Polar Codes and Polar Lattices for the Heegard-Berger Problem

Jinwen Shi, Ling Liu, Deniz Gündüz and Cong Ling
Department of Electronic and Electrical Engineering
Imperial College London, UK
{jinwen.shi12,l.liu12,d.gunduz,c.ling}@imperial.ac.uk

***Abstract*—Explicit coding schemes are proposed to achieve the rate-distortion bound for the Heegard-Berger problem using polar codes. Specifically, a nested polar code construction is employed to achieve the rate-distortion bound for the binary case. 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 for the Gaussian case. The proposed polar lattice is constructed by nesting a quantization polar lattice and an AWGN capacity-achieving polar lattice.**

I. Introduction

The well-known Wyner-Ziv problem is a lossy source coding problem in which a source sequence is compressed to be reconstructed in the presence of a correlated side information at the decoder [1]. An interesting question is whether a reconstruction with a non-trivial distortion quality can still be obtained in the absence of the side information at the receiver. The equivalent coding system for this problem 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, 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. 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.

Polar codes are optimal for the Wyner-Ziv problem [5], as well as general distributed hierarchical source coding problems [6]. The optimality of polar codes for lossy compression of nonuniform sources is shown in [7]. For Gaussian sources, a polar lattice to achieve both the classical and Wyner-Ziv rate-distortion functions is proposed in [8]. Practical codes for the Gaussian Heegard-Berger problem were developed in [9] which hybridize trellis codes and low-density parity-check codes. Here, we propose practical coding schemes using polar codes and polar lattices to achieve the theoretical performance bound in the Heegard-Berger problem.

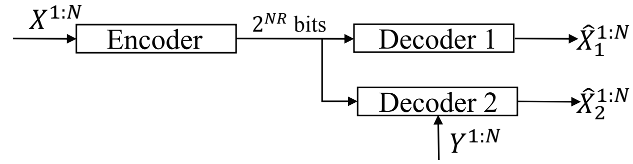

Fig. 1. Illustration of the Heegard-Berger rate-distortion problem.

Notation: P_X denotes the probability distribution of a random variable X taking values in set $\mathcal{X}$. For two positive integers $i < j$, $x^{i:j}$ denotes the vector $\left(x^i, \ldots, x^j\right)$, which represents the realizations of random variables $X^{i:j}$. For a set $\mathcal{F}$ of positive integers, $x_{\mathcal{F}}$ denotes the subvector $\{x^i\}_{i\in\mathcal{F}}$. $\mathcal{F}^c$ and $|\mathcal{F}|$ denote the complement and cardinality of set $\mathcal{F}$, respectively. For a positive integer N, we define $[N] \triangleq \{1, \ldots, N\}$. All logarithms are base two, and information is measured in bits.

II. Problem Statement

A. Heegard-Berger Problem

Let $(\mathcal{X}, \mathcal{Y}, P_{XY})$ 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 $\mathcal{Y}$.

Definition 1. An (N, M, D_1, D_2) Heegard-Berger code for source X with side information Y consists of an encoder $f^{(N)} : \mathcal{X}^N \to [M]$ and two decoders $g_1^{(N)} : [M] \to \hat{\mathcal{X}}_1^N$; $g_2^{(N)} : [M] \times \mathcal{Y}^N \to \hat{\mathcal{X}}_2^N$, where $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2$ are finite reconstruction alphabets, such that

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

where $\mathbb{E}$ is the expectation operator, and $d(\cdot,\cdot) < \infty$ is a per-letter distortion measure.

In this paper, we set $d(\cdot,\cdot)$ to be the Hamming distortion for binary sources, and the squared error distortion for Gaussian sources.

Definition 2. Rate R is said to be $\{(D_1, D_2) - achievable\}$, 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 HBRDF, $R_{HB}(D_1, D_2)$, 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 the following theorem.

Theorem 3. *([2, Theorem 1])*

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

where $\mathcal{P}\left(D_1, D_2\right)$ is the set of all auxiliary random variables $\left(U_1, U_2\right) \in \mathcal{U}_1 \times \mathcal{U}_2$ jointly distributed with the generic random variables $\left(X, Y\right)$, such that: i) $Y \leftrightarrow X \leftrightarrow \left(U_1, U_2\right)$ form a Markov chain; ii) $|\mathcal{U}_1| \leq |\mathcal{X}| + 2$ and $|\mathcal{U}_2| \leq \left(|\mathcal{X}| + 1\right)^2$; iii) there exist functions φ_1 and φ_2 such that $\mathbb{E}d\left(X, \varphi_1\left(U_1\right)\right) \leq D_1$ and $\mathbb{E}d\left(X, \varphi_2\left(U_1, U_2, Y\right)\right) \leq D_2$.

B. Doubly Symmetric Binary Sources (DSBS)

Let X be a binary DMS, i.e., $\mathcal{X} = \{0, 1\}$, with uniform distribution. The binary side information is specified by $Y = X \oplus Z$, where Z is a Bernoulli random variable with $P_Z\left(z = 1\right) = p < 0.5$, and $\oplus$ denotes modulo two addition.

The HBRDF for DSBS can be characterized over four regions [3]. Region I ($0 \leq D_1 < 0.5$ and $0 \leq D_2 < \min\left(D_1, p\right)$) 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 $D_2 \geq \min\left(D_1, p\right)$) is also degenerate since the problem boils down to the classical lossy compression problem for the first decoder; Region IV ($D_1 > 0.5$ and $D_2 > p$) can be trivially achieved without coding. 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 $G\left(u\right) \triangleq h\left(p * u\right) - h(u)$ from [1], defined over the domain $0 \leq u \leq 1$, where $h(u)$ is the binary entropy function $h(u) \triangleq -u \log u - (1-u) \log\left(1 - u\right)$, and $p * u$ is the binary convolution for $0 \leq p, u \leq 1$, defined as $p * u \triangleq p\left(1 - u\right) + u\left(1 - p\right)$. The explicit calculation of HBRDF for DSBS in Region I is given in [4] as follows:

Define the function

$$S_{D_1}\left(\alpha, \mu, \theta, \theta_1\right) \triangleq$$
$$1 - h\left(D_1 * p\right) + \left(\theta - \theta_1\right) G\left(\alpha\right) + \theta_1 G\left(\mu\right) + \left(1 - \theta\right) G\left(\gamma\right),$$

where

$$\gamma \triangleq \begin{cases} \frac{D_1 - \left(\theta - \theta_1\right)\left(1 - \alpha\right) - \theta_1 \mu}{1 - \theta} & \theta \neq 1 \\ 0.5 & \theta = 1 \end{cases},$$

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

The next theorem characterizes the HBRDF in Region I.

Theorem 4. *[4, Theorem 2] For $0 \leq D_1 < 0.5$ and $0 \leq D_2 < \min\left(D_1, p\right)$, we have $R_{HB}\left(D_1, D_2\right) = \min S_{D_1}\left(\alpha, \mu, \theta, \theta_1\right)$, where the minimization is subject to the constraint $\left(\theta - \theta_1\right) \alpha + \theta_1 \mu + \left(1 - \theta\right) p = D_2$.*

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

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

	$(u_1, x) = (0, 0)$	$(u_1, x) = (0, 1)$	$(u_1, x) = (1, 0)$	$(u_1, x) = (1, 1)$
$u_2 = 0$	$\frac{1}{2}\theta_1\left(1 - \mu\right)$	$\frac{1}{2}\theta_1 \mu$	$\frac{1}{2}\left(\theta - \theta_1\right) \cdot \left(1 - \alpha\right)$	$\frac{1}{2}\left(\theta - \theta_1\right) \alpha$
$u_2 = 1$	$\frac{1}{2}\left(\theta - \theta_1\right) \alpha$	$\frac{1}{2}\left(\theta - \theta_1\right) \cdot \left(1 - \alpha\right)$	$\frac{1}{2}\theta_1 \mu$	$\frac{1}{2}\theta_1\left(1 - \mu\right)$
$u_2 = 2$	$\frac{1}{2}\left(1 - \theta\right) \cdot \left(1 - \gamma\right)$	$\frac{1}{2}\left(1 - \theta\right) \gamma$	$\frac{1}{2}\left(1 - \theta\right) \gamma$	$\frac{1}{2}\left(1 - \theta\right) \cdot \left(1 - \gamma\right)$

TABLE I
JOINT DISTRIBUTION $P_{X,U_1,U_2}\left(x, u_1, u_2\right)$ [4].

Fig. 2. The optimal forward test channel for Region I-B. The crossover probability η for the BSC between U_2 and U_1 satisfies $D_2 * \eta = D_1$.

Then the following corollary given by [4] specifies the HBRDF in region $D_2 \leq \min\left(d_c, D_1\right)$ and $D_1 \leq 0.5$. This region belongs to Region I and we refer to it as Region I-B.

Corollary 5. *(Region I-B [4, Corollary 2]) For distortion pairs $\left(D_1, D_2\right)$ satisfying $D_1 \leq 0.5$ and $D_2 \leq \min\left(d_c, D_1\right)$, we have*

$$R_{HB}\left(D_1, D_2\right) = 1 - h\left(D_1 * p\right) + G\left(D_2\right). \quad (1)$$

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. 2.

We first propose a polar code design that achieves the HBRDF in Region I-B for DSBSs in Section III-A. We then provide a polar code construction that achieves the HBRDF in the entire Region I in Section III-B.

C. Gaussian Sources

Suppose $Y = X + Z$, where X and Z are independent (zero-mean) Gaussian random variables with variances σ_X^2 and σ_Z^2, respectively, i.e., $X \sim \mathcal{N}\left(0, \sigma_X^2\right)$ and $Z \sim \mathcal{N}\left(0, \sigma_Z^2\right)$. The explicit expression for $R_{HB}\left(D_1, D_2\right)$ 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, Z_1 and Z_2 are independent zero-mean Gaussian. We have $Z_i \sim \mathcal{N}\left(0, D_i\right),\ i = 1, 2$.

For $D_1 \leq \sigma_X^2$ and $D_2 \geq \frac{D_1 \sigma_Z^2}{D_1 + \sigma_Z^2}$, the problem degenerates into a classical lossy compression problem for Decoder 1, and the HBRDF is given by $R_{HB}\left(D_1, D_2\right) = \frac{1}{2} \log\left(\frac{\sigma_X^2}{D_1}\right)$. 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 $R_{HB}\left(D_1, D_2\right) = \frac{1}{2} \log\left(\frac{\sigma_X^2 \sigma_Z^2}{D_2\left(\sigma_X^2 + \sigma_Z^2\right)}\right)$. 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 classical and Wyner-Ziv rate-distortion functions for Gaussian sources, introduced in [8], 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]:

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

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 $R_{HB}\left(D_1, D_2\right)$ for DSBS in Region I. First, we give a brief overview of polar codes.

Let $G_2 \triangleq \left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$, 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 $\mathcal{C}_N\left(\mathcal{F}, u_{\mathcal{F}}\right)$ [5] is a linear code defined by $\mathcal{C}_N\left(\mathcal{F}, u_{\mathcal{F}}\right) = \left\{v^{1:N}G_N : v_{\mathcal{F}} = u_{\mathcal{F}}, v_{\mathcal{F}^c} \in \{0,1\}^{|\mathcal{F}^c|}\right\}$, 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 $\mathcal{C}_N\left(\mathcal{F}, u_{\mathcal{F}}\right)$ 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 $X \in \mathcal{X} = \{0,1\}$ and output $Y \in \mathcal{Y}$, the Bhattacharyya parameter Z is defined as $Z\left(X|Y\right) \triangleq 2\sum_y \sqrt{P_{X,Y}\left(0,y\right)P_{X,Y}\left(1,y\right)}$.

A. Polar Code Construction for Region I-B

As for Region I-B, we observe from the optimal test channel in Fig. 2 that the minimum rate for Decoder 1 to achieve the target distortion D_1 is $R_1 = I\left(U_1; X\right) = 1 - h\left(D_1\right)$. It is shown in [10, Theorem 3] that polar codes are optimal for lossy source coding for the binary symmetric source. 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 $\hat{X}_1^{1:N}$ 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. 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 at Decoder 2 is very 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 in compressing a source $X^{1:N}$ in the presence of some side information $Y^{1:N}$ using polar codes. The code $\mathcal{C}_s$ with corresponding frozen set $\mathcal{F}_s$ is designed to be a good source code for distortion D_2. Further, the code $\mathcal{C}_c$ with corresponding frozen set $\mathcal{F}_c$ is designed to be a good channel code for BSC($D_2 * p$). It has been shown in [5] that $\mathcal{F}_c \supseteq \mathcal{F}_s$, because the test channel BSC($D_2 * p$) is degraded with respect to BSC(D_2). In this case, the encoder transmits to the decoder the subvector that belongs to the index set $\mathcal{F}_c \backslash \mathcal{F}_s$. 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 $\left(U_1, U_2\right) \in \mathcal{P}\left(D_1, D_2\right)$,

$$\begin{aligned} I\left(X; U_2|U_1, Y\right) &= I\left(U_2; X, U_1, Y\right) - I\left(U_2; Y, U_1\right) \\ &= I\left(U_2; X, U_1\right) - I\left(U_2; Y, U_1\right). \end{aligned} \quad (3)$$

The second equality holds since $Y \leftrightarrow X \leftrightarrow \left(U_1, U_2\right)$ form a Markov chain. Motivated by (3), the code $\mathcal{C}_{s_2}$ with corresponding frozen set $\mathcal{F}_{s_2}$ is designed to be a good source code for the source pair $\left(X^{1:N}, U_1^{1:N}\right)$ with reconstruction $U_2^{1:N}$. T_s denotes the test channel for this source code. Additionally, code $\mathcal{C}_{c_2}$ with corresponding frozen set $\mathcal{F}_{c_2}$ is designed to be a good channel code for the test channel T_c with input $U_2^{1:N}$ and output $\left(Y^{1:N}, U_1^{1:N}\right)$. According to [5, Lemma 4.7], in order to show the nested structure between $\mathcal{C}_{s_2}$ and $\mathcal{C}_{c_2}$, we need to show that T_c is stochastically degraded with respect to T_s.

Proposition 6. *$T_c : U_2 \to \left(Y, U_1\right)$ is stochastically degraded with respect to $T_s : U_2 \to \left(X, U_1\right)$, if the random variables $\left(X, Y, U_1, U_2\right)$ agree with the forward test channel as shown in Fig. 2.*

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

$$\begin{aligned} &P_{Y,U_1|U_2}\left(y, u_1|u_2\right) \\ &= P_{Y|U_2}\left(y|u_2\right)P_{U_1|U_2}\left(u_1|u_2\right) \\ &= \sum_x P_{X,Y|U_2}\left(x, y|u_2\right)P_{U_1|U_2}\left(u_1|u_2\right) \\ &= \sum_x P_{X|U_2}\left(x|u_2\right)P_{Y|X,U_2}\left(y|x, u_2\right)P_{U_1|U_2}\left(u_1|u_2\right) \\ &= \sum_x P_{X,U_1|U_2}\left(x, u_1|u_2\right)P_{Y|X}\left(y|x\right), \end{aligned}$$

completing the proof. ■

Therefore, we can claim that $\mathcal{F}_{c_2} \supseteq \mathcal{F}_{s_2}$, and rather than sending the entire vector that belongs to the index set $\mathcal{F}_{s_2}^c$, the encoder sends only the subvector that belongs to $\mathcal{F}_{c_2} \backslash \mathcal{F}_{s_2}$ to Decoder 2, since Decoder 2 can extract some information on $U_2^{1:N}$ from the side information $\left(U_1^{1:N}, Y^{1:N}\right)$. In summary, the coding scheme 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 $\mathcal{C}_{s_1} = \mathcal{C}_N\left(\mathcal{F}_{s_1}, \bar{0}\right) = \left\{w^{1:N}G_N : w_{\mathcal{F}_{s_1}} = \bar{0}, w_{\mathcal{F}_{s_1}^c} \in \{0,1\}^{|\mathcal{F}_{s_1}^c|}\right\}$, and the encoder transmits the compressed sequence $w_{\mathcal{F}_{s_1}^c}$ to the decoders. The encoder is also able to recover $U_1^{1:N}$ from $\mathcal{C}_{s_1}$. Next, the encoder applies lossy compression jointly for sources $\left(X^{1:N}, U_1^{1:N}\right)$ with reconstruction $U_2^{1:N}$ and target distortion D_2 and $d\left(U_1, U_2\right) = \eta$. We then construct $\mathcal{C}_{s_2} = \mathcal{C}_N\left(\mathcal{F}_{s_2}, \bar{0}\right)$. Finally, the encoder applies channel coding to the symmetric test channel T_c with input $U_2^{1:N}$ and output $\left(Y^{1:N}, U_1^{1:N}\right)$. We derive $\mathcal{C}_{c_2} = \mathcal{C}_N\left(\mathcal{F}_{c_2}, u_{\mathcal{F}_{c_2}}\left(\bar{v}\right)\right)$, where $u_{\mathcal{F}_{c_2}}(\bar{v})$ is defined by $u_{\mathcal{F}_{s_2}} = \bar{0}$ and $u_{\mathcal{F}_{c_2}\setminus\mathcal{F}_{s_2}} = \bar{v}$ for $\bar{v} \in \{0,1\}^{|\mathcal{F}_{c_2}\setminus\mathcal{F}_{s_2}|}$. The encoder sends the subvector $u_{\mathcal{F}_{c_2}\setminus\mathcal{F}_{s_2}}$ to the decoders.

Decoding: Decoder 1 receives $w_{\mathcal{F}^c_{s_1}}$ and outputs the reconstruction sequence $u_1^{1:N} = w^{1:N}G_N$. Decoder 2 receives $u_{\mathcal{F}_{c_2}\setminus\mathcal{F}_{s_2}}$, and hence, it can derive $u_{\mathcal{F}_{c_2}}$. Moreover, Decoder 2 can also recover $U_1^{1:N}$ from $w_{\mathcal{F}^c_{s_1}}$. Decoder 2 applies the successive cancellation (SC) decoding algorithm to obtain the codeword $U_2^{1:N}$ from the realizations of $\left(Y^{1:N}, U_1^{1:N}\right)$.

Next we present the rates that can be achieved in the proposed scheme. From the polarization theorem for lossy source coding in [10], we know that the required rate for Decoder 1 is $\frac{|\mathcal{F}^c_{s_1}|}{N} \xrightarrow{N\to\infty} I(U_1;X) = 1 - h(D_1)$.

From the polarization theorems for source and channel coding [5], the code rate required for reliable decoding at Decoder 2 can be derived by $\frac{|\mathcal{F}_{c_2}|-|\mathcal{F}_{s_2}|}{N} \xrightarrow{N\to\infty} I(U_2;X,U_1) - I(U_2;Y,U_1) = G(D_2) - G(D_1)$. Therefore, the total rate will be asymptotically given by $\frac{|\mathcal{F}^c_{s_1}|+|\mathcal{F}_{c_2}|-|\mathcal{F}_{s_2}|}{N} \xrightarrow{N\to\infty} 1 - h(D_1 * p) + G(D_2)$ for Region I-B.

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 $O(N \log N)$.

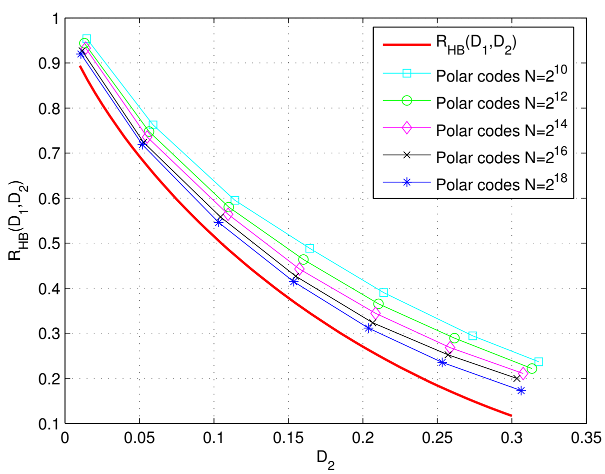

Fig. 3. Simulation of $R_{HB}(D_1, D_2)$ corresponding to D_2 for Region I-B.

Note that, 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. 3 for $n = 10, 12, 14, 16, 18$. It shows that the performances achieved by polar codes approaches the HBRDF as n increases.

B. Coding Scheme for Entire Region I

Now, we present a coding scheme that can achieve the HBRDF given by Theorem 4 for the entire Region I.

From the optimal test channel structure shown in Table I, U_2 is a ternary random variable, i.e., $\mathcal{U}_2 = \{0,1,2\}$. Therefore, we express U_2 as two binary random variables U_a and U_b, where $U_2 = 2U_b + U_a$, 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 $Y^{1:N}$ 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.

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 $T_{C_a} : U_a \to (Y, U_1)$ is degraded with respect to $T_{S_a} : U_a \to (X, U_1)$. We can observe from Table I that U_a and U_b can be nonuniform.

Let $K^{1:N} = U_a^{1:N}G_N$, and for $0 < \beta < 0.5$, the frozen set $\mathcal{F}_{S_a}$ ($\mathcal{F}_{C_a}$), the information set $\mathcal{I}_{S_a}$ ($\mathcal{I}_{C_a}$), and the shaping set $\mathcal{S}_{S_a}$ ($\mathcal{S}_{C_a}$) can be identified as

$$
\begin{aligned}
\mathcal{F}_{S_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}, X^{1:N}, U_1^{1:N}\right) \geq 1 - 2^{-N^\beta} \right\} \\
\mathcal{I}_{S_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}, X^{1:N}, U_1^{1:N}\right) < 1 - 2^{-N^\beta} \right\} \\
&\qquad \cap \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}\right) > 2^{-N^\beta} \right\} \\
\mathcal{S}_{S_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}\right) \leq 2^{-N^\beta} \right\}, \\
\mathcal{F}_{C_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}, Y^{1:N}, U_1^{1:N}\right) \geq 1 - 2^{-N^\beta} \right\} \\
\mathcal{I}_{C_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}, Y^{1:N}, U_1^{1:N}\right) \leq 2^{-N^\beta} \right\} \\
&\qquad \cap \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}\right) \geq 1 - 2^{-N^\beta} \right\} \\
\mathcal{S}_{C_a} &= \left\{ i \in [N]: Z\left(K^i|K^{1:i-1}\right) < 1 - 2^{-N^\beta} \right\} \cup \\
&\quad \left\{ i \in [N]: 2^{-N^\beta} < Z\left(K^i|K^{1:i-1}, Y^{1:N}, U_1^{1:N}\right) < 1 - 2^{-N^\beta} \right\}
\end{aligned}
$$

By [5, Lemma 4.7] and the channel degradation, we have $\mathcal{F}_{S_a} \subseteq \mathcal{F}_{C_a}$, $\mathcal{I}_{C_a} \subseteq \mathcal{I}_{S_a}$ and $\mathcal{S}_{S_a} \subseteq \mathcal{S}_{C_a}$. In addition, we observe that the proportion $\frac{|\mathcal{S}_{C_a}\setminus\mathcal{S}_{S_a}|}{N} \to 0$, as $N \to \infty$.

Encoding: The encoder first applies lossy compression to $X^{1:N}$ with target distortion D_1 to obtain $U_1^{1:N}$, and treats $\left(X^{1:N}, U_1^{1:N}\right)$ as a joint source sequence to evaluate $K_{\mathcal{I}_{S_a}}$ by randomized rounding with respect to $P_{K^i|K^{1:i-1},X^{1:N},U_1^{1:N}}$. The encoder sends $K_{\mathcal{I}_{S_a}\setminus\mathcal{I}_{C_a}}$ to the decoders. $K_{\mathcal{F}_{S_a}}$ is determined uniformly from $\{0,1\}$ and pre-shared between the encoder and the decoders. For $i \in \mathcal{S}_{S_a}$, we use the maximum a posteriori (MAP) decision, i.e. $k^i = \arg\max_k P_{K^i|K^{1:i-1}}\left(k|k^{1:i-1}\right)$.

Decoding: Using the pre-shared $K_{\mathcal{F}_{S_a}}$ and received $K_{\mathcal{I}_{S_a}\setminus\mathcal{I}_{C_a}}$, Decoder 2 recovers $K_{\mathcal{I}_{C_a}}$ and $K_{\mathcal{S}_{S_a}}$ from the side information sequences $Y^{1:N}$ by SC decoding and the MAP rule, respectively. $K_{\mathcal{I}_{C_a}\cup\mathcal{S}_{S_a}}$ and $K_{\mathcal{S}_{C_a}\setminus\mathcal{S}_{S_a}}$ can be recovered with vanishing error probability, since their Bhattacharyya parameters are arbitrarily small when $N \to \infty$. Hence, we obtain $K^{1:N}$, and the reconstruction is given by $U_a^{1:N} = K^{1:N}G_N$.

Note that, from [7], $K_{\mathcal{S}_{C_a}}$ should be covered by a pre-shared random mapping. However, it is shown in [8, Theorem 2] that replacing the random mapping with MAP decision for $K_{\mathcal{S}_{S_a}}$ preserves the optimality. By the nested constructions, we can achieve $\frac{|\mathcal{I}_{S_a} \setminus \mathcal{I}_{C_a}|}{N} \xrightarrow{N\to\infty} I\left(X; U_a | U_1, Y\right)$. The encoding and decoding complexity is $O\left(N \log N\right)$.

For the second level, the encoder and Decoder 2 first recover $U_a^{1:N}$. Consequently, the encoder treats $\left(X^{1:N}, U_1^{1:N}, U_a^{1:N}\right)$ as a joint source, and Decoder 2 treats $\left(Y^{1:N}, U_1^{1:N}, U_a^{1:N}\right)$ as a joint side information. Likewise, according to $Y \leftrightarrow X \leftrightarrow (U_1, U_a, U_b)$, we have the test channel $T_{C_b} : U_b \to (Y, U_1, U_a)$ is degraded with respect to $T_{S_b} : U_b \to (X, U_1, U_a)$, and the rate of the second level approaches $I\left(X; U_b | U_1, U_a, Y\right) = I\left(U_b; X, U_1, U_a\right) - I\left(U_b; Y, U_1, U_a\right)$ by the same arguments. Hence, this coding scheme can achieve the optimal HBRDF, as long as the optimal parameters α, μ, θ, and θ_1 that achieve the minimum value of $S_{D_1}\left(\alpha, \mu, \theta, \theta_1\right)$ can be specified. Finally, we state the achievability of the HBRDF for DSBS for the entire Region I in the following theorem.

Theorem 7. *Consider target distortions $0 \leq D_1 < 0.5$ and $0 \leq D_2 < \min\left(D_1, p\right)$ for DSBS X when side information Y is available only at Decoder 2. For any $0 < \beta' < \beta < 0.5$ and any rate $R > \min S_{D_1}\left(\alpha, \mu, \theta, \theta_1\right)$, there exist a polar code $\mathcal{C}_1$ with rate $R_1 < I(X; U_1)$ and a two-level polar code $\mathcal{C}_2$ with rate $R_2 < I\left(X; U_2 | U_1, Y\right)$, with $R_1 + R_2 < R$, which together achieve the expected distortions $D_1 + O\left(2^{-N^{\beta}}\right)$ at Decoder 1 and $D_2 + O\left(2^{-N^{\beta'}}\right)$ at Decoder 2, respectively, if P_{X,U_1,U_2} is as given in Table I.*

Proof: The full proof is given in Theorem 10 in [11]. ■

IV. Polar Lattices for Gaussian Sources

It is shown in [8] that polar lattices can achieve the optimal rate-distortion performance for both the classical 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 the AWGN capacity achieving polar lattices [12] and the rate-distortion optimal ones [8]. Here we show that the HBRDF for the non-degenerate region specified in (2) can also be achieved by a similar nested code structure.

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

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

where $\mathbf{B}$ is the full rank generator matrix. The Voronoi region of Λ is defined by $\mathcal{V}\left(\Lambda\right) \triangleq \{z : Q_\Lambda\left(z\right) = 0\}$, where $Q_\Lambda\left(z\right) \triangleq \arg\min_{\lambda\in\Lambda} \|\lambda - z\|$ is the nearest-neighbor quantizer associated with Λ. A measurable set $\mathcal{R}\left(\Lambda\right) \subset \mathbb{R}^n$ is a fundamental region of Λ if $\cup_{\lambda\in\Lambda}\left(\mathcal{R}\left(\Lambda\right) + \lambda\right) = \mathbb{R}^n$ and if $\left(\mathcal{R}\left(\Lambda\right) + \lambda\right) \cap \left(\mathcal{R}\left(\Lambda\right) + \lambda'\right)$ has measure 0 for any $\lambda \neq \lambda'$ in Λ. For $\sigma > 0$ and $c \in \mathbb{R}^n$, the Gaussian distribution of variance σ^2 centered at c is defined as $f_{\sigma,c}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|x-c\|^2}{2\sigma^2}}$, $x \in \mathbb{R}^n$. Let $f_{\sigma,0}(x) = f_\sigma(x)$ for short. The Λ-periodic function is defined as $f_{\sigma,\Lambda}(x) = \sum_{\lambda\in\Lambda} f_{\sigma,\lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda\in\Lambda} e^{-\frac{\|x-\lambda\|^2}{2\sigma^2}}$. Note that, when x is restricted to the fundamental region $\mathcal{R}\left(\Lambda\right)$, $f_{\sigma,\Lambda}(x)$ is actually a probability density function (PDF) of the Λ-aliased Gaussian noise [13].

The flatness factor of a lattice Λ is defined as $\epsilon_\Lambda(\sigma) \triangleq \max_{x\in\mathcal{R}(\Lambda)} |V(\Lambda) f_{\sigma,\Lambda}(x) - 1|$, where $V(\Lambda) = |\det\left(\mathbf{B}\right)|$ denotes the volume of a fundamental region of Λ [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 Λ.

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

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

where $f_{\sigma,c}(\Lambda) = \sum_{\lambda\in\Lambda} f_{\sigma,c}(\lambda)$. 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 $n = 1$.

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 σ_X^2 and target distortion D_1 as in [8]. As a result, for a target distortion D_1 and any rate $R_1 > \frac{1}{2}\log\left(\sigma_X^2/D_1\right)$, there exists a polar lattice with rate R_1, such that the average distortion is asymptotically close to D_1 as blocklength is sufficiently large [8, Theorem 4]. Therefore, both decoders can recover U_1 and (U_1, Y) 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, $X' = Z_1 \sim \mathcal{N}\left(0, D_1\right)$ is Gaussian and independent of U_1 and Z. Let

$$\gamma \triangleq \frac{D_1\sigma_Z^2}{D_1\sigma_Z^2 - D_2\left(D_1 + \sigma_Z^2\right)},$$

and consider an auxiliary Gaussian random variable U' defined as $U' = X' + Z_4$, where $Z_4 \sim \mathcal{N}\left(0, \gamma D_2\right)$. Moreover, we define $Y' \triangleq Y - U_1 = X' + Z$ and $Y' \sim \mathcal{N}\left(0, D_1 + \sigma_Z^2\right)$. 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 $Z_3 \sim \mathcal{N}\left(0, \alpha\sigma_Z^2\right)$. We can also write $U' = \alpha Y' + Z_5$, where $Z_5 \sim \mathcal{N}\left(0, \gamma D_2 + \alpha\sigma_Z^2\right)$, which requires an AWGN capacity-achieving code from $\alpha Y'$ to U'.

The final reconstruction of Decoder 2 is given by $\hat{X}_2 = U_1 + \alpha Y' + \frac{1}{\gamma}\left(U' - \alpha Y'\right)$. Note that $\frac{1}{\gamma}\left(U' - \alpha Y'\right)$ is a scaled version of Z_5, which is independent of Y'. Thus, the variance of $\alpha Y' + \frac{1}{\gamma}\left(U' - \alpha Y'\right)$ is $\alpha D_1 + \frac{1}{\gamma^2}\left(\gamma D_2 + \alpha\sigma_Z^2\right) = D_1 - D_2$. Therefore, we have $X - U_1 = \hat{X}_2 - U_1 + \mathcal{N}\left(0, D_2\right)$ as we desired. Furthermore, the required data rate for Decoder 2 is then given by $R_2 > I\left(U'; X'\right) - I\left(U'; \alpha Y'\right) = \frac{1}{2}\log\left(\frac{D_1\sigma_Z^2}{D_2\left(D_1+\sigma_Z^2\right)}\right)$.

Note that U' is a continuous Gaussian random variable

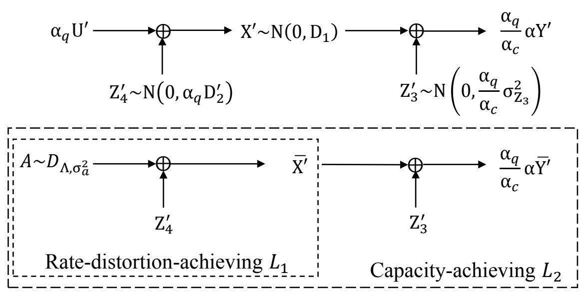

Fig. 4. A reversed solution of test channels to construct polar lattices.

which is impractical for the design of polar lattices. Hence, we use the discrete Gaussian distribution $D_{\Lambda,\sigma^2_{U'}}$ to replace it. Before that, we need to perform MMSE rescaling on U' for the test channels $X' \to U'$ and $Y' \to U'$ described in previous paragraphs with scales α_q and α_c, respectively. Consequently, a reversed version of the test channel can be derived, as depicted in Fig. 4, where $D'_2 \triangleq \gamma D_2$, $\alpha_q = \frac{D_1}{D_1+D'_2}$ and $\alpha_c = \frac{D_1^2}{(D_1+D'_2)(D_1+\sigma_Z^2)}$. The reconstruction of X at Decoder 2 is determined as $\hat{X}_2 = U_1 + \alpha_q U' + \eta\left(\frac{\alpha_q}{\alpha_c}\alpha Y' - \alpha_q U'\right)$, $\eta = \frac{D_2}{\sigma_Z^2}$, such that $X = \hat{X}_2 + \mathcal{N}(0, D_2)$.

Based on the reversed test channel, we can replace the continuous Gaussian random variable $\alpha_q U'$ with a discrete Gaussian distributed variable $A \sim D_{\Lambda,\sigma_a^2}$, where $\sigma_a^2 = \alpha_q^2 \sigma^2_{U'}$. Let $\bar{X}' \triangleq A + \mathcal{N}(0, \alpha_q D'_2)$ and $\frac{\alpha_q}{\alpha_c}\alpha\bar{Y}' = \bar{X}' + \mathcal{N}\left(0, \frac{\alpha_q}{\alpha_c}\sigma^2_{Z_3}\right)$. We also define $\bar{B} \triangleq \frac{\alpha_q}{\alpha_c}\alpha\bar{Y}'$, whose variance is $\sigma_b^2 = \frac{\alpha_q^2}{\alpha_c^2}\alpha^2\sigma^2_{Y'}$. By [8, Lemma 1], the distributions of $\bar{X}'$ and $\bar{Y}'$ can be made arbitrarily close to the distributions of X' and Y', respectively. Therefore, polar lattices can be designed for the source $\bar{X}'$ and the side information $\bar{Y}'$ at Decoder 2. Specifically, a rate-distortion 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 $A \to \frac{\alpha_q}{\alpha_c}\alpha\bar{Y}'$, as shown in Fig. 4. According to [8, Lemma 2], L_2 is nested within L_1, i.e., $L_2 \subseteq L_1$. The code constructions of L_1 and L_2 can be adapted from [8, Section V] and we omit them here for brevity. In the end, the reconstruction of Decoder 2 is given by $\check{X}_2 = U_1 + A + \eta\left(\bar{B} - A\right)$.

The rate for Decoder 1 is $R_1 \to \frac{1}{2}\log\left(\frac{\sigma_X^2}{D_1}\right)$ according to [8, Theorem 4]. The rate R_{L_1} of L_1 approaches $\frac{1}{2}\log\left(\frac{D_1}{\alpha_q D'_2}\right)$ when the flatness factor is negligible. By [12, Theorem 7], the rate R_{L_2} of L_2 approaches $\frac{1}{2}\log\left(\frac{\sigma_b^2}{\alpha_q D'_2 + \frac{\alpha_q}{\alpha_c}\sigma^2_{Z_3}}\right)$ with a negligible flatness factor. Since $L_2 \subseteq L_1$, the rate for Decoder 2 is $R_2 = R_{L_1} - R_{L_2}$, and the total rate is given by $R_1 + R_2 \to \frac{1}{2}\log\left(\frac{\sigma_X^2\sigma_Z^2}{D_2(D_1+\sigma_Z^2)}\right)$ which is the same as (2).

Next, we give the main theorem of the Gaussian Heegard-Berger problem for the non-degenerate region.

Theorem 8. *Let (X, Y, Z, D_1, D_2) be as specified in Section II-C. For any rate $R_1 > \frac{1}{2}\log\left(\frac{\sigma_X^2}{D_1}\right)$, there exists a polar lattice code at rate R_1 with sufficiently long blocklength, whose expected distortion is arbitrarily close to D_1. For any $0 < \beta' < \beta < 0.5$, there exist nested polar lattices L_1 and L_2 with $R_2 \triangleq R_{L_1} - R_{L_2}$ arbitrarily close to $\frac{1}{2}\log\left(\frac{D_1\sigma_Z^2}{D_2(D_1+\sigma_Z^2)}\right)$ such that the expected distortion D_{Q_2} satisfies $D_{Q_2} \leq D_2 + O\left(2^{-N^{\beta'}}\right)$.*

Proof: The full proof is given in Theorem 12 in [11]. ■

V. Discussion

We reserve for a future work the study of the optimality of polar codes and polar lattices for the Kaspi problem [15] which is a generalization of the Heegard-Berger problem, where the encoder may have access to the side information. The explicit rate-distortion functions for the Gaussian case and binary erasure case have been given in [16] and [17], respectively.

References

[1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[2] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, 1985.

[3] K. Kerpez, "The rate-distortion function of a binary symmetric source when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 448–452, May 1987.

[4] C. Tian and S. N. Diggavi, "A calculation of the Heegard-Berger rate-distortion function for a binary source," in *Proc. 2006 IEEE Inform. Theory Workshop*, Oct 2006, pp. 342–346.

[5] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.

[6] M. Ye and A. Barg, "Polar codes for distributed hierarchical source coding," *Advances in Mathematics of Communications*, vol. 9, no. 1, pp. 87–103, 2015.

[7] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[8] L. Liu and C. Ling, "Polar lattices for lossy compression," Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.05683

[9] S. Ramanan and J. M. Walsh, "Practical codes for lossy compression when side information may be absent," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 3048–3051, May 2011.

[10] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, 2010.

[11] J. Shi, L. Liu, D. Gündüz, and C. Ling, "Polar codes and polar lattices for the Heegard-Berger problem," Oct 2017. [Online]. Available: https://arxiv.org/abs/1702.01042v2

[12] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.0187

[13] G. D. Forney Jr., M. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[14] C. Ling and J.-C. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[15] A. H. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, 1994.

[16] E. Perron, S. N. Diggavi, and I. E. Telatar, "The Kaspi rate-distortion problem with encoder side-information: Gaussian case," Nov 2005. [Online]. Available: https://infoscience.epfl.ch/record/59938

[17] ——, "The Kaspi rate-distortion problem with encoder side-information: Binary erasure case," 2006. [Online]. Available: https://infoscience.epfl.ch/record/96000