Weighted Polar Codes for Channels with State

Yanxiao Liu
Department of Information Engineering
The Chinese University of Hong Kong
Hong Kong, China
yanxiaoliu@link.cuhk.edu.hk

Chih Wei Ling
Department of Information Engineering
The Chinese University of Hong Kong
Hong Kong, China
chihweiLing@link.cuhk.edu.hk

Cheuk Ting Li
Department of Information Engineering
The Chinese University of Hong Kong
Hong Kong, China
ctli@ie.cuhk.edu.hk

***Abstract*—In this paper, we introduce a new class of polar codes, called weighted polar codes. Recently, it has been observed that by attaching weights to parity-check bits of error-correcting channel codes (instead of fixing parity-check bits to zeros), the error rate can be reduced for channels with state information and/or cost constraints. We extend the weighted construction to polar codes, a class of codes that have efficient encoding and decoding algorithms and are provably capacity-achieving. Experiment results show that for channels with state information (Gelfand-Pinsker problem), the weighted polar codes have smaller error rates compared to the nested polar codes in short blocklength regime.**

***Index Terms*—Information theory, finite blocklength, Gelfand-Pinsker problem, channels with state, polar codes.**

I. Introduction

Error-correcting codes are used in channel coding to correct errors during information transmission. Conventionally, the codebook of an error-correcting code, which consists of sequences that may be chosen as the channel inputs, is a fixed set, and whether a sequence is a codeword is a binary choice. However, it has been shown in [1]–[3] that by attaching weights (probabilities) to the codewords which correspond to the likelihood of the codewords to be chosen, we can have advantages in various settings, e.g., smaller error rates in channels with state information available to the encoder noncausally, i.e., the Gelfand-Pinsker problem [4], [5]. By this construction, the codebook is a "fuzzy set" instead, called a *weighted codebook* in [2].

However, the weighted code in [1], [2] is unstructured, and therefore no efficient coding algorithms could be provided. In [3], structured weighted codes based on linear codes are proposed. The performance is consistently better than the traditional nested linear codes for channels with state [6], and the codes were proved to be capacity-achieving for any (symmetric or asymmetric) channel with state. While the construction in [3] can be applied on sparse parity-check matrices, the design of efficient coding algorithms does not appear to be straightforward.

In this paper, we extend the weighted construction to the polar codes. Compared with the nested polar codes [7], [8] for the channels with state, which divide the bits into three groups, namely *information bits* (that are set to be the bits in the message), *fixed bits* (that are fixed to zero) and *flexible bits* (that can adapt to the channel state), the proposed codes, called *weighted polar codes*, blur the boundary between fixed bits and flexible bits, by treating them as a single group (the *weighted bits*), and attaching a weight to each of these bits representing the "degree of flexibility" of the bit. The proposed codes utilize mutual information to choose information bits and assign weights to the weighted bits. Experiment results show that the proposed codes can attain lower block error rates compared to the nested polar codes in short blocklength regime (e.g. ultra-small packet lengths $N = 512$ to $N = 2048$), and therefore the code is practical for low-latency required communication settings.

A. Coding with Side or State Information

For the channels with state information (Gelfand-Pinsker) setting [4], [5], [9], there is a state sequence that indicates the channel statistics, and varies throughout the information transmission. The state sequence is available to the encoder non-causally, but is unknown to the decoder. This setting was originally studied on memory with stuck-at faults [9]. It was then generalized to the discrete memoryless channel with discrete memoryless state in [4], [5]. In [10] the capacity of the Gaussian channel with additive Gaussian state was studied, which is called *writing on dirty paper*. The finite blocklength cases were studied in [2], [11]–[13].

It was shown in [6], [14] that the channel coding with state information has a duality with the source coding with decoder side information, i.e., the Wyner-Ziv problem [15]. The construction of encoders for the Wyner-Ziv problem was studied in [16], and the rate loss was discussed in [17]. The Wyner-Ziv problem with multiple sources was discussed in [18].

It was shown in [19] that nested code constructions, i.e., a collection of good source codes are nested inside a good channel code (or vice versa) can achieve the optimal rates of the Gelfand-Pinsker problem and the Wyner-Ziv problem. However, the random constructions of nested codes are impractical. Nested linear code constructions were given in [6] for the information embedding problem [20]. The sparse graphical codes can be used to generate practical nested linear codes in [21].

Weighted codebook constructions [1], [2] were proved to achieve the optimal rates for the Gelfand-Pinsker problem and

This work was partially supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No.s: CUHK 24205621 (ECS), CUHK 14209823 (GRF)].

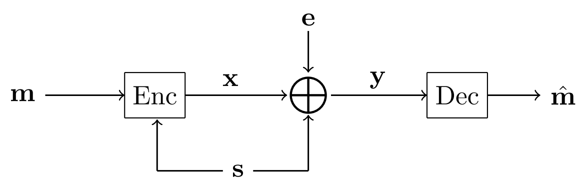

Fig. 1. Block diagram representation of a channel with state information.

the Wyner-Ziv problem, and attain the best known second-order asymptotic results. In [3], it was shown that a linear weighted code construction can achieve the optimal rates for the Gelfand-Pinsker problem and the Wyner-Ziv problem.

B. Polar Codes

Polar codes [22] are the first practical codes that can be proved to achieve the capacity of a wide range of channels. The construction is based on *channel polarization*, which refers to the fact that when the blocklength increases, the fraction of "reliable" channels is approaching the channel capacity. The speed of polarization was discussed in [23]. Originally, the polar codes were constructed by calculating the Bhattacharyya parameters and ranking sub-channels based on them. Various improvements on the code constructions were proposed, for example, the polar weight construction [24] and the Gaussian approximation construction [25]. A summary of polar codes construction methods can be found in [26]. An important improvement on the efficiency of polar codes is the list decoding of polar codes, which keeps a list of possible codewords during the successive cancellation decoding [27].

In [7], it was shown that polar codes are also optimal for source coding problem, and therefore are optimal for settings combining both channel and source coding with or without side information, e.g., the Gelfand-Pinsker problem and the Wyner-Ziv problem. They used a nested construction of polar codes to achieve the capacity. In [8], list decoding with cyclic redundancy check aided successive cancellation coding was used for the Gelfand-Pinsker problem.

In [28], the information bits play roles like frozen bits: the so-called *half-frozen bits* are placed at where errors hardly occur within the frozen bits and also where errors are easy to occur within the information bits. Soft-output decoding of polar codes was proposed in [29]. Moreover, polar coding for secure Wyner-Ziv problem was studied in [30].

The remainder of the paper is organized as follows. After introducing the problem setting in Sec. II, we review the polar codes and nested polar codes in Sec. III. We then show the weighted polar codes in Sec. IV, and the experiments will be discussed in Sec. V.

II. Problem Formulation

We review the Gelfand-Pinsker problem [4] as follows. Suppose there is a state sequence available noncausally at the encoder, denoted as $\mathbf{S} = [S_1, \ldots, S_n]$, $S_i \in \mathcal{S}$, such that $S_1, \ldots, S_n \overset{iid}{\sim} P_S$. The encoder observes $\mathbf{S}$ and the message $\mathbf{m}$, and encodes them to the codeword $\mathbf{X} \in \mathbb{F}_2^n$. The input is subject to a cost constraint, namely $\mathbf{E}[\sum_{i=1}^n c(S_i, X_i)] \leq nD$ where $c : \mathcal{S} \times \mathbb{F}_2 \to [0, \infty)$ is the cost function. The codeword is then sent through a channel $P_{Y|S,X}(y|s,x)$ which is memoryless. The decoder observes the output of the channel, denoted as $\mathbf{Y} = [Y_1, \ldots, Y_n]$, and decodes it to $\hat{\mathbf{m}}$. The problem is illustrated in Figure 1.

III. Polar Codes and Nested Polar Codes

Polar codes were invented in [22] for channel coding and extended to source coding [7]. They are also proved to be optimal for the Gelfand-Pinsker problem and the Wyner-Ziv problem by using a nested construction [7]. We briefly review the polar codes and the nested polar codes.

A. Polar Codes for Channel and Source Coding

Suppose $W(y|x)$ is a discrete memoryless channel, where $x \in \{0, 1\}$ and $y \in \mathcal{Y}$. The Bhattacharyya parameter of W is defined as $Z(W) = \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$. We use an upper case bold letter $\mathbf{U}$ and a lower case bold letter $\mathbf{u}$ to denote a random vector and its realization, respectively. Unless otherwise specified, we assume a vector has length $N = 2^n$, where $n \in \mathbb{N}_+$. We use $\mathbf{u}_i^j$ to denote the random vector $\mathbf{u}_i^j = [u_i, u_{i+1}, \ldots, u_j]$ for $i < j$. Write $\mathbf{u}_F$ for the vector consisting of the entries of $\mathbf{u}$ with indices in F, without changing the order within them.

Given $G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, we apply the transform $G_2^{\otimes n}$ (the n^{th} Kronecker product of G_2) to the vector $\mathbf{u}$, and send the entries of $\mathbf{u}G_2^{\otimes n}$ through independent copies of W. When N increases, the channels polarize [22] and become either noiseless or purely noisy. There are $I(W)$ fraction of the channels $\{W_N^{(i)}\}$ become noiseless with $Z(W_N^{(i)})$ close to 0, which are used to transmit messages with a rate equal to the channel capacity. The set of indices of these sub-channels is called the *information set*.

Suppose $\mathbf{U}$ is uniform over $\{0, 1\}^N$, $\mathbf{X} = \mathbf{U}G_2^{\otimes n}$ is the channel input and $\mathbf{Y}$ is the output. The channel from $\mathbf{U}$ to $\mathbf{Y}$ is

$$P_{\mathbf{Y}|\mathbf{U}}(\mathbf{y}|\mathbf{u}) := \prod_{i=0}^{N-1} W(y_i|(\mathbf{u}G_2^{\otimes n})_i).$$

Moreover, $W_N^{(i)} : \{0, 1\} \to \{0, 1\}^i$ is the channel with input u_i and output $(\mathbf{y}, \mathbf{u}_0^{i-1})$, where the transition probabilities are

$$W_N^{(i)}(\mathbf{y}, \mathbf{u}_0^{i-1}|u_i) := P(\mathbf{y}, \mathbf{u}_0^{i-1}|u_i).$$

The successively cancellation (SC) coding is shown in Algorithm 1. Given a frozen set F, the bits are decoded from bit 0 to bit $N - 1$, where the likelihood is calculated based on $W_N^{(i)}(\mathbf{y}, \mathbf{u}_0^{i-1}|u_i)$. If $i \in F$, we set $\hat{u}_i$ to a frozen value, which is usually 0; otherwise, we calculate

$$L_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}) = \frac{W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|u_i = 0)}{W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|u_i = 1)}, \tag{1}$$

and make a decision $\hat{u}_i = 0$ if $L_N^{(i)} > 1$ and $\hat{u}_i = 1$ if $L_N^{(i)} \leq 1$.

The SC encoder for source coding performs similar coding steps [7]. A polar code $C(F)$ is defined by $C(F) := \{\mathbf{x} = \mathbf{u}G_2^{\otimes n}\}$, where $F \subseteq \{0, \dots, N-1\}$.

Algorithm 1 High-Level Description of the SC Coding [22]

Input: the received vector $\mathbf{y}$
Output: $\hat{\mathbf{u}}$
procedure DECODER
  for $i = 1, \dots, N-1$ **do**
    calculate $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|0), W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|1)$
    if u_i is frozen **then**
      set $\hat{u}_i$ to the frozen value
    else
      if $\frac{W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|0)}{W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|1)} > 1$ **then**
        set $\hat{u}_i = 0$
      else
        set $\hat{u}_i = 1$

B. Nested Polar Codes

The nested polar codes [7] were proposed to solve problems including both channel and source coding, e.g., the channels with state. Fix $0 < p < D < 1/2$, we construct two polar codes: $\mathcal{C}_c$ and $\mathcal{C}_s$ based on two sets F_c and F_s, respectively:

$$F_s = \{i : Z_N^{(i)}(D) \geq 1 - (\delta/N)^2\},$$
$$F_c = \{i : Z_N^{(i)}(p) \geq \delta/N\},$$

and $Z_N^{(i)}(D)$, $Z_N^{(i)}(p)$ are the Bhattacharyya parameters of the i-th sub-channel after the polarization of BSC(D) and BSC(p), respectively. The coding scheme is as follows.

The bits are divided into three groups: the information bits $F_s \backslash F_c$, the fixed bits $F_c \cap F_s$, and the flexible bits $F_s^{\complement} \cap F_c^{\complement}$. There are also bits not in these three groups, called the retransmission bits $F_c \backslash F_s$, though it has been proved that the number of such bits vanishes asymptotically [8], and can be retransmitted without affecting the channel rate [7]. Therefore we can assume $F_c \subseteq F_s$. The source polar code $\mathcal{C}_s(F_s)$ first sets the fixed bits to zero, i.e., $\mathbf{u}_{F_c \cap F_s} = \mathbf{0}$, and the information bits to $\mathbf{m}$, i.e., $\mathbf{u}_{F_s \backslash F_c} = \mathbf{m}$. Then the encoder observes the state sequence $\mathbf{s}$ and obtains $\mathbf{s}' \in \mathcal{C}_s(F_s)$ by SC encoding, while being able to choose the values of the flexible bits according to $\mathbf{s}$. The encoder transmits $\mathbf{x} = \mathbf{s} \oplus \mathbf{s}'$ through BSC(p), and the cost constraint is $c(\mathbf{x}) \leq D$. The decoder receives $\mathbf{y} = \mathbf{x} \oplus \mathbf{s} \oplus \mathbf{z} = \mathbf{s}' \oplus \mathbf{z}$ where $\mathbf{z}$ is the noise vector, and decode $\hat{\mathbf{u}}_{F_m^c}$ using the SC decoding in Algorithm 1.

IV. WEIGHTED POLAR CODES

In this section, we design a class of codes for channels with state, called *weighted polar codes*. For channels with state, the rate $I(X;Y) - I(X;S)$ is achievable due to the Gelfand-Pinsker theorem [4], [5]. This motivates us to design the weighted polar codes by measuring the mutual information. Suppose $(X_i, Y_i, S_i) \sim P_{X,Y,S}$ i.i.d. for $i = 1, \dots, N$. Consider

$$\begin{aligned} & n(I(X;Y) - I(X;S)) \\ & = I(\mathbf{X}_0^{n-1}; \mathbf{y}) - I(\mathbf{X}_0^{n-1}; \mathbf{S}_0^{n-1}) \\ & = \sum_{i=0}^{n-1} \big(I(X_i; Y_i | \mathbf{X}_0^{i-1}) - I(X_i; S_i | \mathbf{X}_0^{i-1}) \big). \end{aligned} \tag{2}$$

Therefore, to design a code that can achieve the rate $I(X;Y) - I(X;S)$ we define

$$\alpha_i := I(X_i; Y_i | \mathbf{X}_0^{i-1}),$$
$$\beta_i := I(X_i; S_i | \mathbf{X}_0^{i-1}),$$

which stand for the mutual information between channel input and output and the mutual information between channel input and the state, respectively. Because of channel polarization [22], both α_i and β_i polarize to the values 0 and 1, and hence the values of $\alpha_i - \beta_i$ will polarize to three numbers: -1, 0 or 1.

We will use such a polarization to design the weighted polar codes. Before that, we first illustrate the nested polar codes in the context of α_i and β_i. For the nested polar codes, the bits are divided to four groups:

- The **information bits** $F_s \backslash F_c$, which are the bits with $\alpha_i \approx 1$ and $\beta_i \approx 0$. According to (2), these bits correspond to the sub-channels which contribute the most to the capacity of the channel with state, and hence they should be used to transmit the message.
- The **fixed bits** $F_c \cap F_s$, which are the bits with $\alpha_i \approx 0$ and $\beta_i \approx 0$. These bits are fixed to zero in the nested polar codes.
- The **flexible bits** $F_s^{\complement} \cap F_c^{\complement}$, which are the bits with $\alpha_i \approx 1$ and $\beta_i \approx 1$. The encoder is free to choose these bits according to the state sequence.
- The **retransmission bits** $F_c \backslash F_s$, which are the bits with $\alpha_i \approx 0$ and $\beta_i \approx 1$. The number of such bits vanishes asymptotically [8].

In our coding scheme, instead of dividing the bits into four groups (or three groups ignoring the retransmission bits), we only divide the bits into two groups: those with $\alpha_i - \beta_i \approx 1$ which contribute to the capacity (2), and those with $\alpha_i - \beta_i \approx 0$ which do not contribute. We now describe these two groups in detail.

- The **information bits** are the bits with $\alpha_i - \beta_i \approx 1$, implying $\alpha_i \approx 1$ and $\beta_i \approx 0$. These bits correspond to the sub-channels which contribute to the capacity. To transmit a message of k bits, we choose the k bits with the largest $\alpha_i - \beta_i$ to be the message bits. Intuitively, they are similar to the information bits $F_s \backslash F_c$ in the nested polar codes, though selecting the information bits according to $\alpha_i - \beta_i$ would result in slightly different choices.
- The **weighted bits** are the bits with $\alpha_i - \beta_i \approx 0$. This covers both cases $(\alpha_i, \beta_i) \approx (0, 0)$ (fixed bits in nested polar codes) and $(\alpha_i, \beta_i) \approx (1, 1)$ (flexible bits in nested polar codes). According to (2), these bits correspond to the sub-channels which does not contribute

to the capacity. Since both fixed bits and flexible bits are "unused", it is unnecessary to have a hard partition rule between them. Instead, we attach a weight q_i to every weighted bit, which represents how flexible the bit is.

Note that $\alpha_i - \beta_i \approx -1$ is rare for the same reason that retransmission bits are rare [8], and no bit with $\alpha_i - \beta_i \approx -1$ has been observed throughout our experiments.

In practice, the quantities α_i and β_i are approximated by the lower bound in terms of the Bhattacharyya parameter in [22, Proposition 1]:

$$I(W) \geq \log \frac{2}{1 + Z(W)}.$$

An alternative is to use the upper bound estimate $I(W) \leq \sqrt{1 - Z(W)^2}$, though the lower bound estimate is observed in experiments to give better performance.

We then discuss the method of attaching weights. The weight q_i of the i-th sub-channel corresponds to the prior probability of the bit being 1. If $q_i = 0$, then the bit will be frozen to 1. If $q_i = 1/2$, there will be no prior preference to either 0 or 1, and the bit will behave as a flexible bit that the encoder can choose depending on the state sequence. During the weighted SC coding, we make a decision $\hat{u}_i = 0$ if

$$L_N^{(i)}(\mathbf{u}, \hat{\mathbf{u}}_0^{i-1}) > \frac{q_i}{1 - q_i}.$$

Compared to Algorithm 1, a weight $\frac{1-q_i}{q_i}$ is attached on the likelihood (1). The strategy is to assign larger q_i to the bits that depend more on the state sequence. Therefore, we let the weight sequence being an increasing function of β_i. If $\beta_i \approx 1$, we take $q_i \approx 0.5$, i.e., high flexibility. If β_i is small, we take a small q_i so that it is more biased. We want q_i to increase quickly to be close to 0.5 when β_i is small, and then increases slowly until $\beta_i = 1$. Therefore, we take

$$q_i = \frac{1 - (1 - \beta_i)^b}{2} \tag{3}$$

for $i = 1, \ldots, N - k$, where $b > 1$ is a parameter.

Given input $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{q}$, the encoder and decoder perform weighted successively cancellation coding by Algorithm 2. We can see that the complexity of encoding or decoding is still $O(N \log(N))$.

In summary, the advantages of weighted polar codes over the nested polar codes are as follows.

1) Compared with the nested polar codes, we treat fixed bits and flexible bits in a uniform manner, by attaching a degree of flexibility to each bit.
2) The encoder is flexible to trade-off between choosing a high-weight sequence, which has higher likelihood of being chosen by the decoder, and choosing a low-weight sequence, which may be closer to the state sequence. This explains the performance gain of our weighted code construction for channels with states, and possibly Wyner-Ziv problem and asymmetric channels [3].
3) Experiments show that in short blocklength regime, e.g., $512 \leq N \leq 2048$, the weighted polar codes achieve smaller costs and block error rates, which will be discussed in Section V.

We remark that as $N \to \infty$, α_i and β_i will polarize to 0 or 1, and hence q_i will polarize to 0 or $1/2$, and the weighted polar codes will be approximately the same as the nested polar codes. Intuitively, this means the weighted polar codes should also be capacity-achieving like the nested polar codes. A rigorous proof is left for future study. For N that is not large, there will be a larger portion of weighted bits with q_i not close to 0 or $1/2$, which can lead to an improvement in the

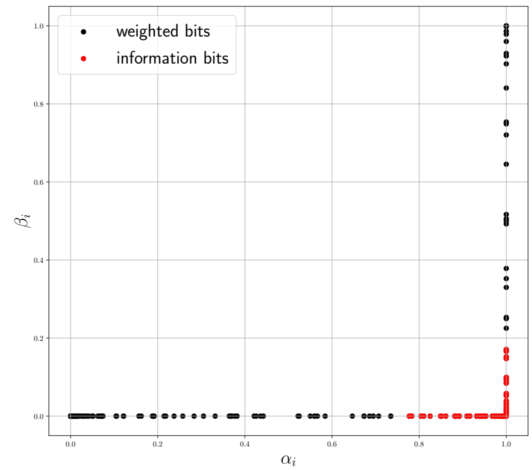

Fig. 2. The plot of α_i and β_i for the binary-Hamming information embedding setting in Section V, where $D = 0.3$, $p = 0.05$, with blocklength $N = 256$ and message length 100. The red points are the information bits, and others are the weighted bits.

Algorithm 2 High-Level Description of Weighted SC Coding

Input: $\mathbf{u}$, message $\mathbf{m}$, information set F_m. In encoding, $\mathbf{u} = \mathbf{s}$; in decoding, $\mathbf{u} = \mathbf{y}$. A weight sequence $\mathbf{q}$.
Output: a decoded codeword $\hat{\mathbf{u}}$.
procedure ENCODER AND DECODER
  for $i = 1, \ldots, N - 1$ **do**
    calculate $W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|0), W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|1)$
    if Encoding **then**
      if $i \in F_m$ **then**
        set $\hat{u}_i = m_i$
      else
        if $\frac{W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|0)}{W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|1)} > \frac{q_i}{1-q_i}$ **then**
          set $\hat{u}_i = 0$
        else
          set $\hat{u}_i = 1$
    if Decoding **then**
      for $i \in F_m$ **do**
        set $q_i = 1/2$
      if $\frac{W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|0)}{W_N^{(i)}(\mathbf{u}_0^{N-1}, \hat{\mathbf{u}}_0^{i-1}|1)} > \frac{q_i}{1-q_i}$ **then**
        set $\hat{u}_i = 0$
      else
        set $\hat{u}_i = 1$

block error rate as measured in the experiments in Section V.

V. Experiments

We consider the binary-Hamming information embedding setting [6]. Consider a state sequence $\mathbf{S}$ with entries $S_i \sim \text{Bern}(1/2)$ and the channel is a binary symmetric channel (BSC) with crossover probability p. The transmission is required to satisfy an expected cost constraint $\mathbf{E}[|\{i \in \{1,\ldots,n\} : X_i \neq S_i\}|] \neq nD$ for $0 < D < 1$, similar to [7], [8]. The D is called the maximum average cost (or distortion) per symbol. While satisfying the average cost constraint, we want to achieve the optimal tradeoff between the code rate $R = k/N$ and the error probability $\mathbf{P}(\mathbf{M} \neq \hat{\mathbf{M}})$. It has been shown in [6] that the capacity C of this problem is $C = \mathfrak{C}[g(D)]$ such that

$$g(D) = \begin{cases} 0, & \text{if } 0 \leq D < p, \\ H(D) - H(p), & \text{if } p \leq D \leq 1/2, \end{cases}$$

and $\mathfrak{C}[g(D)]$ is the upper concave envelope of $g(D)$.

We compare the weighted polar codes with attaching weights by (3) and choosing the parameter $b = 15$, to the nested polar code [7], [8] for the the cases with blocklength N from 512 to 2048, crossover probability $p = 0.05$.

Figure 3 shows the tradeoff between the cost and the block error rate measured in the experiments. In the plots, for each data point, we perform 2×10^4 trials to compute the block error rate and the average cost. We observe that the weighted polar codes (WPC) outperform the nested polar codes (NPC) almost consistently. When D is not too close to $1/2$, there are many weighted bits and we have performance improvement. However, when D is close to $1/2$, there might be only one or two weighted bits, and the performance is similar with the nested polar codes. We compare them on high information rate ranges (i.e., K/N is large), since when the rate is low both of them have small and close error probabilities.

We remark that having performance improvement in short blocklength (e.g. from $N = 512$ to $N = 2048$) cases show that the weighted polar codes can be applied to practical delay-constrained communication settings. While the performance improvement appears not huge, it would be unrealistic to expect a huge improvement by an order since the nested polar codes are already capacity-achieving.

VI. Conclusion

In this paper, we propose a class of polar codes, called the weighted polar codes, for the channels with state information. Each parity check bit is attached by a weight, which enables us to deal with the fixed bits and flexible bits in the nested polar codes in a uniform manner. Experiment results show that we have smaller block error rates when the blocklength is short, and therefore the code is practical for delay-constrained communication settings. Weighted polar code constructions for other settings, e.g., source coding with side information and asymmetric channels, are left for future research. Moreover, more general cases (compared with the binary symmetric channel in this paper) can also be studied in the future.

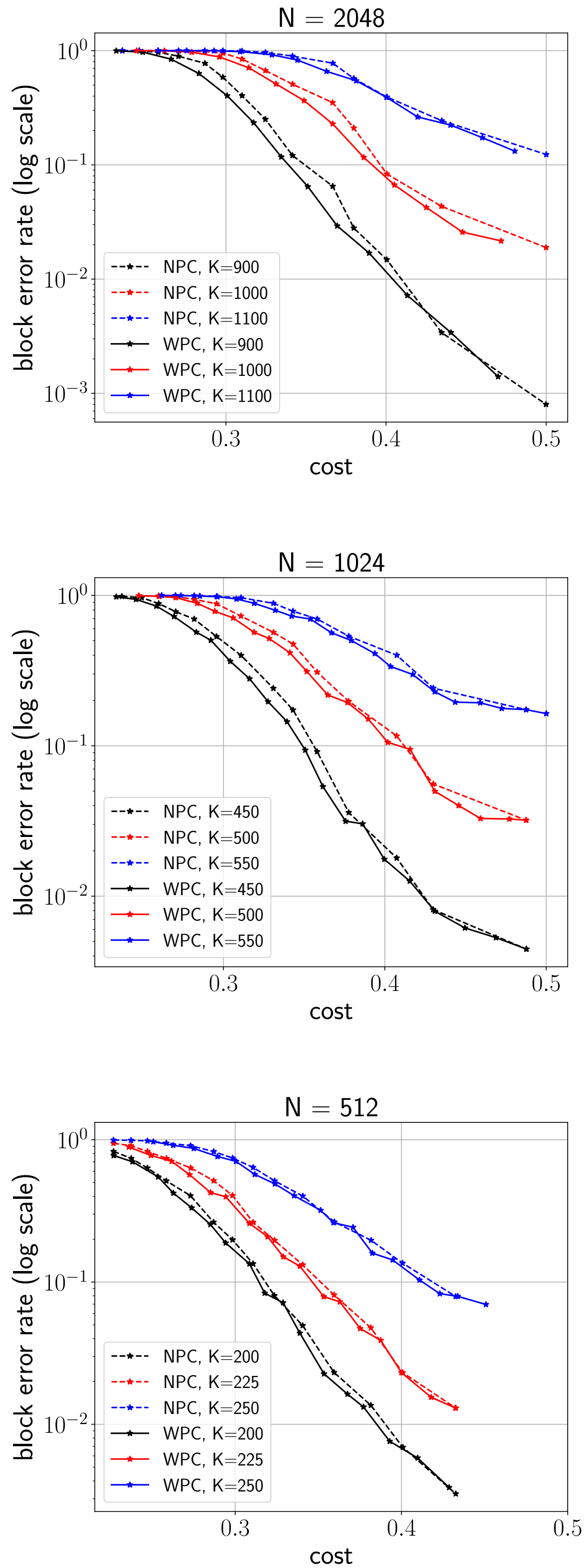

Fig. 3. Comparison of the block error rate (log scale) of weighted polar codes (WPC) and nested polar codes (NPC) on blocklength from $N = 512$ to $N = 2048$. The crossover probability is $p = 0.05$ and we perform 2×10^4 trials for each data point. Weight sequence chooses $b = 15$ in (3).

REFERENCES

[1] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6967–6978, Nov 2018.

[2] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the Poisson matching lemma," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.

[3] C. W. Ling, Y. Liu, and C. T. Li, "Weighted parity-check codes for channels with state and asymmetric channels," *arXiv preprint arXiv:2201.10171*, 2022.

[4] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[5] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 731–739, 1983.

[6] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1159–1180, 2003.

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

[8] B. Beilin and D. Burshtein, "On polar coding for side information channels," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 673–685, 2020.

[9] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Peredachi Inf.*, vol. 10, no. 2, pp. 52–60, 1974.

[10] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[11] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1–8.

[12] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *2013 IEEE ISIT*, July 2013, pp. 1287–1291.

[13] Y. Liu and C. T. Li, "One-shot coding over general noisy networks," in *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024.

[14] J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between distributed source coding and data hiding," in *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers (Cat. No. CH37020)*, vol. 2. IEEE, 1999, pp. 1503–1507.

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

[16] P. Mitran and J. Bajcsy, "Coding for the wyner-ziv problem with turbo-like codes," *Proc. ISIT'02*, p. 91, 2002.

[17] R. Zamir, "The rate loss in the wyner-ziv problem," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 2073–2084, 1996.

[18] M. Gastpar, "The wyner-ziv problem with multiple sources," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2762–2768, 2004.

[19] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[20] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, 1998.

[21] E. Martinian and M. Wainwright, "Low density codes achieve the rate-distortion bound," in *Data Compression Conference (DCC'06)*. IEEE, 2006, pp. 153–162.

[22] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[23] V. Guruswami and P. Xia, "Polar codes: Speed of polarization and polynomial gap to capacity," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 3–16, 2014.

[24] G. He, J.-C. Belfiore, I. Land, G. Yang, X. Liu, Y. Chen, R. Li, J. Wang, Y. Ge, R. Zhang *et al.*, "Beta-expansion: A theoretical framework for fast and recursive construction of polar codes," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[25] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221–3227, 2012.

[26] Y. Zhou, R. Li, H. Zhang, H. Luo, and J. Wang, "Polarization weight family methods for polar code construction," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–5.

[27] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

[28] K. Nagano and T. Saba, "Polar coding with introducing half-frozen bits," in *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2018, pp. 1–6.

[29] L. Xiang, Y. Liu, Z. B. K. Egilmez, R. G. Maunder, L.-L. Yang, and L. Hanzo, "Soft list decoding of polar codes," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 921–13 926, 2020.

[30] T. V. Minh, T. J. Oechtering, and M. Skoglund, "Polar code for secure wyner-ziv coding," in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2016, pp. 1–6.