_{\sigma, \Lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda \in \Lambda} e^{-\frac{1}{2\sigma^2} \|x - \lambda\|^2}.$$  

Note that, when $x$ is restricted to the fundamental region $R(\Lambda)$, $f_{\sigma, \Lambda}(x)$ is actually a probability density function (PDF) of the $\Lambda$-aliased Gaussian noise [13].

The flatness factor of a lattice $\Lambda$ is defined as

$$e_{\sigma}(\Lambda) = \max_{x \in R(\Lambda)} |V(\Lambda) f_{\sigma, \Lambda}(x) - 1|,$$

where $V(\Lambda) = |\det(B)|$ denotes the volume of a fundamental region of $\Lambda$ [13]. It can be interpreted as the maximum variation of $f_{\sigma, \Lambda}(x)$ with respect to the uniform distribution over a fundamental region of $\Lambda$.

We define the discrete Gaussian distribution over $\Lambda$ centered at $c$ as the discrete distribution taking values in $\Lambda$ as

$$D_{\Lambda, \sigma, c}(\lambda) = \frac{f_{c, \sigma}(\lambda)}{f_{\sigma, \Lambda}(\lambda)}, \quad \forall \lambda \in \Lambda,$$

where $f_{\sigma, \Lambda}(\lambda) = \sum_{\lambda \in \Lambda} f_{\sigma, \Lambda}(\lambda)$. For convenience, we write $D_{\Lambda, \sigma} = D_{\Lambda, \sigma, 0}$. It has been shown in [14] that lattice Gaussian distribution preserves many properties of the continuous Gaussian distribution when the flatness factor is negligible. To keep the notations simple, we always set $c = 0$ and $n = 1$.

A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by $\Lambda/\Lambda'$) of $\Lambda$ into equivalence groups modulo $\Lambda'$. The order of the partition equals the number of the cosets. If the order is two, we call this a binary partition. Let $\Lambda(\Lambda_0)/\Lambda_1/ \cdots /\Lambda_{r-1}/\Lambda' = (\Lambda_r)$ for $r \geq 1$ be an $n$-dimensional lattice partition chain. For each partition $\Lambda_{l-1}/\Lambda_l$ $(1 \leq l \leq r)$ a code $C_l$ over $\Lambda_{l-1}/\Lambda_l$ selects a sequence of coset representatives $a_l$ in a set $A_l$ of representatives for the cosets of $\Lambda_l$. This construction requires a set of nested linear binary codes $C_l$ with blocklength $N$ and dimension of information bits $k_l$, and $C_1 \subset C_2 \cdots \subset C_r$.  


B. Polar Lattice Construction for Gaussian Sources

For the Gaussian Heegard-Berger problem, let $(X,Y,Z,Z_1,Z_2,U_1,U_2)$ be chosen as specified in Section II-C. Given $Y$ as the side information for Decoder 2, the HBRDF is given by (2). To achieve the HBRDF at Decoder 1, we can design a quantization polar lattice for source $X$ with variance $\sigma_X^2$ and target distortion $D_1$ as in [6]. As a result, for a target distortion $D_1$ and any rate $R_1 > \frac{1}{2} \log (\sigma_Z^2/D_1)$, there exists a multilevel polar lattice with rate $R_1$, such that the average distortion is asymptotically close to $D_1$ when the length $N \to \infty$ and the number of levels $r = O(\log \log N)$ [6, Theorem 4]. Therefore, both decoders can recover $U_1^{1:N}$ and $(U_1^{1:N}, Y^{1:N})$ can be regarded as the side information at Decoder 2.

As for Decoder 2, we first need a code that achieves the rate-distortion requirement for source $X' \triangleq X - U_1$ with Gaussian reconstruction alphabet $U'$. In fact, $X' = Z_1 \sim N(0,D_1)$ is Gaussian and independent of $U_1$ and $Z$. Let
\[
\gamma \triangleq \frac{D_1 \sigma_Z^2}{D_1 \sigma_Z^2 - D_2 (D_1 + \sigma_Z^2)},
\]
and consider an auxiliary Gaussian random variable $U''$ defined as
\[
U'' = X' + Z_3,
\]
where $Z_3 \sim N(0,D_1^Z)$ and $D_1^Z \triangleq \gamma D_2$. Moreover, we define $Y' \triangleq Y - U_1 = X' + Z$ and $Y' \sim N(0,D_1 + \sigma_Z^2)$. Then we can apply the minimum mean square error (MMSE) rescaling parameter $\alpha = \frac{D_1}{D_1 + \sigma_Z^2}$ to $Y'$. As a result, we obtain $X' = \alpha Y'' + Z_3$, where $Z_3 \sim N(0,\alpha \sigma_Z^2)$. We can also write $U'' = \alpha Y'' + Z_3$, where $Z_3 \sim N(0,\gamma D_2 + \alpha \sigma_Z^2)$, which requires an AWGN capacity-achieving code from $\alpha Y''$ to $U''$. This test channel is depicted in Fig. 5.

The final reconstruction at Decoder 2 is given by
\[
\hat{X}_2 = U_1 + \alpha_q U'' + \frac{1}{\gamma}(U'' - \alpha Y''),
\]
Note that $\frac{1}{\gamma}(U'' - \alpha Y'')$ is a scaled version of $Z_3$, which is independent of $Y'$. Thus, the variance of $\alpha Y'' + \frac{1}{\gamma}(U'' - \alpha Y'')$ is $\alpha D_1 + \frac{1}{\gamma}(\gamma D_2 + \alpha \sigma_Z^2) = D_1 + D_2$. Therefore, we have $X - U_1 = \hat{X}_2 - U_1 + N(0,D_2)$ as we desired. Furthermore, the required data rate for Decoder 2 is then given by $R_2 > I(U''; Y') - I(U''; \alpha Y'') = \frac{1}{2} \log \left(\frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\right)$.

Note that $U''$ is a continuous Gaussian random variable which is impractical for the design of polar lattices. Hence, we use the discrete Gaussian distribution $D_{\alpha \sigma_Z^2} U''$ to replace it. Before that, we need to perform MMSE rescaling on $U''$ for test channels $X' \to U'$ and $Y' \to U'$ with scales $\alpha_q$ and $\alpha_c$, respectively, and consider $X' = Z_4 \sim N(0,D_4)$, where $D_4 \triangleq \alpha_q \gamma D_2$. We also have
\[
\begin{align*}
\alpha_q U'' & \to X' \sim N(0,D_4), \\
Z_4 & \sim N(0,\alpha_q \sigma_Z^2), \\
\end{align*}
\]
and
\[
\begin{align*}
\alpha_q U & \to \alpha_q Y' \sim N(0,\alpha_q \sigma_Z^2), \\
Z & \sim N(0,\alpha_q \sigma_Z^2), \\
\end{align*}
\]
Consequently, a reversed version of the test channel in Fig. 5 can be derived, as depicted in Fig. 6, where
\[
\alpha_q = \frac{D_1}{D_1 + D_2^Z} = \frac{D_1 (\sigma_Z^2 - D_2) - D_2 \sigma_Z^2}{D_1 (\sigma_Z^2 - D_2)}
\]
and
\[
\alpha_c = \frac{D_1^Z}{(D_1 + D_2) (D_1 + \sigma_Z^2)} = \frac{D_1 (\sigma_Z^2 - D_2) - D_2 \sigma_Z^2}{(D_1 + \sigma_Z^2) (\sigma_Z^2 - D_2)}.
\]
The reconstruction of $X$ at Decoder 2 is as given in the following proposition.

**Proposition 7.** If we use the reversed test channel shown in Fig 6, the reconstruction of $X$ at Decoder 2 is given by
\[
\hat{X}_2 = U_1 + \alpha_q U'' + \eta \left(\frac{\alpha_q \alpha Y' - \alpha_q U''}{\alpha_c}\right), \quad \eta = \frac{D_2}{\sigma_Z^2}.
\]

**Proof:** It suffices to prove that $X - U_1 = \hat{X}_2 - U_1 + N(0,D_2)$. Since we have $X' = \alpha_q U'' + N(0,\alpha_q \sigma_Z^2)$, as illustrated in Fig 6, showing $\hat{X}_2 - U_1 = \alpha_q U'' + N(0,\alpha_q \sigma_Z^2)$ would complete the proof. We can see from Fig 6 that $\frac{\alpha_q \alpha_c \alpha Y' - \alpha_q U''}{\alpha_c}$ is Gaussian distributed with zero mean and variance $\alpha_q \sigma_Z^2 + \frac{\alpha_q^2 \sigma_Z^2}{\alpha_c}$, and it is independent of $U''$. We also have
\[
\begin{align*}
\alpha_q D_2^Z - D_2 & = \frac{D_1 (\sigma_Z^2 - D_2) - D_2 \sigma_Z^2}{D_1 (\sigma_Z^2 - D_2)} \frac{D_1 D_2 \sigma_Z^2}{D_1 (\sigma_Z^2 - D_2)} - D_2 \\
& = \frac{D_2}{\sigma_Z^2 - D_2},
\end{align*}
\]
and
\[
\begin{align*}
\eta^2 \left(\frac{\alpha_q \sigma_Z^2}{\alpha_c} + \frac{\alpha_q \sigma_Z^2}{\alpha_c}\right) & = \left(\frac{D_2}{\sigma_Z^2 - D_2}ight)^2 \left(\frac{D_2 \sigma_Z^2}{\sigma_Z^2 - D_2} + \frac{D_1 + \sigma_Z^2}{D_1} \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}\right) \\
& = \frac{D_2}{\sigma_Z^2 - D_2},
\end{align*}
\]
as desired. 

Based on the reversed test channel, we can replace the continuous Gaussian random variable $\alpha_q U'$ with a discrete
Gaussian distributed variable \( A \sim D_{A, \sigma_a^2} \), where \( \sigma_a^2 = \alpha_a^2 \sigma_{I_a}^2 \).

Let \( X' = X + N(0, \sigma_a D_1) \) and \( \alpha_a \sigma_a Y' = X' + N(0, \sigma_c^2 \sigma_{Z_c}^2) \).

We also define \( \bar{B} \equiv \alpha_a \sigma_a Y' \), whose variance is \( \sigma_b^2 = \frac{\alpha_a \sigma_a^2 \sigma_c^2 \sigma_{Z_c}^2}{\sigma_a^2 + \sigma_c^2} \).

By [14], the distributions of \( \bar{X}' \) and \( Y' \) can be made arbitrarily close to the distributions of \( X' \) and \( Y' \), respectively. Therefore, polar lattices can be designed for the source \( X' \) and the side information \( Y' \) at Decoder 2. Specifically, a rate-distortion bound-achieving polar lattice \( L_1 \) is constructed for the source \( \bar{X}' \) with distortion \( \alpha_q D_2' \), and an AWGN capacity-achieving polar lattice \( L_2 \) is constructed for the channel \( A \sim \alpha_a \sigma_a Y' \), as shown in Fig. 6. In the end, the reconstruction of Decoder 2 is \( \bar{X}_2 = U_1 + A + \eta(B - A) \).

Even though the quantization noise of \( L_1 \) is not an exact Gaussian distribution, it is shown in [6, Theorem 4] that the two distributions can be arbitrarily close when \( N \) is sufficiently large. Therefore, \( B - A \) can be treated as Gaussian noise independent of \( A \). By Proposition 7, \( \eta \) scales \( B - A \) to \( N(0, \sigma_q D_2' - D_2) \) and by [14], the distributions of \( \bar{X}_2 \) and \( \bar{X}_2' \) can be arbitrarily close, which gives an average distortion close to \( D_2 \).

According to [12, Lemma 10], \( L_1 \) and \( L_2 \) can be equivalently constructed for the MMSE-rescaled channel with Gaussian noise variances

\[
\bar{\sigma}_q^2 = \frac{\sigma_a^2 \sigma_c^2 D_2}{D_1 (\sigma_a^2 - D_2)},
\]

and

\[
\bar{\sigma}_c^2 = \frac{\sigma_a^2 \sigma_c^2}{\sigma_b^2} \left( \frac{\alpha_a \sigma_a^2}{\alpha_c} \right),
\]

The coding strategy for \( L_1 \) and \( L_2 \) can be adapted from [6, Section V]. We briefly describe it for completeness. First, choose a good constellation \( D_{A, \sigma_a^2} \) such that the flatness factor \( \epsilon_A(\bar{\sigma}_q^2) \) is negligible. Let \( A/\mathcal{A}_1 \cdots /\mathcal{A}_{r-1} /\mathcal{A}_r / \cdots \) denote a one-dimensional binary partition chain labeled by bits \( A_1 /A_2 / \cdots /A_{r-1} /A_r / \cdots \). Therefore, \( P_{A_1} \) and \( A_{1:r} \) approach \( D_{A, \sigma_a^2} \) and \( A \), respectively, as \( r \to \infty \). Consider \( N \) i.i.d. copies of \( A \), and let \( K_i^{1:N} \equiv A_i:G_N \) for each level \( 1 \leq l \leq r \). For \( 0 < \beta < 0.5 \), the frozen set \( \mathcal{I}_l^Q(\mathcal{I}_l^C) \), information set \( \mathcal{I}_l^Q \) \( \mathcal{I}_l^C \), and shaping set \( S_l^Q \) \( \mathcal{S}_l^C \) for \( L_1 \) (\( L_2 \)) at level \( l \) can be adapted from [6, Equation (35)] \( [6, \text{Equation (36)}] \), respectively, by replacing \( \bar{X} \) with \( X' \).

Furthermore, according to [6, Lemma 2], \( L_2 \) is nested within \( L_1 \), i.e., \( L_2 \subseteq L_1 \). By the fact \( \sigma_c^2 \leq \sigma_a^2 \) and [12, Lemma 3], the partition channel \( A_l:G \) with noise variance \( \sigma_c^2 \) is degraded with respect to the one with noise variance \( \sigma_a^2 \). Therefore, we have \( \mathcal{I}_l^Q \subseteq \mathcal{I}_l^C \), \( \mathcal{I}_l^C \subseteq \mathcal{I}_l^Q \), and by the definition of shaping set, we observe that \( \mathcal{S}_l^Q \subseteq \mathcal{S}_l^C \).

The encoder can recover the auxiliary codeword \( U_1^{1:N} \) for Decoder 1, and obtains the realizations \( X_1^{1:N} \) \( Y_1^{1:N} \) of \( X_1^{1:N} = X'_{1:N} - U_1^{1:N} \) \( Y_1^{1:N} = Y_{1:N} - U_1^{1:N} \) from given realizations of variables \( X_1^{1:N} \) \( Y_1^{1:N} \), respectively. The encoder recovers \( k_i^{1:N} \) from \( l = [r] \) successively according to the random rounding quantization rules given in [6, Equations (13), (14), (17) and (18)]. Note that \( X_1^{1:N} \) as realization of \( X_{1:N} \) is acceptable since the distributions of \( X' \) and \( \bar{X}' \) are arbitrarily close. Also, according to [11], replacing the random rounding rule with MAP decision to obtain \( k_i^{0:Q} \) will not affect [12, Theorem 5] and [12, Theorem 6]. Consequently, the coding scheme for Decoder 2 for the Gaussian Heegard-Berger problem can be summarized as following:

**Encoding:** From the \( N \)-dimensional i.i.d. source vector \( X_1^{1:N} \), the encoder recovers the auxiliary codeword \( U_1^{1:N} \) employing a quantization polar lattice for source \( X \) with variance \( \sigma_X^2 \) and distortion \( D_1 \), and obtains \( X_1^{1:N} \) \( Y_1^{1:N} \). Next, the encoder evaluates \( K_i^{0:Q} \) by random rounding and sends \( k_i^{0:Q} \) to the decoders.

**Decoding:** By the pre-shared \( k_i^{0:Q} \) and received \( k_i^{0:Q} \) \( k_i^{0:Q} \) from the side information \( X_1^{1:N} \) with vanishing error probability, by using SC decoding for Gaussian channels [12]. At each level, Decoder 2 obtains \( k_i^{1:N}, A_1^{1:N} \) can be recovered according to [6, Equation (38)]. Finally, the reconstruction of Decoder 2 is

\[
\bar{X}_2^{1:N} = U_1^{1:N} + A_1^{1:N} + \eta(B - A).
\]
to \( \frac{1}{2} \log \left( \frac{D_2 \sigma_X^2}{D_1 (D_1 + \sigma_Z^2)} \right) \) such that the expected distortion \( D_{Q_2} \) satisfies \( D_{Q_2} \leq D_2 + O \left( 2^{-N \eta'} \right) \).

**Proof:** The achievability of the rate-distortion function for Decoder 1 follows from [6, Theorem 4]. The proof of the achievability for Decoder 2 can be adapted from [6, Theorem 5], by considering the test channel depicted in Fig. 6 and the reconstruction as given by (9). It is worth mentioning that the requirements \( \epsilon (\sigma_X^2, \sigma_Z) = O \left( 2^{-\sqrt{N \eta'}} \right) \) and \( r = O (\log N) \) are given by [6, Proposition 2] to guarantee that Decoder 2 can recover \( X'_{1C} \) with sub-exponentially decaying error probability by using SC decoding.

For the simulations, we first plot the theoretical \( R_{HB}(D_1, D_2) \) for the non-degenerate region specified by (2), where we set \( \sigma_X = 1 \) and \( \sigma_Z = 1 \). The rate-distortion function \( R_{HB}(D_1, D_2) \) for this region (i.e., \( D_1 \leq \sigma_X^2 = 1 \) and \( D_2 \leq D_1 \sigma_Z^2 \)) can be depicted as shown in Fig. 7. We also highlighted the rate-distortion bound in Fig. 7 when \( D_1 = 0.5 \), which is the same rate-distortion bound in Fig. 8 marked by red. In our simulation, we fixed \( D_1 = 0.5 \), and vary the values of \( D_2 \) in the non-degenerate region. We construct polar lattices with the number of levels \( r = 6 \), and set the blocklength of the polar codes in each level as \( N = 2^{10}, 2^{12}, \ldots, 2^{18} \).

The simulation consists in two parts. First, we employ the lossy compression of \( N \) i.i.d. sources \( X^{1:N} \sim \mathcal{N}(0, 1) \) with distortion \( D_1 = 0.5 \), and the reconstruction at Decoder 1 is denoted by \( U^{1:N}_1 \). Second, we generate the side information \( Y^{1:N} = X^{1:N} + Z^{1:N} \), where \( Z^{1:N} \) are \( N \) i.i.d. standard Gaussian random variables. Then we use \( Y^{1:N} \) and the same realization of sources \( X^{1:N} \) to compute \( X'^{1:N} = X^{1:N} - U^{1:N}_1, Y'^{1:N} = Y^{1:N} - U^{1:N}_1 \) and \( \tilde{B}^{1:N} \). Therefore, we can construct a nested polar lattice structure consisting of a rate-distortion achieving lattice \( L_1 \) with source \( X'^{1:N} \) and reconstruction \( A^{1:N} \), and a capacity-achieving lattice \( L_2 \) with input \( A^{1:N} \) and output \( \tilde{B}^{1:N} \), as depicted in Fig. 6. Finally, the reconstruction at Decoder 2 is denoted by \( \tilde{X}'^{1:N} = U^{1:N}_2 + A^{1:N} + \eta (\tilde{B}^{1:N} - A^{1:N}) \), where \( \eta = \frac{D_2}{\sigma_Z^2} = D_2 \). The coding process is iterated for \( 10^4 \) times until the average distortion of both decoders are stable and converged. The simulation results are presented in Fig. 8 that the performances achieved by polar lattices approach the HBRDF for the Gaussian case as \( N \) increases.

**V. CONCLUSION**

We presented nested polar codes and polar lattices that achieve the rate-distortion function for the binary and Gaussian Heegard-Berger problems, respectively. Different from the code constructions for the Wyner-Ziv problem [5] and [6], we took advantage of the reconstruction at Decoder 1 to build the nested structure that achieves the rate-distortion function for Decoder 2. The proposed schemes achieve the HBRDF in the entire non-degenerate regions for both DSBS and Gaussian sources.

Finally, the Kaspi problem in [15] is regarded as a generalization of the Heegard-Berger problem, where the encoder may also have access to the side information. The explicit rate-distortion functions for the Kaspi problem with Gaussian and binary sources have been given in [16] and [17], respectively. We will study the construction of polar codes and polar lattices for the Kaspi problem in our future work.

**APPENDIX A**

**PROOF OF THEOREM 5**

First, we show that the distortion \( D_2 \) can be achieved. Since \( U^{1:N}_b = W^{1:N}_b G_N \) gives a one-to-one mapping between \( W^{1:N} \) and \( U^{1:N}_b \), expression (8) is equivalent to

\[
\forall \left( P_{u^{1:N}_a, u^{1:N}_b, x^{1:N}, u^{1:N}_i, q_{u^{1:N}_a, u^{1:N}_b, x^{1:N}, u^{1:N}_i}} \right) = O \left( 2^{-N \eta'} \right).
\]

From the coding scheme presented in Section III-B, we assume that \( K_{Ic_a} \) and \( W_{Ic_2} \) can be correctly decoded by using side information, and \( K_{S_{a}} \) and \( W_{S_{b}} \) can be recovered by the MAP rule. Therefore, Decoder 2 can recover \( U^{1:N}_a \) and \( U^{1:N}_b \).
with the joint distribution \( Q_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \). Denote by \( Q_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \) and \( Q_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \) the resulting distribution when the encoder performs compression at each level, i.e., compresses \( X_1^{1:N} \) to \( W_{S_a} \). Let \( P_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \) denote the joint distribution when the encoder does not perform compression. For simplicity, we denote random variables \( U_1^{1:N} \), \( Q_1^{1:N} \) and \( U_b^{1:N} \) by \( U_1^{1:N} \).

Next we show that the decoder can recover \( U_a^{1:N} \) and \( U_b^{1:N} \) with a sub-exponentially decaying block error probability.

\[
\mathbb{P}
\begin{align*}
\left( P_{U_a^{1:N}, U_b^{1:N}, Y_1^{1:N}} Q_{U_a^{1:N}, U_b^{1:N}, Y_1^{1:N}} \right) \\
= \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} \left| P \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N}, Y_1^{1:N} \right) - Q \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N}, Y_1^{1:N} \right) \right| \\
= \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} \left| P \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) - Q \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) \right| \\
= \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} P \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) - Q \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) \\
= \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} \left| P \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) - Q \left( u_1^{1:N} \mid u_{1:a}^{1:N}, x_1^{1:N} \right) \right| \\
= \mathcal{O} \left( 2^{-N^{2p}} \right).
\end{align*}
\]

Let \( E_{b}^p \) and \( E_{b}^p \) denote the expectation error probability as a result of the distribution \( P_{U_a^{1:N}, U_b^{1:N}, Y_1^{1:N}} \) at level 1 and 2, respectively. Take \( E_{b}^p \) as an example to show the decoding error probability. Let \( E_i \) denote the set of random variables \( \left( u_b^{1:N}, u_a^{1:N}, y_1^{1:N} \right) \) such that the SC decoding error occurred at the \( i \)th bit. Hence the block error event is defined by \( E_b^p \) \( \subseteq \bigcup_{i \in E_b} E_i \), and the expectation of decoding block error probability over all random mapping is given by

\[
E_{b}^p \left[ P_e \right] = \sum_{u_b^{1:N}, u_a^{1:N}, y_1^{1:N}} \mathbb{P} \left( u_b^{1:N}, u_a^{1:N}, y_1^{1:N} \right) I \left( u_b^{1:N}, u_a^{1:N}, y_1^{1:N} \right) \subseteq \sum_{i \in E_b} \sum_{u_b^{1:N}, u_a^{1:N}, y_1^{1:N}} \mathbb{P} \left( u_b^{1:N}, u_a^{1:N}, y_1^{1:N} \right) I \left( u_b^{1:N}, u_a^{1:N}, y_1^{1:N} \right)
\]

Note the last equality holds due to the constrains \((\theta - \theta_1) + \theta_1 \mu + (1 - \theta) p = D_2 \) of Theorem 2. This is reasonable because \( D_2 \) is achieved when the encoder does not perform any compression. Combined with (11), the expected distortion \( D_Q \) achieved by \( Q_{U_a^{1:N}, U_b^{1:N}, Y_1^{1:N}} \) satisfies

\[
\begin{align*}
D_Q - D_P &= \frac{1}{N} \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} \left( Q_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} - P_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \right) \\
&\leq \frac{1}{N} \sum_{u_1^{1:N}, u_{1:a}^{1:N}, y_1^{1:N}} \left| P_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} - Q_{U_a^{1:N}, U_b^{1:N}, X_1^{1:N}, Y_1^{1:N}} \right| \\
&= \mathcal{O} \left( 2^{-N^{2p}} \right).
\end{align*}
\]
\[ P_e^{HB} = E_P [P_e] = \sum_{u_1^{1:N},u_2^{1:N}} \left( Q \left( U_1^{1:N}, Y_1^{1:N} \right) - P \left( U_1^{1:N} \right) \right) \cdot I \left( U_1^{1:N}; Y_1^{1:N} \right) \in \mathcal{E} \] 
\[ \leq 2V \left( P_{U_1^{1:N}} Y_1^{1:N}, Q_{U_1^{1:N}} Y_1^{1:N} \right) \leq O \left( 2^{-\beta N^2} \right). \]

As for the rates, we have 
\[ \frac{I_{S_0}}{N} \xrightarrow{N \to \infty} I(U_a; X, U_1) \] and 
\[ \frac{I_{C_a}}{N} \xrightarrow{N \to \infty} I(X; U_a U_1 Y). \]

For the second level, 
\[ \frac{I_{S_0}}{N} \xrightarrow{N \to \infty} I(U_b; X, U_1 U_a) \] and 
\[ \frac{I_{C_a}}{N} \xrightarrow{N \to \infty} I(U_b; Y, U_1 U_a). \] Thus, we have
\[ \frac{I_{S_0}}{N} - \frac{I_{C_a}}{N} \xrightarrow{N \to \infty} I(U_b; Y, U_1 U_a, Y). \]

Finally, the rate of Decoder 2 is
\[ \frac{I_{S_0}}{N} - \frac{I_{C_a}}{N} + \frac{I_{S_b}}{N} - \frac{I_{C_b}}{N} \xrightarrow{N \to \infty} I(X; U_2 U_1 Y). \]

**References**


