Extracting Wyner’s Common Randomness Using Polar Codes

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

Abstract—Explicit construction of polar codes for the Gray-Wyner network is studied. We show that Wyner’s common information plays an essential role in constructing polar codes for both lossless and lossy Gray-Wyner coding problems. For discrete joint sources, extracting Wyner’s common randomness can be viewed as a lossy compression problem, which is accomplished by extending polar coding for a single source to a pair of joint sources with doubled alphabet size. We show that the lossless Gray-Wyner region can be achieved by an upgraded or degraded version of the common randomness. For joint Gaussian sources, we prove that extracting Wyner’s common randomness is equivalent to lossy compression for a single Gaussian source, meaning that Wyner’s common randomness can be extracted by utilizing polar lattice for quantization and represented by a lattice Gaussian random variable.

I. INTRODUCTION

In this paper we shall be concerned with extracting the common information contained in a pair of joint sources \((X, Y)\). There are different ways to characterize the amount of common information in literature. Apart from Shannon’s mutual information [1] and Gács and Körner’s common randomness [2], Wyner proposed an alternative definition to quantify the common information for \((X, Y)\) with finite alphabet [3] as

\[
C(X, Y) = \inf_{W} I(X; Y | W) - I(X; W) - I(Y; W),
\]

where the infimum is taken over all \(W\), such that \(X - W - Y\) forms a Markov chain.

Wyner’s definition was originated from his earlier work in the Gray-Wyner network [4] depicted in Fig. 1 which demonstrates Wyner’s first approach [3] to the explanation of \(C(X, Y)\). This network model contains an encoder that observes a pair of sequences \((X^N, Y^N)\) and outputs three messages \(W_0, W_1\) and \(W_2\) with rates \(R_0, R_1, R_2\) respectively. Decoder 1 reconstructs \(X_i^N\) by observing \((W_0, W_1)\) and decoder 2 reconstructs \(Y_i^N\) from \((W_0, W_2)\). Wyner also gave a second approach to explain the common information. In that model a common message \(W\) is sent to two independent processors. The processors generate output sequences separately according to distributions \(P_{X_i | W}(x_i | w)\) and \(P_{Y_i | W}(y_i | w)\). The output sequences \(X_i^N\) and \(Y_i^N\) have joint probability

\[
P_{X_i^N, Y_i^N}(x_i^N, y_i^N) = \sum_{w \in W} P_{X_i^N | W}(x_i^N | w) P_{Y_i^N | W}(y_i^N | w).
\]

Wyner showed that \(C(X, Y)\) is equal to the minimum rate on the shared message, with constraints that the sum of rates equals the joint entropy or that the joint distribution \(P_{X_i^N, Y_i^N}(x_i^N, y_i^N)\) is arbitrarily close to \(P_{X_i^N, Y_i^N}(\hat{x}_i^N, \hat{y}_i^N)\).

Wyner and Gács-Körner’s works on common information can be considered as two different viewpoints of the lossless Gray-Wyner region. Their works were then expanded by [5], [6] into the lossy case, where the output sequences \((\hat{X}_1^N, \hat{Y}_1^N)\) may have certain distortion.

Polar codes firstly proposed by Arikan in [7] have become widely studied due to their achievability of the Shannon bound with low complexity. Polar codes are also used for both lossy and lossless compression. For discrete sources, [8] provides constructions using polar codes for lossless and lossy compression. [9] gives a solution to the lossy compression for nonuniform sources. For memoryless Gaussian sources, [10] proposed a polar lattice construction to achieve the rate-distortion bound.

The use of polar codes for the common randomness (i.e. G point in Fig. 2) was recently proposed in [11]. This paper discussed polarization from the perspective of the maximal correlation of two discrete sources. Furthermore, it proved that polar codes are optimal to extract Wyner’s common randomness of discrete sources. However, we will investigate the entire best-known lossless Gray-Wyner region in [4], [5] for discrete sources. In addition, an explicit construction based on polar codes and polar lattice is given to achieve the lossy Gray-Wyner region [6] for both discrete and Gaussian sources when the distortion is small. We also give an explanation on the results in [6, Theorem 1] that the common information defined in lossless Gray-Wyner coding remains the same in lossy case when the distortion is small.

The paper is organized as follows: Section II presents the background of lossy and lossless compression using polar codes. The construction of polar codes for lossless Gray-Wyner network is investigated in Section III. In Section IV, we discuss polar codes and polar lattices for discrete sources and Gaussian sources for the lossy Gray-Wyner network. Finally, the paper is concluded in Section V.
Notations: All random variables (RVs) are denoted by capital letters. Let $P_X$ denote the probability distribution of a RV $X$ taking value $x$ in a set $X$. $X_N^i$ denotes a series of vectors $(X_1, ..., X_N)$. For a set $I$, $I^c$ denotes its complement, and $|I|$ represents its cardinality. $X_T$ denotes the subvector $\{X_i\}_{i \in T}$. For an integer $N$, $[N]$ will be used to denote the set of all integers from 1 to $N$. The information is measured in bits and $b(\cdot)$ denotes the binary entropy function.

II. POLAR CODES FOR SOURCE CODING

A. Polar Codes for Lossless Source Coding\(^9\)

Let $X_N^i$ be $N$ i.i.d. drawings of a RV $X$ which is a Bernoulli source with crossover probability $p$ (Ber($p$)), for $N = 2^n$ where an integer $n \geq 1$. Apply the polarizing transformation $U_N^i = X_N^i G_N$ where $G_N = G_2^\otimes n$, $\otimes$ denotes the Kronecker product and $G_2 = [1 \ 0]$. Fix $R > H(X)$ and let $F$ denote the frozen set such that $|F| = \lfloor NR \rfloor$ and $H(U_i|U_1^{-1}) \geq H(U_j|U_1^{-1})$ for all $i \in F$ and $j \notin F$. We denote the information set by $I = F^c$.

The Successive Cancellation (SC) encoder introduced in \(^[12]\) stores $u_F$ and computes $u_T$ following the encoding rules \(^9\) explained in Section II-B without considering the side information $Y_N^i$. If $u_i \neq u_i$ for $i \in I$, an estimation error occurs and the index $i$ needs to be announced to the decoder. The set of error indices is denoted by $T$. The encoder outputs $(u_F, T)$. The decoder puts $\hat{u_i} = u_i$ for $i \in F$, then estimates $u_T$ using the same rule to the SC encoder \(^9\). If $i \in T$, the decision is flipped. Then the decoder outputs $\hat{x}_N^i = \hat{u}_N^i G_N$. It has been shown in \(^[12]\) that the error rate tends to zero for any rate $R > H(X)$.

Since the entropy $H(U_i|U_1^{-1})$ is complicated to calculate when $N$ becomes very large, the Bhattacharyya parameter is often used. For source coding with side information, assume $(X, Y) \in \{0, 1\} \times \mathcal{Y}$ be a pair of RVs. The Bhattacharyya parameter \(^13\) is defined as $Z(X | Y) = 2 \sum_y P_Y(y) \sqrt{P_{X|Y}(0|y) P_{X|Y}(1|y)}$. Also $Z(X | Y)$ and $H(X | Y)$ are related by \(^13\) Proposition 2) which indicates that $H(X | Y)$ is near 0 if $1$ and only if $Z(X | Y)$ is near 0 or 1, respectively. Thus the parameters \{$Z(U_i|U_1^{-1})\}^N$ and \{$Z(U_i|U_1^{-1})\}^N$ polarize simultaneously. Furthermore \(^9\) shows $Z(U_i|U_1^{-1})$ is equal to the Bhattacharyya parameter of a symmetric channel defined in \(^7\). Therefore we can apply the method of constructing polar codes for symmetric channels \(^14\) to asymmetric channels.

B. Polar Codes for Lossy Source Coding\(^9\)

In this part, we discuss polar lossy source coding for a nonuniform source. We model the source as a sequence of i.i.d. realizations of a RV $Y \in \mathcal{Y}$. Let $X$ denote the reconstruction space. Let the distortion function denote by $d : \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}_+$. The rate-distortion function is given by $R(D) = \min_{e_{uv}(d(x,y)) \leq D} I(X, Y)$.

We assume an information set $I \subset [N]$ and a frozen set $I^c$. Similarly to lossless source coding, $U_N^i$ is defined as $U_N^i = X_N^i G_N$. The frozen set can be identified by $Z(U_i|U_1^{-1}, Y_N^i) \geq 1 - 2^{-N^\beta}$ or $Z(U_i|U_1^{-1}) \leq 2^{-N^\beta}$ for all $i \in I^c$. Then the information set satisfies $|I| = NR$, where $R$ is the encoding rate. From \(^9\) such an information set exists if $R > I(X, Y) = R(D)$, $\beta < 1/2$ and $N$ is sufficiently large.

After the indexes for information set and frozen set are identified, the encoder determines $u_1^i$ from a given source sequence $y_1^N$ by

\[
\begin{align*}
    u_i &= \begin{cases} 
        0, & \text{with probability } P_{U_i|U_1^{-1}, Y_N^i}(0 | u_1^{-1}, y_1^N) \\
        1, & \text{with probability } P_{U_i|U_1^{-1}, Y_N^i}(1 | u_1^{-1}, y_1^N)
    \end{cases}
\end{align*}
\]

for $i \in I$ and

\[
\begin{align*}
    u_i &= \begin{cases} 
        \hat{u}_i, & \text{if } Z(U_i|U_1^{-1}, Y_N^i) \geq 1 - 2^{-N^\beta} \\
        \arg \max_u P_{U_i|U_1^{-1}}(u | u_1^{-1}), & \text{if } Z(U_i|U_1^{-1}) \leq 2^{-N^\beta}
    \end{cases}
\end{align*}
\]

for $i \in I^c$, where $\hat{u}_i$ is determined beforehand uniformly from $\{0, 1\}$. Then the encoder sends $u_1$ to the decoder and the decoder outputs the reconstructed sequence $x_1^N = u_1^N G_N$.

Finally, from \(^9\) the average distortion results in $\sum_{i=1}^N d(y_i, x_i)$, where $d(x,y)$ denotes Hamming distance for discrete RVs or Euclidean distance for continuous RVs. Following that, we give a formal definition to this model.

Definition 1. An $(N, M_0, M_1, M_2)$ code is defined as follows: An encoder is a mapping $f_E : \mathcal{X}^N \times \mathcal{Y}^N \rightarrow I_{M_0} \times I_{M_1} \times I_{M_2}$, where $I_m = \{0, 1, 2, \ldots, M_m - 1\}$ for $i = 0, 1, 2$. A decoder is a pair of mappings $f_D^{(X)} : I_{M_0} \times I_{M_1} \rightarrow \mathcal{X}^N$ and $f_D^{(Y)} : I_{M_0} \times I_{M_2} \rightarrow \mathcal{Y}^N$. Let $f_E(X_1^N, Y_1^N) = (W_0, W_1, W_2)$, then $X_1^N = f_D^{(X)}(W_0, W_1)$ and $Y_1^N = f_D^{(Y)}(W_0, W_2)$. The average distortion between the inputs and outputs are $(\Delta_X, \Delta_Y)$, where $\Delta_X = \frac{1}{N} \sum_{i=1}^N d(X_i, \hat{X}_i)$ and $\Delta_Y = \frac{1}{N} \sum_{i=1}^N d(Y_i, \hat{Y}_i)$.

The achievable rate region based on $(N, M_0, M_1, M_2)$ code is defined as following.

Definition 2. For lossless coding, a triple $(R_0, R_1, R_2)$ is said to be achievable if there exists an $(N, M_0, M_1, M_2)$ code with $M_i \leq 2^{N(R_i+\epsilon)}$, $\epsilon = 0, 1, 2$ and $\Delta = \max(\Delta_X, \Delta_Y) \leq \epsilon$, for arbitrary $\epsilon > 0$ and sufficiently large $N$. Denote $R$ as the set of achievable rate.

Theorem 3. (\(^[12], Theorem 2\)) If $(R_0, R_1, R_2) \in R$, then $R_0 + R_1 + R_2 \geq H(X, Y)$, $R_0 + R_1 \geq H(X)$ and $R_0 + R_2 \geq H(Y)$.

Let us consider a pair of DSBS $(X, Y)$ where $X = \mathcal{Y} = \{0, 1\}$ and $Y = X \oplus Z$, $Z \sim Ber(a_0)$. In this case $H(X, Y) = \ldots$
This point can be trivially achieved since $H(X, Y) = H(X) + H(Y | X) = H(X) + H(Z)$ where $X$ does not need to be compressed. Encoder sends $x^N_1$ and compresses $z^N_1 = x^N_1 \oplus y^N_1$ as introduced in Section II-A. Decoder 1 and 2 reconstruct $\hat{x}^N_1$ and $\hat{z}^N_1$ with error probability tends to zero when $N$ is sufficiently large. Then $\hat{y}^N_1 = \hat{x}^N_1 \oplus \hat{z}^N_1$. Due to the symmetry the role of source $X$ and $Y$ can be exchanged.

- **Point G:**

  Firstly we extract the common information from source $(X_1^N, Y_1^N)$. Since $R_0 \geq I(W; XY) = 1 + h(a_0) - 2h(a_1)$, the encoder applies lossy compression to the joint sources $(x^N_1, y^N_1)$ as introduced in Section II-B with reconstruction $w^N_1$. Differently from traditional lossy compression with a single source, the test channel for the joint sources is $P_{XY|W}(xy|w) = P_{X|W}(x|w)P_{Y|W}(y|w)$.

  In this way the average distortion between $w^N_1$ and $x^N_1$ or $w^N_1$ and $y^N_1$ will tend to $a_1$ simultaneously. Hence the lossy-compressed sequence $u_2$ is the common message and is sent to both decoders, where $|I| \geq (1 + h(a_0) - 2h(a_1))N$. Since $H(X | W) = H(Y | W) = h(a_1)$, we apply lossless compression to source $x^N_1$ together with $w^N_1$ as side information and send the lossless compressed sequence privately to Decoder 1. Symmetrically source $Y$ is operated in the same way.

  At the decoder side, both decoders reconstruct $w^N_1$ from $u_2$ by the lossy decoding rule introduced in Section II-B. Then decoder 1 receives the compressed sequence from its private channel and derives a reconstructed sequence $\hat{x}^N_1$. Then the source can be reconstructed as $\hat{x}^N_1 = \hat{x}^N_1 \oplus w^N_1$. Decoder 2 can operate the same way. Therefore $X$ and $Y$ can be reconstructed with error rate tends to zeros when $N$ is sufficiently large. Notice that a similar method was also given in [11].

- **Points on dashed line AG:**

  On this line, the common branch carries more information than point G. To show how much additional amount of information to send over the common branch, we keep the relation that $X - W - Y$ is a Markov chain. However, if we move the intermediate variable $W$ closer to source $(X,Y)$, there will be two new RVs $(X', Y')$ between the source and common variable $W$ correspondingly. Hence we assume the test channel is a BSC($d_1$) between $X$ and $X'$ $(Y', Y')$, and a BSC($d_2$) between $W$ and $(X', Y')$ where $0 \leq d_1, d_2 \leq a_1$ as shown in Fig. 2(b). Therefore $X - X' - W - Y' - Y$ forms a Markov chain.

  In this case, the rate for the common channel is $R_0 \geq I(X'Y'; XY) = I(X'Y'W; XY) = I(XY; W) + I(X'; X|W) + I(Y'; Y|W) = 1 - 2h(d_1) + h(a_0)$. Instead of extracting the common variable $W$ for the common channel, decoders can reconstruct $(X', Y')$ and retrieve more information from the common channel because $(X', Y')$ is closer to

---

1. [1]
2. [2]
3. [3]
4. [4] Pangloss plane which constrains the triples $a_1, b, c$ to satisfy $\sum_{i=0}^2 R_i = H(X, Y)$. The dashed line AG in Fig. 3 is the Pangloss bound for the lossless Gray-Wyner coding problem. Point A refers to the case where only the common channel is used. Therefore the problem is the same as joint compression for source $(X,Y)$ and $(R_0, R_1, R_2) = (H(X,Y), 0, 0)$. Point G refers to the case where $R_0$ is the smallest rate that achieves lossless compression for source $(X,Y)$ and constrains the total rate equals $H(X, Y)$. In other words, $R_0$ achieves Wyner’s common information $C(X, Y)$ at point G and $(R_0, R_1, R_2) = (1 + h(a_0) - 2h(a_1), h(a_1), h(a_1))$. Next we demonstrate how to achieve the Pangloss bound using polar codes without time-sharing.

---

A. Polar codes for Pangloss bound

Theorem 3 indicates a fact that the Gray-Wyner network defined in Definition 2 cannot perform better than the situation where the receivers can collaborate. This situation refers to the Pangloss plane which constrains the triples $(R_0, R_1, R_2)$ to satisfy $\sum_{i=0}^2 R_i = H(X, Y)$. The dashed line AG in Fig. 3 is the Pangloss bound for the lossless Gray-Wyner coding problem. Point A refers to the case where only the common channel is used. Therefore the problem is the same as joint compression for source $(X,Y)$ and $(R_0, R_1, R_2) = (H(X,Y), 0, 0)$. Point G refers to the case where $R_0$ is the smallest rate that achieves lossless compression for source $(X,Y)$ and constrains the total rate equals $H(X, Y)$. In other words, $R_0$ achieves Wyner’s common information $C(X, Y)$ at point G and $(R_0, R_1, R_2) = (1 + h(a_0) - 2h(a_1), h(a_1), h(a_1))$. Next we demonstrate how to achieve the Pangloss bound using polar codes without time-sharing.

- **Point A:**

  This point can be trivially achieved since $H(X, Y) = H(X) + H(Y | X) = H(X) + H(Z)$ where $X$ does not need to be compressed. Encoder sends $x^N_1$ and compresses $z^N_1 = x^N_1 \oplus y^N_1$ as introduced in Section II-A. Decoder 1 and 2 reconstruct $\hat{x}^N_1$ and $\hat{z}^N_1$ with error probability tends to zero when $N$ is sufficiently large. Then $\hat{y}^N_1 = \hat{x}^N_1 \oplus \hat{z}^N_1$. Due to the symmetry the role of source $X$ and $Y$ can be exchanged.

- **Point G:**

  Firstly we extract the common information from source $(X_1^N, Y_1^N)$. Since $R_0 \geq I(W; XY) = 1 + h(a_0) - 2h(a_1)$, the encoder applies lossy compression to the joint sources $(x^N_1, y^N_1)$ as introduced in Section II-B with reconstruction $w^N_1$. Differently from traditional lossy compression with a single source, the test channel for the joint sources is $P_{XY|W}(xy|w) = P_{X|W}(x|w)P_{Y|W}(y|w)$.

  In this way the average distortion between $w^N_1$ and $x^N_1$ or $w^N_1$ and $y^N_1$ will tend to $a_1$ simultaneously. Hence the lossy-compressed sequence $u_2$ is the common message and is sent to both decoders, where $|I| \geq (1 + h(a_0) - 2h(a_1))N$. Since $H(X | W) = H(Y | W) = h(a_1)$, we apply lossless compression to source $x^N_1$ together with $w^N_1$ as side information and send the lossless compressed sequence privately to Decoder 1. Symmetrically source $Y$ is operated in the same way.

  At the decoder side, both decoders reconstruct $w^N_1$ from $u_2$ by the lossy decoding rule introduced in Section II-B. Then decoder 1 receives the compressed sequence from its private channel and derives a reconstructed sequence $\hat{x}^N_1$. Then the source can be reconstructed as $\hat{x}^N_1 = \hat{x}^N_1 \oplus w^N_1$. Decoder 2 can operate the same way. Therefore $X$ and $Y$ can be reconstructed with error rate tends to zeros when $N$ is sufficiently large. Notice that a similar method was also given in [11].
(X, Y) than W. Then both sources can be losslessly reconstructed by applying lossless compression to source (X, Y) with side information (X', Y'). Therefore the rate of the private channels is \( R_1 = R_2 \geq H(X|X') = H(Y'|Y') = h(d_1) \).

Now we show a construction using polar codes that achieves the rate bound. Similarly to G point, we can firstly retrieve the common message \( w_1^N \) by applying lossy compression to joint source sequences \( (x_1^N, y_1^N) \). The compressed rate approaches \( 1 + h(a_0) - 2h(a_1) \). Notice that \( Z_1 = X \oplus W \) is a Ber\((a_1)\) source. Then the encoder applies lossless compression to the nonuniform source \( x_1^N \oplus w_1^N \) with distortion \( d_1 \) as introduced in Section [II-B]. We operate the same to source \( Y \). Thus the additional rate approaches \( 2(h(a_1) - h(d_1)) \) when \( N \) is sufficiently large. For the private channels, the encoder first reconstructs \( X' \) and \( Y' \). Then the encoder applies lossless compression to source \( X \oplus X' \) and \( Y \oplus Y' \) with rate approaching \( h(d_1) \). It is known that polar codes are optimal for lossy and lossless compression [9], [13]. Therefore the average distortion for source \( (X_1^N, Y_1^N) \) approaches zero when \( N \) is sufficiently large.

### B. Polar codes for \( R_0 < C(X,Y) \) (Curve GB)

In this part we explain how to achieve the dashed curve GB in Fig. 3 using polar codes. From Theorem 3 the lower boundary of \( R \) should lie above the lines \( R_0 + 2R_1 = 1 + h(a_0) \) and \( R_0 + R_1 = 1 \). In the above section, we have explained polar code constructions to achieve this lower boundary of \( R \) called Pangloss bound when \( R_0 \geq C(X,Y) \) (e.g. dashed line AG in Fig. 3). However the lower boundary of \( R \) remains unknown when \( R_0 < C(X,Y) \). To the best of our knowledge, the tightest lower boundary was given in [4] for the case as follows:

Consider a degradation applied on the common variable \( W \), where \( W \) is considered the input to BSC(\( \rho \)) with output \( W' \) as shown in Fig. 2(C). As a result \( (X, Y) - W - W' \) forms a Markov chain and \( a_1 \leq \beta = a_1 \cdot \rho \leq 1/2 \). From this Markov chain, the transition probability reads

\[
P_{XY|W'}(x; y|w') = \sum_{w \in W} P_{W|W'}(w|w') P_{X|W}(x|w) P_{Y|W}(y|w'),
\]

(2)

Then the triple \( (R_0, R_1, R_2) \in R \) satisfies

\[
\begin{align*}
R_0 & = I(W'; XY) = (1 - a_0) (1 - h(\beta - 0.5a_0)) \\
R_1 & = R_2 = H(X|W') = H(Y'|W') = h(\beta).
\end{align*}
\]

(3)

As \( \beta \in [a_1, \frac{1}{2}] \), the family of rate triples can generate the dashed curve GB in Fig. 3. Achieving the triple in (3) is quite similar to the encoding and decoding for point G using polar codes. Firstly we apply polar lossy compression to joint sources \( (X_1^N, Y_1^N) \) with distortion \( \beta \) and derive reconstruction \( W' \). The test channel used in lossy compression is specified in [2], which is the major difference from the construction of point G. Then send the compressed sequence over the common channel. After that, we apply lossless compression to source \( X \) and \( Y \) with \( W' \) as side information, and send the compressed sequences through private channels to decoder 1 and 2 accordingly. Alternatively we can derive the common randomness \( W \) by lossy compression of \( (X, Y) \). After that, we apply symmetric lossy compression to \( W \) with distortion \( \rho \) to obtain \( W' \). Moreover it is trivial to achieve point \( B \) when \( R_1 = R_2 = 1 \) and \( R_0 = 0 \).

Together with the result from [III-A] all points from point A to B along the dashed line in Fig. 3 can be achieved by polar codes. Moreover the above models can be extended to achieving the lower boundary for the more general sources mentioned in [15].

### IV. POLAR CODES FOR LOSSY GRAY-WYNER CODING

#### A. Common Information Preserves in Small Distortion Region

In this section we show how to achieve the common information \( C(\Delta_1, \Delta_2) \) which is defined as the smallest common rate \( R_0 \) such that the total rate meets the rate-distortion bound. Based on Definition 1 we define the lossy Gray-Wyner coding as follows:

The rate-distortion function for source \( (X, Y) \) is

\[
R_{XY}(\Delta_1, \Delta_2) = \min I(X', Y'; X, Y),
\]

where the minimum is taken over all test channels \( P_{X', Y'; XY}(x'y'|xy) \) such that \( Ed(X', X) \leq \Delta_1 \) and \( Ed(Y', Y) \leq \Delta_2 \).

**Definition 4.** For any \( \Delta_1, \Delta_2 \geq 0 \), a number \( R_0 \) is said to be \( (\Delta_1, \Delta_2) \)-achievable if for any \( \varepsilon > 0 \) we can find a sufficiently large \( N \) such that there exists a \( (N, M_0, M_1, M_2) \) code with \( M_0 \leq \frac{N^{R_0}}{2} \), \( \sum_{i=0}^{N} \frac{1}{2} \log M_i \leq R_{XY}(\Delta_1, \Delta_2) + \varepsilon \), \( \Delta_X \leq \Delta_1 + \varepsilon \), and \( \Delta_Y \leq \Delta_2 + \varepsilon \). Then \( C(\Delta_1, \Delta_2) \) is defined as the infimum of all \( R_0 \) that is \( (\Delta_1, \Delta_2) \)-achievable.

Following that we provide an operational meaning to the conclusion of [6] Theorem 1 that \( C(\Delta_1, \Delta_2) = C(X,Y) \) in some neighborhood of the origin \( \{(\Delta_1, \Delta_2): 0 \leq \Delta_1, \Delta_2 \leq \gamma \} \). Again we use DSBS sources \( (X,Y) \) as an example.

The relation between lossy and lossless Gray-Wyner coding is not difficult to find if we recall the construction of the line AG in Section [III-A] We apply the same test channel where \( X \rightarrow X' \rightarrow W \rightarrow Y' \rightarrow Y \) forms a Markov chain shown in Fig. 2(b). The rate of the shared branch in lossless Gray-Wyner coding equals \( I(X'Y'; XY) = I(X;X') + I(X';X) + I(Y';Y) \). Thus we send three sub-sequences on this branch.

Notice that if we consider \( (X', Y') \) as the output of the decoders, the above rate is sufficient for lossy Gray-Wyner coding. In the lossy case, we only require to recover the source \( (X, Y) \) with distortions \( (\Delta_1, \Delta_2) \). So we consider the intermediate RVs \( (X', Y') \) as the reconstruction variables.

Again, as in Section [III-A] we only consider the plane in the \( (R_0, R_1, R_2) \) space where \( R_1 = R_2 \) and \( \Delta_1 = \Delta_2 = \Delta \). The encoder applies the same lossy compression as that to the line AG to extract \( W \) and sends on the shared branch with rate \( R_0 \geq I(XY; W) \). Then the encoder applies asymmetric lossy compression to source \( X + W \) with distortion \( \Delta \) where
\[ a_1 = d_2 \triangle. \] To achieve \( \triangle_1 = \triangle_2 = \triangle \), the additional rate we should send is \( I(X'; X \mid W) \) and \( I(Y'; Y \mid W) \) over either private channels or the shared channel. Then the distortion between \( X(Y) \) and \( X'(Y') \) tends to \( \triangle \) when \( N \) is sufficiently large. As a result the total rate \( \sum_{i=0}^{2} R_i = I(X|Y'; X'Y) = 1 + h(a_0) - 2h(\triangle) = R_{XY}(\triangle, \triangle), \) for \( 0 \leq \triangle \leq a_1 \). This indicates that the lossy Pangloss bound \([4]\) can be achieved when \( C(\triangle, \triangle) = C(X, Y) \) as long as \( 0 \leq \triangle \leq a_1 \).

\section{B. Common Information for Joint Gaussian Sources}

The common information was generalized to the joint Gaussian sources in \([1]\). Let \( X, Y \) be two bivariate Gaussian variables with zero mean and covariance matrix \([\rho 1]\). The common variable of \((X, Y)\) is described by a Gaussian variable \( W \) with mean \( 0 \) and variance \( \rho \) such that

\[
\begin{cases}
X = W + \sqrt{1-\rho} N_1 \\
Y = W + \sqrt{1-\rho} N_2
\end{cases}
\]

(4)

where \( N_1 \) and \( N_2 \) are standard Gaussian variables and \( N_1, N_2 \) are independent of each other. Clearly, the common information is given by \( I(X, Y; W) = \frac{1}{2} \log \frac{1+\rho}{1-\rho} \).

However, extracting a continuous common message from \((X, Y)\) is difficult. We show that a discrete version of \( W \) is also eligible to convey the common message of two joint Gaussian variables according to Wyner’s second approach to the characterization of common information \([3]\).

An \( n \)-dimensional lattice is a discrete subgroup of \( \mathbb{R}^n \) which can be described by

\( \Lambda = \{ \lambda = Bz : z \in \mathbb{Z}^n \} \),

where \( B \) the full rank generator matrix. For \( \sigma > 0 \) and \( \epsilon \in \mathbb{R}^n \), the Gaussian distribution of variance \( \sigma^2 \) centered at \( \epsilon \) is defined as

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

Let \( f_{\sigma, 0}(x) = f_{\sigma}(x) \) for short.

The \( \Lambda \)-periodic function is defined as

\[ f_{\sigma, \Lambda}(x) = \sum_{\lambda \in \Lambda} e^{-\frac{|x-\lambda|^2}{2\sigma^2}}. \]

We note that \( f_{\sigma, \Lambda}(x) \) is a probability density function (PDF) if \( x \) is restricted to the fundamental region \( R(\Lambda) \). It is actually the PDF of the \( \Lambda \)-aliased Gaussian noise \([16]\).

The flatness factor of a lattice \( \Lambda \) is defined as \([16]\)

\[ \epsilon_{\Lambda}(\sigma) \triangleq \max_{x \in R(\Lambda)} |V(\Lambda) f_{\sigma, \Lambda}(x) - 1|, \]

where \( V(\Lambda) = |\det(B)| \) denotes the volume of a fundamental region of \( \Lambda \). 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 \( \epsilon \) as the discrete distribution taking values in \( \lambda \in \Lambda \) as

\[ D_{\Lambda, \sigma, \epsilon}(\lambda) = f_{\sigma, \epsilon}(\lambda) \]

(5)

for \( \forall \lambda \in \Lambda \), where \( f_{\sigma, \epsilon}(\Lambda) = \sum_{\lambda \in \Lambda} f_{\sigma, \epsilon}(\lambda) \). For convenience, we write \( D_{\Lambda, \sigma} = D_{\Lambda, \sigma, 0} \). It has been shown in \([17]\) 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 \( \epsilon = 0 \) and \( n = 1 \) (one-dimensional lattice \( \Lambda \)) in this work.

\textbf{Lemma 5.} Let \( W \) be a \( \text{RV which follows a discrete Gaussian distribution} \( D_{\Lambda, \sqrt{\rho}} \). Consider two continuous variables \( X \) and \( Y \)

\[
\begin{aligned}
\{ \bar{X} = W + \sqrt{1-\rho} N_1 \\
\bar{Y} = W + \sqrt{1-\rho} N_2
\end{aligned}
\]

where \( N_1 \) and \( N_2 \) are the same as that given in \([4]\). Let \( f_{X,Y}(x,y) \) and \( f_{X,Y}(x,y) \) denote the joint PDF of \((X, \bar{Y})\) and \((X, Y)\), respectively. If \( \epsilon = \epsilon_{\Lambda}(\sqrt{\frac{(1-\rho)}{1+\rho}}) < \frac{1}{2} \), the statistical distance between \( f_{X,\bar{Y}}(x,y) \) and \( f_{X,Y}(x,y) \) is upper-bounded by

\[
\int_{\mathbb{R}^2} |f_{X,\bar{Y}}(x,y) - f_{X,Y}(x,y)| \ dx \ dy \leq 4\epsilon,
\]

and the mutual information \( I(X, \bar{Y}; W) \) satisfies

\[
I(X, \bar{Y}; W) \geq \frac{1}{2} \log \frac{1+\rho}{1-\rho} - 5\epsilon \log(e).
\]

According to Wyner’s second approach, \( W \) is an eligible candidate of the common message of \((X, Y)\) when \( \epsilon \to 0 \).

\textbf{Proof:} Since \( \bar{X} - \bar{W} - \bar{Y} \) is a Markov chain, we have

\[ f_{X,\bar{Y}}(x,y) = \sum_{a \in \Lambda} f_{X,Y}(x,y,a) \]

(6)

\[ = \sum_{a \in \Lambda} f_{\bar{X}|\bar{Y}}(x|a) f_{\bar{Y}|\bar{W}}(y|a) \]

\[ = \frac{1}{f_{\sqrt{\rho}}(\Lambda)} \sum_{a \in \Lambda} \frac{1}{\sqrt{2\pi\rho}} \exp\left( -\frac{a^2}{2\rho} \right) \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\left( -\frac{(y-a)^2}{2(1-\rho)} \right) \]

\[ \cdot \exp\left( -\frac{(x-a)^2}{2(1-\rho)} \right) \]

\[ = \frac{1}{2\pi\sqrt{1-\rho^2}} \sum_{a \in \Lambda} \frac{1}{\sqrt{2\pi\rho}} \exp\left( -\frac{(a - \rho(x+y))^2}{2(1+\rho)} \right) \]

\[ \cdot f_{\sqrt{\rho}}(\Lambda) \]

(7)

where \( \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left( -\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)} \right) = f_{X,Y}(x,y) \) is the PDF of two joint Gaussian variables. By the definition of the flatness factor, we have
Since \( \epsilon_A(\sigma) \) is a monotonically decreasing function of \( \sigma \) (see [18, Remark 2]), we have \( \epsilon(\sqrt{\rho}) \leq \epsilon \) and hence

\[
|V(\Lambda) f_{\sqrt{\rho}}(\Lambda) - 1| \leq \epsilon. \tag{7}
\]

Combining (5), (6) and (7) gives us

\[
f_{X,Y}(x, y)(1 - 2\epsilon) \leq f_{X,Y}(x, y) \cdot \frac{1 - \epsilon}{1 + \epsilon} \leq f_{\tilde{X}, \tilde{Y}}(x, y),
\]

and

\[
f_{\tilde{X}, \tilde{Y}}(x, y) \leq f_{X,Y}(x, y) \cdot \frac{1 + \epsilon}{1 - \epsilon} \leq f_{X,Y}(x, y)(1 + 4\epsilon),
\]

when \( \epsilon < \frac{1}{2} \). Finally,

\[
\int_{\mathbb{R}^2} |f_{\tilde{X}, \tilde{Y}}(x, y) - f_{X,Y}(x, y)| \, dx \, dy \\
\leq 4\epsilon \int_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx \, dy = 4\epsilon.
\]

Similarly, the Kullback-Leibler divergence between \( f_{\tilde{X}, \tilde{Y}}(x, y) \) and \( f_{X,Y}(x, y) \) can be upper-bounded as

\[
\mathbb{D}(f_{\tilde{X}, \tilde{Y}} \| f_{X,Y}) = \int_{\mathbb{R}^2} f_{\tilde{X}, \tilde{Y}}(x, y) \log \frac{f_{\tilde{X}, \tilde{Y}}(x, y)}{f_{X,Y}(x, y)} \, dx \, dy \\
\leq \int_{\mathbb{R}^2} f_{\tilde{X}, \tilde{Y}}(x, y) \log(1 + 4\epsilon) \, dx \, dy \\
= \log(1 + 4\epsilon) \tag{8}
\]

For any \( \sqrt{\frac{\rho(1 - \rho)}{1 + \rho}} > 0 \), \( \epsilon \) can be made arbitrarily small by scaling \( \Lambda \). Therefore, when \( \epsilon \to 0 \), \( \tilde{W} \) can be viewed as the common message according to Wyner’s second approach. To see that \( I(\tilde{X}, \tilde{Y}; \tilde{W}) \) can be arbitrarily close to the common information, we rewrite \( \mathbb{D}(f_{\tilde{X}, \tilde{Y}} \| f_{X,Y}) \) as

\[
\mathbb{D}(f_{\tilde{X}, \tilde{Y}} \| f_{X,Y}) \\
= \int_{\mathbb{R}^2} f_{\tilde{X}, \tilde{Y}}(x, y) \log f_{\tilde{X}, \tilde{Y}}(x, y) \, dx \, dy \\
= - \int_{\mathbb{R}^2} f_{\tilde{X}, \tilde{Y}}(x, y) \log f_{X,Y}(x, y) \, dx \, dy - h(\tilde{X}, \tilde{Y}) \\
= - \int_{\mathbb{R}^2} f_{\tilde{X}, \tilde{Y}}(x, y) \log \left( \frac{1}{2\pi \rho} \right) \\
\cdot \exp \left( - \frac{x^2 + y^2 - 2\rho xy}{2(1 - \rho^2)} \right) \, dx \, dy - h(\tilde{X}, \tilde{Y}) \\
= \log \left( 2\pi \sqrt{1 - \rho^2} \right) + \frac{E_{\tilde{X}, \tilde{Y}}[x^2 + y^2 - 2\rho xy]}{2(1 - \rho^2)} \log(e) - h(\tilde{X}, \tilde{Y}) \\
= \log \left( 2\pi \sqrt{1 - \rho^2} \right) + \frac{1 + E_{\tilde{W}}[w^2]}{1 + \rho} \log(e) - h(\tilde{X}, \tilde{Y}).
\]

Note that \( \epsilon_A(\sqrt{\rho}) \leq \epsilon \). By [17, Lemma 5] and [17, Remark 3], it is easy to make \( E_{\tilde{W}}[w^2] \geq \rho(1 - 2\epsilon) \). Then we have

\[
\mathbb{D}(f_{\tilde{X}, \tilde{Y}} \| f_{X,Y}) \geq \log \left( 2\pi \sqrt{1 - \rho^2} \right) + (1 - \epsilon) \log(e) - h(\tilde{X}, \tilde{Y}) \\
= h(X, Y) - h(X, \tilde{Y}) - \epsilon \log(e).
\]

Using (8), we obtain

\[
I(X, Y; W) - I(\tilde{X}, \tilde{Y}; \tilde{W}) = h(X, Y) - h(\tilde{X}, \tilde{Y}) \\
\leq \log(1 + 4\epsilon) + \epsilon \log(e) \\
\leq 5\epsilon \log(e).
\]

Similar to [10], using \( D_{\Lambda, \sqrt{\rho}} \) as the reconstruction distribution, we can design quantization polar lattices to extract the common randomness according to “Construction D”. The only difference is that the size of the source alphabet is doubled in this work. The next theorem shows that the design of polar lattices for extracting the common randomness of a pair of joint Gaussian sources is exactly the same as that for quantizing a Gaussian source, which means that the technique proposed in [10] can be directly employed to our work.

**Theorem 6.** The construction of a polar lattice for extracting the common randomness of a pair of joint Gaussian sources \((X, Y)\) is equivalent to the construction of a rate-distortion bound achieving polar lattice for a Gaussian source \( X + Y/2 \).

**Proof:** Let \( \tilde{W} \) be labelled by bits \( \tilde{W}_1, \ldots, \tilde{W}_r (\tilde{W}_1') \) according to a binary partition chain \( \Lambda/\Lambda_1/ \cdots /\Lambda_{r-1}/\Lambda_r \). Then, \( D_{\Lambda, \sqrt{\rho}} \) induces a distribution \( P_{\tilde{W}_1} \) whose limit corresponds to \( D_{\Lambda, \sqrt{\rho}} \) as \( r \to \infty \).

By the chain rule of mutual information,

\[
I(\tilde{X}, \tilde{Y}; W_1') = \sum_{\ell=1}^{r} I(\tilde{X}, \tilde{Y}; W_\ell | W_1^{\ell-1}),
\]

we obtain \( r \) binary-input channels \( V_\ell \) for \( 1 \leq \ell \leq r \). Given the realization \( w_1^{\ell-1} \) of \( W_1^{\ell-1} \), denote \( A_\ell(w_1^{\ell-1}) \) the coset of
\( \Lambda_\ell \) indexed by \( w_1^{\ell-1} \) and \( w_\ell \). According to [19], the channel transition PDF of the \( \ell \)-th channel \( V_\ell \) is given by

\[
f_{X_\ell, Y_\ell | \bar{W}_\ell, \bar{W}^{\ell-1}_1}(x, y | w_\ell, w_1^{\ell-1}) = \frac{1}{\sqrt{2\pi\rho}} \sum_{a \in A_\ell(w_\ell)} f_{Y_\ell} (a) f_{X_\ell | \bar{W}_\ell}(x | y; a)
\]

\[
= \frac{1}{\sqrt{2\pi\rho}} \sum_{a \in A_\ell(w_\ell)} \frac{1}{\sqrt{2\pi\rho}} \exp \left( -\frac{(a - \rho)^2}{2\rho} \right) \frac{1}{\sqrt{2\pi(1-\rho)}} \exp \left( -\frac{(y - a)^2}{2(1-\rho)} \right)
\]

\[
\cdot \frac{1}{\sqrt{2\pi(1-\rho)}} \exp \left( -\frac{(a - \rho x + y - 2\rho xy)^2}{2\rho^2(1-\rho)} + \frac{1}{\rho} \right)
\]

Using the channel symmetrization technique for asymmetric channels [10], the channel transition PDF of the symmetrized channel \( V'_\ell \) is

\[
f_{X'_\ell}(y, y, w_1^{\ell-1}, w_\ell \oplus \bar{w}_\ell | \bar{W}_\ell) = \frac{f_{X_\ell, Y_\ell | \bar{W}_\ell, \bar{W}^{\ell-1}_1}(x, y | w_\ell, w_1^{\ell-1})}{f_{X'_\ell}(y, y, w_1^{\ell-1}, w_\ell \oplus \bar{w}_\ell | \bar{W}_\ell)}
\]

\[
= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left( -\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)} \right) \frac{1}{\sqrt{2\pi(1-\rho)}} \exp \left( -\frac{(a - \rho^2 x + y - 2\rho xy)^2}{2\rho^2(1-\rho)} + \frac{1}{\rho} \right)
\]

Comparing with the \( \Lambda_{\ell-1} / \Lambda_\ell \) channel [20 eqn. (13)], we see that the symmetrized channel (9) is equivalent to a \( \Lambda_{\ell-1} / \Lambda_\ell \) channel with noise variance \( \frac{\rho^2(1-\rho)}{1+\rho} \) in the sense of the likelihood ratio.

Recall that \( X, Y \) are bivariate Gaussian with zero mean and covariance matrix \( \begin{bmatrix} 1 & \rho \end{bmatrix} \). It is well known that \( \Lambda_{\ell-1} / \Lambda_\ell \) is Gaussian with zero mean and variance \( \frac{\rho^2(1-\rho)}{1+\rho} \). Now consider the construction of a polar lattice to quantize \( X + Y \) using the reconstruction distribution \( D_{\Lambda, \sqrt{\rho}} \). The MMSE re-scaled coefficient and noise variance are given by \( \frac{2\rho}{1+\rho} \) and \( \frac{\rho^2(1-\rho)}{1+\rho} \), which is the same as that in [6].

V. Conclusion

Explicit construction of polar codes and polar lattices for both lossy and lossless Gray-Wyner network coding is proposed. For a joint Gaussian source, we show that a rate-distortion bound achieving polar lattice designed for a Gaussian source can be directly used to extract the common randomness. For future work, we will extend the result of Theorem [6] into multiple joint Gaussian RVs. The common information of multiple Gaussian sources has been given in [6].

REFERENCES

[1] C. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, July 1948.
[2] P. Gács and J. Körner, “Common information is far less than mutual information,” Problems of Control and Information Theory, vol. 2, no. 2, pp. 149–162, 1973.
[3] A. D. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, March 1975.
[4] R. Gray and A. Wyner, “Source coding for a simple network,” Bell System Technical Journal, vol. 53, no. 9, pp. 1681–1721, 1974.
[5] K. B. Viswanatha, E. Akyol, and K. Rose, “The lossy common information of correlated sources,” Information Theory, IEEE Transactions on, vol. 60, no. 6, pp. 3238–3253, 2014.
[6] G. Xu, W. Liu, and B. Chen, “Wyner’s common information for continuous random variables - A lossy source coding interpretation,” in Proc. Annual Conference on Information Sciences and Systems, Baltimore, MD, March 2011.
[7] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.
[8] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2009.
[9] 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.
[10] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, “Semantically secure coset codes and multilevel coset codes,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.
[11] N. Goela, “Polarized random variables: Maximal correlations and common information,” in Proc. 2014 IEEE Int. Symp. Inform. Theory, Honolulu, USA, June 2014, pp. 1643–1647.
[12] H. S. Cronie and S. B. Korada, “Lossless source coding with polar codes,” in Proc. 2010 IEEE Int. Symp. Inform. Theory, Austin, TX, June 2010, pp. 904–908.
[13] E. Arikan, “Source polarization,” 2010 IEEE International Symposium on Information Theory Proceedings, 2010.
[14] I. Tal and A. Vardy, “How to construct polar codes,” IEEE Trans. Inf. Theory, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.
[15] H. S. Witsenhausen, “Values and bounds for the common information of two discrete random variables,” SIAM Journal on Applied Mathematics, vol. 31, no. 2, pp. 313–333, 1976.
[16] G. D. Forney Jr., M. Trotti, 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.
[17] C. Ling and J. Belfiore, “Achieving AWGN channel capacity with lattice Gaussian coding,” IEEE Trans. Inf. Theory, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.
[18] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, “Semantically secure lattice codes for the Gaussian wiretap channel,” IEEE Trans. Inf. Theory, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.
[19] U. Wachsmann, R. Fischer, and J. Huber, “Multilevel codes: Theoretical concepts and practical design rules,” IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1361–1391, July 1999.
[20] 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