2022 IEEE the 14th International Conference on Wireless Communications and Signal Processing

Optimality of Polar Codes in Additive Steganography under Constant Distortion Profile

Qiyi Yao, Weiming Zhang, and Nenghai Yu
School of Cyber Science and Technology
University of Science and Technology of China
Hefei, China
zhangwm@ustc.edu.cn

2022 14th International Conference on Wireless Communications and Signal Processing (WCSP) | 978-1-6654-5085-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/WCSP55476.2022.10039228

***Abstract*—Steganography aims to hide information in cover media where steganographic coding acts as a vital part. None of the existing steganographic coding algorithms have been proved to be optimal so far. Recently, capacity-achieving polar codes have been used to devise steganographic coding algorithms which are evaluated by experimental simulations without theoretical analysis. In this paper, we prove that under the constant distortion profile, polar codes-based steganographic coding methods could achieve the theoretical rate-distortion bound for additive steganography when the code length goes to infinity.**

***Index Terms*—Steganography, polar codes, Bhattacharyya parameters, BSC, channel capacity.**

I. Introduction

Steganography is an information hiding technique [1] which aims to embed a secret message into a cover object with slight modifications so that the stego object would not be distinguished by a passive adversary [2]. The commonly used steganographic scheme is called adaptive steganography [3] which can be divided into two steps: distortion calculation step and steganographic coding step. In the distortion calculation step, a distortion function denoting the distortion introduced by each cover element modification is defined which will be an input in the steganographic coding step. We only consider the *constant distortion profile* where the modification distortion of each cover element is the same in this paper. A distortion function is called *additive* when the modification of each element in cover media is considered to be independent, i.e., the total distortion after information embedding is the sum of the modification distortion of each element. The aim of the steganographic coding step is to embed the secret message while minimizing the total distortion.

Adaptive steganographic coding is defined as the coding on the adaptive distortion model where the modification distortion of the elements in cover media could be any positive real number and could be different from each other. The early steganographic coding methods aimed to minimize the number of modifications [4]–[7], which corresponds to the constant distortion profile in the context of adaptive steganography. For general adaptive steganographic coding, there are two practical near-optimal codes so far: Syndrome-Trellis Codes (STCs) [8] and Steganographic Polar Codes (SPCs) [9]. STC is designed using linear convolutional codes as well as the Viterbi decoding algorithm and SPC is designed using polar codes [10] as well as its corresponding Successive Cancellation List (SCL) decoding algorithm [11].

Designing provably optimal codes has always been an essential problem in information theory and its relevant application fields including steganography. Although researchers have devised a lot of good steganographic codes, none of which have been proved to be optimal so far, even under the constant distortion profile: they have all been evaluated by experimental simulations. Since the appearance of capacity-achieving polar codes and SPC, we could partially solve the hard problem. In this paper, we prove that polar codes-based steganographic coding methods could achieve the theoretical rate-distortion bound for additive steganography under the constant distortion profile when the code length goes to infinity.

This work was supported in part by the Natural Science Foundation of China under Grant 62002334, 62072421, and 62121002, Anhui Science Foundation of China under Grant 2008085QF296, and by Anhui Initiative in Quantum Information Technologies under Grant AHY150400.
(Corresponding author: Weiming Zhang.)

II. Preliminaries

In this section, we introduce the basic knowledge of adaptive steganographic coding and the polar codes-based steganographic coding method SPC. Random variables will be typeset in capital letters while their corresponding realizations will be in lower case throughout the paper.

A. Adaptive Steganographic Coding

We use $\boldsymbol{X}$ to denote the cover random variable and use $\boldsymbol{Y}$ to denote the stego random variable after embedding. The realization $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ is used to denote the cover vector with length N, realization $\boldsymbol{y} = (y_1, y_2, ..., y_N)$ is used to denote the stego vector and $\boldsymbol{m}$ with length m is used to denote the secret message. We only consider the additive *binary embedding* situation in this paper, thus, the three vectors are carried in GF(2). Note that the q-ary embedding could be implemented using multi-layered binary embedding [8], [12], [13] and if each binary embedding is optimal, the multi-layered embedding is optimal.

1) Adaptive Distortion Model: An additive distortion function D is used to measure the modification distortion brought by the embedding procedure. It has the form

$$D(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \rho_i(x_i, y_i), \tag{1}$$

where ρ_i are functions representing the cost of replacing the cover element x_i with y_i. Note that the situation under the constant distortion profile that we consider in this paper employs the common assumption: $\rho_i(y_i = x_i) = 0$. Therefore, we will use $\boldsymbol{\rho} = \{\rho_1(y_1 = \overline{x_1}), ..., \rho_N(y_N = \overline{x_N})\}$ to represent the distortion profile in the rest of the paper. In addition, because changing each element in the cover vector under the constant distortion profile brings the same distortion, by simply assigning $\rho_i(y_i = \overline{x_i}) = 1$, $D(\boldsymbol{x}, \boldsymbol{y})$ equals the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

We use $\pi(\boldsymbol{y}|\boldsymbol{x})$ to denote the probability distribution of the modifications, $\pi(\boldsymbol{y}|\boldsymbol{x}) \triangleq P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x})$. For any given cover $\boldsymbol{x}$, the sender could send at most

$$H(\pi) = -\sum_{\boldsymbol{y}} \pi(\boldsymbol{y}|\boldsymbol{x}) log_2 \pi(\boldsymbol{y}|\boldsymbol{x}) \tag{2}$$

bits information on average. Then, considering all possible cover vectors, by averaging over the whole space of $\boldsymbol{X}$, the average information is

$$E_P(H) = \sum_{\boldsymbol{x}} P(\boldsymbol{x}) H(\pi). \tag{3}$$

By generating the stego according to π, the average distortion

$$E_\pi(D) = \sum_{\boldsymbol{y}} \pi(\boldsymbol{y}|\boldsymbol{x}) D(\boldsymbol{x}, \boldsymbol{y}) \tag{4}$$

would be introduced for a given $\boldsymbol{x}$ [3], [8]. Then, the average distortion over the whole space of $\boldsymbol{X}$ is

$$E_P(D) = \sum_{\boldsymbol{x}} P(\boldsymbol{x}) E_\pi(D). \tag{5}$$

2) Performance Bounds: There are two forms of problems in the context of minimizing distortion under the adaptive distortion model for a given cover $\boldsymbol{x}$:

1) **Payload-limited sender (PLS):** embed m bits *fixed average payload* while minimizing the average distortion,

$$\underset{\pi}{minimize}\, E_\pi(D) \qquad subject\ to\ H(\pi) = m. \tag{6}$$

2) **Distortion-limited sender (DLS):** maximize the average payload while introducing a *fixed average distortion* D_ϵ,

$$\underset{\pi}{maximize}\, H(\pi) \qquad subject\ to\ E_\pi(D) = D_\epsilon. \tag{7}$$

The PLS problem is more commonly used in steganography compared to the DLS problem, besides, the two problems are dual to each other meaning that the optimal distribution for the PLS problem is also optimal for the DLS problem when considering the corresponding $E_\pi(D)$ and D_ϵ. The optimal solution has a form of a Gibbs distribution [3]:

$$\pi(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{N} \pi_i(y_i|x_i) \tag{8}$$

where

$$\pi_i(y_i|x_i) = \frac{exp(-\lambda \rho_i(x_i, y_i))}{\sum_{t_i} exp(-\lambda \rho_i(x_i, t_i))}. \tag{9}$$

The parameter $\lambda \in [0, \infty)$ in (9) is obtained by solving the algebraic equations corresponding to the constraints in (17), (18), while in practice, a simple binary search is usually used because of the monotonicity of $H(\pi)$ and $E_\pi(D)$ w.r.t. λ.

B. Steganographic Polar Coding

We will introduce the basic notations of polar codes as well as the method of SPC in this subsection. Most contents are the same as [10] and [9], so refer to them for more details.

1) Polar Codes: For a generic binary-input discrete memoryless channel (B-DMC) $W : \mathcal{X} \to \mathcal{Y}$ with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y}$, and transition probabilities $W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}$, the Bhattacharyya parameter of W is

$$Z(W) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{W(y|0) W(y|1)} \tag{10}$$

and the symmetric capacity of W is

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) log_2 \frac{W(y|x)}{\frac{1}{2} W(y|0) + \frac{1}{2} W(y|1)}. \tag{11}$$

If W is symmetric, $I(W)$ equals the channel capacity. It is proved in [10] that $I(W) = 0$ iff $Z(W) = 1$, and $I(W) = 1$ iff $Z(W) = 0$, which means that using Bhattacharyya parameter to estimate the quality of a channel is reasonable.

We use W^N to denote N uses of W, meaning: $W^N : \mathcal{X}^N \to \mathcal{Y}^N$ with $W^N(y_1^N|x_1^N) = \prod_1^n W(y_i|x_i)$. We use a_1^N to denote the row vector $(a_1, ..., a_N)$ and $a_i^j, 1 \le i \le j \le N$ to denote the subvector $(a_i, ...a_j)$. Given a_1^N and $\mathcal{A} \subset \{1, ..., N\}$, we use $a_\mathcal{A}$ to denote the subvector $\{a_i : i \in \mathcal{A}\}$. For message $u_1^N, N = 2^n, n \ge 0$, the polar encoder encodes u_1^N into x_1^N, and then, x_1^N passes the channel W^N and gets the output y_1^N. Then we have the synthesized channel $W_N : \mathcal{X}^N \to \mathcal{Y}^N$ with transition probabilities

$$W_N(y_1^N|u_1^N) = W^N(y_1^N|x_1^N). \tag{12}$$

We use $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}, 1 \le i \le N$ with transition probabilities

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \triangleq \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(y_1^N|u_1^N) \tag{13}$$

to denote the coordinate channels after *channel splitting* [10]. Channel polarization effect is that as N goes to infinity through powers of two, the fraction of indices $i \in \{1, ..., N\}$ for which $\lim_{N\to\infty} I(W_N^{(i)}) = 1$, $\lim_{N\to\infty} Z(W_N^{(i)}) = 0$ goes to $I(W)$ and for which $\lim_{N\to\infty} I(W_N^{(i)}) = 0$, $\lim_{N\to\infty} Z(W_N^{(i)}) = 1$ goes to $1 - I(W)$ [10, Theorem 1].

Polar codes can be identified by a parameter vector $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$, where K is the code dimension, $\mathcal{A}$ is the set of indices of *information* bits and $u_{\mathcal{A}^c}$ is called the *frozen* bits. One way to decide set $\mathcal{A}$ is to sort the Bhattacharyya parameters of the coordinate channels in increasing order and choose the first K bits. Note that the complexity to obtain the exact Bhattacharyya parameters is $O(N)$ for binary erasure channels (BECs) but exponential for other symmetric B-DMCs. As a consequence, there have been many methods devised to assess the quality of the coordinate channels with

acceptable complexity such as Arikan's heuristic method [14] and density evolution [15], [16].

The successive cancellation (SC) decoder proposed by Arikan [10] uses the transition probabilities of $W_N^{(i)}$ to calculate the likelihood ratio (LR) and then makes the decoding decision successively for bit i from 1 to N. The SC decoder decides bit i according to the following principle

$$\hat{u}_i \triangleq \begin{cases} u_i, & \text{if } i \in \mathcal{A}^c \\ h_i(y_1^N, \hat{u}_1^{i-1}), & \text{if } i \in \mathcal{A} \end{cases} \tag{14}$$

where functions $h_i : \mathcal{Y}^N \times \mathcal{X}^{i-1} \to \mathcal{X}$ are defined as

$$h_i(y_1^N, \hat{u}_1^{i-1}) \triangleq \begin{cases} 0, & \text{if } \frac{W_N^{(i)}(y_1^N, u_1^{i-1}|0)}{W_N^{(i)}(y_1^N, u_1^{i-1}|1)} \geq 1 \\ 1, & \text{otherwise.} \end{cases} \tag{15}$$

2) SPC: In steganography, the SC operation is employed during the encoding process. Therefore, in this paper, we use the expression *SC encoding* henceforth.

1) *Encoding:* For given cover vector $\boldsymbol{x}$ with distortion profile $\boldsymbol{\rho}$ and length N, as well as secret message $\boldsymbol{m}$ with length m, SPC encoder [9] first calculates the optimal modification probabilities $\pi_i(y_i \neq x_i)$ using (9) and sets the embedding rate $\alpha = \frac{m}{N}$. Then the encoder employs Arikan's heuristic method [14] to estimate the Bhattacharyya parameters by setting the erasure probability $p = \alpha$. Then the encoder sorts the indices i in decreasing order of the parameters, chooses the first $N\alpha$ bits to be the frozen indices set $\mathcal{A}^c$, and sets $u_{\mathcal{A}^c} = \boldsymbol{m}$.

   Then, the encoder sees the modification probability as the transition probability of a binary symmetric channel (BSC) and initializes $W_i = \pi_i$. With this initialization, the encoder uses SC encoding expressed by (14) to decode $\boldsymbol{x}$ and determine the information bits $u_{\mathcal{A}}$.

   By combining $u_{\mathcal{A}}$ and $u_{\mathcal{A}^c}$ into u_1^N and encoding u_1^N into $\boldsymbol{y}$ which is close to $\boldsymbol{x}$ using polar encoding [10], the embedding process is finished.
2) *Decoding:* Given the message length m and stego vector $\boldsymbol{y}$, the decoder uses Arikan's heuristic method which is the same as the encoder side to obtain the frozen indices set $\mathcal{A}^c$. Then, by conducting the inverse transformation from $\boldsymbol{y}$ to u_1^N, the decoder extracts the secret message $\boldsymbol{m} = u_{\mathcal{A}^c}$, which completes the extraction process.
3) *Distortion:* The average distortion over the whole space of $\boldsymbol{X}$ brought by the embedding process is given by

$$E_P(D) = E(D(\boldsymbol{X}, \boldsymbol{Y})) = \sum_{\boldsymbol{x} \in \{0,1\}^N} P(\boldsymbol{x}) D(\boldsymbol{x}, \boldsymbol{y}) \tag{16}$$

   where the expectation $E(\cdot)$ is over the randomness of cover random variable $\boldsymbol{X}$.

III. Main Results

We aim to prove the optimality of steganographic polar codes globally in the whole space of $\boldsymbol{X}$, that is to say, the codes can be applied to any cover vector $\boldsymbol{x}$ perfectly. Besides, the binary cover vectors are usually obtained using the least significant bits (LSBs) of images or other media which can be considered random noise. Therefore, it's reasonable to assume that $P(\boldsymbol{x}) = \frac{1}{2^N}$ for any $\boldsymbol{x} \in \{0,1\}^N$. With the assumption, the random variable $\boldsymbol{X}$ could be decomposed into $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_N)$ where $\boldsymbol{X}_i, 1 \leq i \leq N$ is a Bernoulli random variable with $P(1) = 0.5$.

Now we can restate the PLS problem and DLS problem considering the whole space of $\boldsymbol{X}$:

1) PLS:

$$\underset{\pi}{minimize}\, E_P(D) \qquad subject\ to\ E_P(H) = m. \tag{17}$$

2) DLS:

$$\underset{\pi}{maximize}\, E_P(H) \qquad subject\ to\ E_P(D) = D_\epsilon. \tag{18}$$

Here we give a lemma showing that the new form of the two problems is equivalent to the original problems in the situation of additive binary embedding.

Lemma 1: Let cover random variable $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_N)$ where $\boldsymbol{X}_i, 1 \leq i \leq N$ is a Bernoulli random variable with $P(1) = 0.5$. Then $E_P(H) = H(\pi), E_P(D) = E_\pi(D)$ in additive binary embedding.

Proof: For the binary embedding, (9) can be simplified into

$$\pi_i(y_i) = \frac{exp(-\lambda\rho_i(y_i))}{exp(-\lambda\rho_i(y_i = x_i)) + exp(-\lambda\rho_i(y_i = \overline{x}_i))}. \tag{19}$$

We first calculate $H(\pi_i)$ using (2) w.r.t. the optimal probability distribution in binary embedding (19):

$$H(\pi_i) = \begin{cases} H(\pi_i(1|0), \pi_i(0|0)), & \text{if } x_i = 0 \\ H(\pi_i(1|1), \pi_i(0|1)), & \text{if } x_i = 1. \end{cases} \tag{20}$$

Because

$$\pi_i(1|0) = \pi_i(0|1) = \frac{exp(-\lambda\rho_i(y_i = \overline{x}_i))}{exp(-\lambda\rho_i(y_i = x_i)) + exp(-\lambda\rho_i(y_i = \overline{x}_i))}$$

and

$$\pi_i(0|0) = \pi_i(1|1) = \frac{exp(-\lambda\rho_i(y_i = x_i))}{exp(-\lambda\rho_i(y_i = x_i)) + exp(-\lambda\rho_i(y_i = \overline{x}_i))},$$

we have that $H(\pi_i)$ holds the same value no matter $x_i = 0$ or $x_i = 1$. We use $\pi(\boldsymbol{Y}|\boldsymbol{X})$ to denote the probability distribution of the modifications. With the help of the independence property of additive steganography and (8), we have

$$\pi(\boldsymbol{Y}|\boldsymbol{X}) = \prod_{i=1}^{N} \pi_i(\boldsymbol{Y}_i|\boldsymbol{X}_i). \tag{21}$$

Then, we have

$$H(\pi) = \sum_{i=1}^{N} H(\pi_i). \tag{22}$$

As a result, $H(\pi)$ does not depend on $\boldsymbol{x}$.

From (3), we have

$$\begin{aligned} E_P(H) &= \sum_{\boldsymbol{x}} P(\boldsymbol{x}) H(\pi) \\ &= \frac{1}{2^N} H(\pi) \times 2^N \\ &= H(\pi). \end{aligned} \tag{23}$$

Similarly, applying the independence property of additive steganography, $E_\pi(D)$ in (4) can be decomposed as

$$E_\pi(D) = \begin{cases} \sum\limits_{i=1}^{N} \pi_i(1|0)\rho_i(y_i = \overline{x}_i) + \pi_i(0|0)\rho_i(y_i = x_i), \\ \qquad\qquad \text{if } x_i = 0 \\ \sum\limits_{i=1}^{N} \pi_i(0|1)\rho_i(y_i = \overline{x}_i) + \pi_i(1|1)\rho_i(y_i = x_i), \\ \qquad\qquad \text{if } x_i = 1. \end{cases} \tag{24}$$

Therefore, as $\pi_i(0|0) = \pi_i(1|1), \pi_i(1|0) = \pi_i(0|1)$, $E_\pi(D)$ is also independent of $\boldsymbol{x}$. As a result, from (5), we have

$$\begin{aligned} E_P(D) &= \sum_{\boldsymbol{x}} P(\boldsymbol{x}) E_\pi(D) \\ &= \frac{1}{2^N} E_\pi(D) \times 2^N \\ &= E_\pi(D) \end{aligned} \tag{25}$$

which concludes the proof. □

Now we see that the optimal modification probability distribution π does not really depend on $\boldsymbol{x}$. It only depends on the modification distortion profile $\boldsymbol{\rho}$. For convenience, we still use the commonly used pair $H(\pi)$ and $E_\pi(D)$ to represent the payload and average distortion in the following content.

The following lemma states the theoretical bound of the DLS problem under the constant distortion profile.

Lemma 2: Given the fixed average distortion per bit $\overline{D} = \frac{D_\epsilon}{N}$ and the constant distortion profile $\boldsymbol{\rho} = (1, ..., 1)$ with length N, the theoretical bound of the maximum embedding rate $\alpha = \frac{m}{N}$ in binary embedding is $H(\overline{D}) = H(\frac{D_\epsilon}{N})$, i.e., the maximum embedding payload m tends to $NH(\overline{D}) = NH(\frac{D_\epsilon}{N})$ as the cover length N tends to infinity.

Proof: By substituting $\rho_i(y_i \neq x_i) = 1, \rho_i(y_i = x_i) = 0$ into (19), we have

$$\begin{aligned} \pi_i(y_i \neq x_i) &= \frac{exp(-\lambda)}{1 + exp(-\lambda)}, \\ \pi_i(y_i = x_i) &= \frac{1}{1 + exp(-\lambda)}. \end{aligned} \tag{26}$$

Using (4) and the additive property (24), we have

$$E_\pi(D) = D_\epsilon = \sum_{i=1}^{N} \frac{exp(-\lambda)}{1 + exp(-\lambda)} = N\pi_i(y_i \neq x_i). \tag{27}$$

By substituting $\pi_i(y_i \neq x_i) = \frac{D_\epsilon}{N} = \overline{D}$ into (22), we have

$$\begin{aligned} H(\pi) = m &= NH(\frac{D_\epsilon}{N}) = NH(\overline{D}), \\ \alpha &= \frac{m}{N} = H(\frac{D_\epsilon}{N}) = H(\overline{D}) \end{aligned} \tag{28}$$

which concludes the proof. □

In practice, the cover length N is fixed. Then by fixing $\overline{D} = \frac{D_\epsilon}{N}$, the average distortion D_ϵ is also fixed, which corresponds to the conventional form of the DLS problem in (18). Therefore, the lemma fits the common problem practically. It is the same for the PLS problem considering the pair embedding rate α and distortion per bit $\overline{D}$ instead of embedding payload m and average distortion D_ϵ As a consequence, we will use the pair α and $\overline{D}$ in the following content for convenience.

The original SC encoder applies the hard decision function (15) which yields good performance in practice but is hard for theoretical analysis in our context. To make it much easier for the analysis, we slightly modify the method of SC encoding here [17] and then give the main theorem of the paper. We add randomness into (15) by calculating the likelihood ratio

$$l_i(y_1^N, \hat{u}_1^{i-1}) \triangleq \frac{W_N^{(i)}(y_1^N, u_1^{i-1}|0)}{W_N^{(i)}(y_1^N, u_1^{i-1}|1)} \tag{29}$$

and modifying the decision function h_i to be

$$h_i^*(y_1^N, \hat{u}_1^{i-1}) \triangleq \begin{cases} 0, & \text{with probability } \frac{l_i}{1+l_i} \\ 1, & \text{with probability } \frac{1}{1+l_i}. \end{cases} \tag{30}$$

We will use this slightly modified SC encoder [17] in the polar codes-based steganographic coding model from now on.

Under the modified SC encoding, the average distortion can be expressed using

$$E_P(D) = E(D^*(\boldsymbol{X}, \boldsymbol{Y})) = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} P(\boldsymbol{x}) P(\boldsymbol{y}|\boldsymbol{x}) D(\boldsymbol{x}, \boldsymbol{y}) \tag{31}$$

where the expectation is over the randomness of the cover random variable $\boldsymbol{X}$ and the random decisions incurred by (30).

Here we give the main theorem of the paper.

Theorem 1: Let cover random variable $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_N)$ where $\boldsymbol{X}_i, 1 \leq i \leq N$ is a Bernoulli random variable with $P(1) = 0.5$. Given the constant distortion profile $\boldsymbol{\rho} = (1, ..., 1)$ with length N and fixed *design* distortion per bit $0 < \overline{D} = \frac{D_\epsilon}{N} < \frac{1}{2}$, there exists a sequence of steganographic polar codes with length N tending to infinity through powers of two such that under the modified SC encoding, the maximum embedding rate $\alpha = \frac{m}{N}$ tends to $H(\overline{D}) = H(\frac{D_\epsilon}{N})$ while the average distortion per bit tends to $\overline{D} = \frac{D_\epsilon}{N}$.

Proof: Let $\boldsymbol{x} = (x_1, ..., x_N)$ denote a realization of $\boldsymbol{X}$. Let W denote a BSC with transition probability $\overline{D} = \frac{D_\epsilon}{N}$. We use the above settings to assess the quality of the coordinate channels $W_N^{(i)}$, for example, calculate the exact Bhattacharyya parameters $Z(W_N^{(i)})$. Then, we sort the indices i according to their corresponding channel quality, e.g., sort them in decreasing order of Bhattacharyya parameters $Z(W_N^{(i)})$. We choose the first $NH(\frac{D_\epsilon}{N})$ indices to form the frozen set $\mathcal{A}^c$ while the other indices are in $\mathcal{A}$, which completes the step of code construction. With the help of the channel polarization effect [10, Theorem 1], we know that $\lim\limits_{N\to\infty} Z(W_N^{(i)}) = 1, i \in \mathcal{A}^c$ and $\lim\limits_{N\to\infty} Z(W_N^{(i)}) = 0, i \in \mathcal{A}$.

Let $u_{\mathcal{A}^c} = \boldsymbol{m}$. Then use the modified SC encoding to find $u_{\mathcal{A}}$ and obtain the stego vector $\boldsymbol{y}$ after polar encoding.

For this polar codes-based steganographic coding model, we can see that $m = |\mathcal{A}^c| = NH(\frac{D_\epsilon}{N})$ as N goes to infinity through power of two. Thus the embedding rate $\alpha = \frac{m}{N}$ tends to $\frac{NH(\frac{D_\epsilon}{N})}{N} = H(\frac{D_\epsilon}{N}) = H(\overline{D})$. Here we need to verify that the average distortion per bit tends to $\overline{D}$.

Let us analyze the transition from $\boldsymbol{x}$ to $\boldsymbol{y}$. For any $1 \leq i \leq N$, the probability that $y_i \neq x_i$ is $\pi_i(y_i \neq x_i)$ while the probability that $y_i = x_i$ is $\pi_i(y_i = x_i) = 1 - \pi_i(y_i \neq x_i)$. This can be modeled as a BSC($\pi_i(y_i \neq x_i)$). If and only if $\pi_i(y_i \neq x_i) = \frac{D_\epsilon}{N} = \overline{D}$, this BSC is the channel W and $E_\pi(D^*) = D_\epsilon$. As a result, the last thing needed is to verify that $E(D^*(\boldsymbol{X}, \boldsymbol{Y}))$ tends to D_ϵ. In that case, $\pi_i(y_i \neq x_i) = \frac{D_\epsilon}{N}$ and $E_\pi(D^*) = E(D^*(\boldsymbol{X}, \boldsymbol{Y})) = D_\epsilon$ as N goes to infinity through powers of two. Here we can use the theorem in [17] which says that the expected distortion $E(D^*(\boldsymbol{X}, \boldsymbol{Y}))$ does tend to D_ϵ and is independent of the vector $u_{\mathcal{A}^c} = \boldsymbol{m}$, which concludes the proof. □

Applying the results of Lemma 2 and Theorem 1, we have the following corollary.

Corollary 1: The polar codes-based steganographic coding method is optimal for the DLS problem under the constant distortion profile given the exact Bhattacharyya parameters of the coordinate channels after channel splitting.

Polar codes-based steganographic coding is also optimal for the PLS problem which is dual to the DLS problem.

Theorem 2: Let cover random variable $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_N)$ where $\boldsymbol{X}_i, 1 \leq i \leq N$ is a Bernoulli random variable with $P(1) = 0.5$. Given the constant distortion profile $\boldsymbol{\rho} = (1, ..., 1)$ with length N and fixed embedding rate $\alpha = \frac{m}{N}$, there exists a sequence of steganographic polar codes with length N tending to infinity through powers of two with embedding rate α such that under modified SC encoding, the minimum average distortion per bit $\overline{D} = \frac{D_\epsilon}{N}$ tends to $H^{-1}(\alpha)$.

Proof: We use the same model in the proof of Theorem 1. For the same code sequence, using the duality between the PLS problem and DLS problem, we know that

$$\lim_{N\to\infty} H(\frac{D_\epsilon}{N}) = \lim_{N\to\infty} H(\overline{D}) = \alpha. \tag{32}$$

Therefore, by solving the above equation (32), we conclude that the average distortion per bit $\overline{D} = \frac{D_\epsilon}{N}$ tends to $H^{-1}(\alpha)$ which completes the proof. □

Similar to the former result, we conclude the PLS problem with the following corollary.

Corollary 2: The polar codes-based steganographic coding method is also optimal for the PLS problem under the constant distortion profile given the exact Bhattacharyya parameters of the coordinate channels after channel splitting.

IV. Conclusion

In this paper, we proved that using polar codes in steganographic coding contributes to optimal methods for additive steganography under the constant distortion profile. These are the first provably optimal steganographic codes in the history of steganography. Although it is hard to estimate the quality of the coordinate channels after channel splitting in practice, using the suboptimal methods such as Arikan's heuristic method employed in [9] could already yield good performance, which makes polar codes-based steganographic coding practicable in real-world applications.

References

[1] I. Cox, M. Miller, J. Bloom, J.Fridrich and T. Kalker *Digital Watermarking and Steganography*. San Mateo, CA, USA: Margan Kaufmann, 2008.

[2] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[3] T. Filler and J. Fridrich, "Gibbs construction in steganography," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 705–720, Dec. 2010.

[4] A. Westfeld, "High capacity despite better steganalysis (F5-A steganographic algorithm)," in *Proc. Int. Workshop Inf. Hiding*. New York, NY, USA: Springer-Verlag, 2001, pp. 289–302.

[5] M. Van Dijk and F. Willems, "Embedding information in grayscale images," in *Proc. 22nd Symp. Inf. Commun. Theory*, 2001, pp. 147–154.

[6] D. Schönfeld and A. Winkler, "Embedding with syndrome coding based on BCH codes," in *Proc. 8th Workshop Multimedia Secur.*, 2006, pp. 214–223.

[7] R. Zhang, V. Sachnev, and H. J. Kim, "Fast BCH syndrome coding for steganography," in *Proc. Int. Workshop Inf. Hiding*, vol. 2009, pp. 48–58.

[8] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.

[9] W. Li, W. Zhang, L. Li, H. Zhou and N. Yu, "Designing Near-Optimal Steganographic Codes in Practice Based on Polar Codes," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 3948-3962, July 2020.

[10] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051-3073, Jul. 2009.

[11] A. Balatsoukas-Stimming, M. B. Parizi, and A. Burg, "LLR-based successive cancellation list decoding of polar codes," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5165–5179, Oct. 2015.

[12] X. Zhang, W. Zhang, and S. Wang, "Efficient double-layered steganographic embedding," *Electronics Letters*, vol. 43, pp. 482–483, April 2007.

[13] T. Filler and J. Fridrich, "Using non-binary embedding operation to minimize additive distortion functions in steganography," in *Second IEEE International Workshop on Information Forensics and Security,* (Seattle, WA), 2010.

[14] E. Arikan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447-449, Jun. 2008.

[15] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE Int. Symp. on Inf. Theory,* Jun. 2009, pp. 1496-1500.

[16] R. Mori and T. Tanaka, "Performance of Polar Codes with the Construction using Density Evolution," *IEEE Commun. Lett.,* vol. 13, no. 7, pp. 519-521, July 2009.

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