Exploiting Source Redundancy to Improve the Rate of Polar Codes

Ying Wang
ECE Department
Texas A&M University
yingwang@tamu.edu

Krishna R. Narayanan
ECE Department
Texas A&M University
krn@tamu.edu

Anxiao (Andrew) Jiang
CSE Department
Texas A&M University
ajiang@cse.tamu.edu

***Abstract*—We consider a joint source-channel decoding (JSCD) problem where the source encoder leaves residual redundancy in the source. We first model the redundancy in the source encoder output as the output of a side information channel at the channel decoder, and show that this improves random error exponent. Then, we consider the use of polar codes in this framework when the source redundancy is modeled using a sequence of t-erasure correcting block codes. For this model, the rate of polar codes can be improved by unfreezing some of originally frozen bits and that the improvement in rate depends on the distribution of frozen bits within a codeword. We present a proof for the convergence of that distribution, as well as the convergence of the maximum rate improvement. The significant performance improvement and improved rate provide strong evidences that polar code is a good candidate to exploit the benefit of source redundancy in the JSCD scheme.**

I. Introduction

We study how to improve the performance of channel codes using the *natural redundancy* in data. By natural redundancy, we refer to the inherent redundancy in data (e.g., features in languages and images) that is not artificially added for error correction. We focus on compressed languages here. Current works have shown that after compression of texts (at a compression ratio higher than practical systems), lots of natural redundancy still exists [1]. For example, after English texts are compressed by an LZW code with a dictionary of 2^{20} patterns (much larger than LZW dictionaries in many practical systems) and transmitted over a binary erasure channel (BEC), a decoding algorithm using only natural redundancy can reduce the noise in the compressed texts by over 85% for channel erasure rates from 5% to 30%. The natural redundancy can significantly improve the error correction capabilities of ECCs [2]–[7]. The topic is an extension of joint source-channel coding and denoising, which have been studied extensively.

Polar codes have gained lots of attention due to their capacity achieving property, low encoding/decoding complexity and good error floor performance [8]. In [4], we have proposed a joint source-channel coding scheme where the text source is compressed by a Huffman code, and then transmitted through a noisy channel after encoded by a polar code. We have shown that the proposed joint decoder, which combines a language decoder with the polar-code decoder, provides substantial improvement in performance over the CRC-aided list decoding, while the decoding complexity is kept in the same order as the list decoding scheme. The insight of using polar codes is that the language decoder and polar decoder both work over trees and that they can be combined in a computationally efficient way. Another insight comes from the sequential nature of polar decoding. In general, successive cancellation (SC) decoding of polar codes suffers from error propagation. Wrongly decoded bits in early stages may severely degrade the decoding performance of the whole sequence. To alleviate this, we can reorder the words before encoding and put more reliable sub-sequences in the front and suppress the error propagation. By cleverly arranging the order of information bits based on the reliability/recover ability, we are able to largely improve the overall performance. This makes polar codes more naturally suited for exploiting source redundancy than LDPC type codes.

We have shown a significant improvement in the performance of joint decoding of polar codes in [4]. The improvement in the finite length performance inspires us to explore how much rate of channel codes can be improved asymptotically by source redundancy. We first give a general model of decoding with side information, and show that the source redundancy can help improve the error exponent and rate of channel codes. Then we analyze in particular the rate improvement of polar codes. A theoretical model is studied for natural redundancy in compressed languages. That is, we model the source redundancy as a set of t erasure correcting block codes concatenated to the information bits. The block code is a simple and effective way to model the erasure correcting ability of words in the dictionary. Each block code corresponds to a word or a longer text. We show that the rate of polar codes can be improved by the source redundancy. The improvement in the rate depends on the distribution of frozen bits within a codeword. We obtained a lower bound on the improved rate of polar codes in [4], assuming the limit distribution exists. In this work, we formally prove that the distribution of frozen bits converges to a limit distribution. We propose an optimal information-bit allocation algorithm and analyze the maximum improvement in the rate of polar codes with the proposed algorithm.

II. Background

In this section we review polar codes, and the joint decoding scheme of polar codes with source redundancy proposed in [4].

This work was supported in part by the National Science Foundation under grants ECCS-1611285.

A. Polar codes

Let $n = 2^m$ be the length of polar codes. Polar codes are recursively constructed with the generator matrix $G_n = R_n G_2^{\otimes m}$, where R_n is an $n \times n$ bit-reversal permutation matrix, $G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, and $\otimes$ is the Kronecker product. Let u_1^n be the bits to be encoded, x_1^n be the coded bits and y_1^n be the received bits. Also, let $W(y|x)$ be the channel law. The Arikan's polar codes consist of two parts, namely channel combining and channel splitting. For channel combining, n copies of the channel are combined to create the channel

$$W_n(y_1^n|u_1^n) \triangleq W^n(y_1^n|u_1^n G_n) = \prod_{i=1}^{n} W(y_i|x_i)$$

due to the memoryless property of the channel. The channel splitting process splits W_n back into a set of n bit channels

$$W_m^{(i)}(y_1^n, u_1^{i-1}|u_i) \triangleq \frac{1}{2^{n-1}} \sum_{u_{i+1}^n} W_n(y_1^n|u_1^n), \; i = 1, \cdots, n.$$

Let $I(W)$ be the capacity of a discrete memoryless channel (DMC) defined by W. The channels $W_m^{(i)}$ will polarize in the sense that the fraction of bit channels for which $I(W_m^{(i)}) \to 1$ will converge to $I(W)$, and the remaining will have $I(W_m^{(i)}) \to 0$ as $n \to \infty$. The construction of Arikan's polar codes is to find a set of best channels $\mathcal{F}^c$ and transmit information only on those channels.

B. Joint decoding of polar codes for language-based sources

Motivated by the nice properties of polar codes, we study the potential of polar codes in the joint decoding scheme. An advantage of polar codes in the joint scheme is that list decoding of polar codes is based on the tree structure, which can be naturally combined with the tree structure of the dictionary. A joint decoding framework is given in [4]. The decoder employs joint list decoding to take advantage of a-priori information provided by the dictionary. Simulation results show that the scheme significantly outperforms list decoding of CRC-aided polar codes.

III. General Decoding Model With Source Redundancy

In a system where the source is compressed without removing all the redundancy, the leftover redundancy can act as the side information to help improve the decoding performance, while no change is made to the encoding structure. In our work, the source redundancy is modeled as the side information as shown in Fig. 1. There are two parallel channels, one transmitting the normal codewords, and the other transmitting the side information. The source $\mathbf{w}$ is first source encoded into $\mathbf{u}$ and channel encoded into $\mathbf{x}$, and then is transmitted through a channel to the decoder. The decoder receives $\mathbf{y}$ and has side information $\mathbf{v}$ available. $\mathbf{v}$ is correlated with the source $\mathbf{w}$ and the correlation depends on the redundancy left in the source.

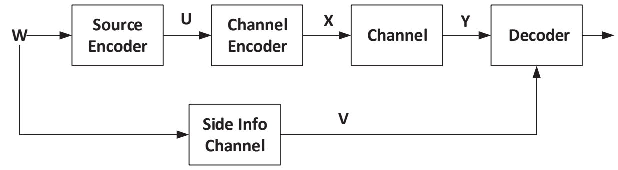

Fig. 1: A joint decoding model

Lemma 1. Assume the source is encoded by source and channel codes and transmitted through a DMC with transition probability $P(y|x)$. Each codeword is chosen i.i.d from distribution $\mathbf{Q}$. Denote n as the length of the code. Assume the side information $\mathbf{v}$ is available at the decoder with prior probability $p(\mathbf{v})$. The source $\mathbf{w}$ is transmitted through a channel with transition probability $p(\mathbf{v}|\mathbf{w})$. The average error probability of the ensemble is bounded by

$$\bar{P}_e \leq 2^{-nE_0(\rho,\mathbf{Q})+E_s(\rho)}, \rho \in (0,1],$$

where $E_0(\rho, \mathbf{Q})$ and $E_s(\rho)$ are defined as

$$E_0(\rho, \mathbf{Q}) = -\log \sum_j \left(\sum_k Q(k) P(j|k)^{1/(1+\rho)} \right)^{1+\rho},$$

$$E_s(\rho) = \log \left(\sum_{\mathbf{v}} p(\mathbf{v}) \left(\sum_{\mathbf{w}} p(\mathbf{w}|\mathbf{v})^{1/(1+\rho)} \right)^{1+\rho} \right).$$

Proof. Let $\bar{P}_{e,\mathbf{v}}$ denote the average error probability given side information $\mathbf{v}$. From Exercise 5.16 of [9] we know that if the prior probability of the source is available at the decoder, with MAP decoding the average error probability of the ensemble given $\mathbf{v}$ is bounded by

$$\bar{P}_{e,\mathbf{v}} \leq \left(\sum_{\mathbf{w}} p(\mathbf{w}|\mathbf{v})^{1/(1+\rho)} \right)^{1+\rho} 2^{-nE_0(\rho,\mathbf{Q})}. \tag{1}$$

The average error probability is derived as

$$\bar{P}_e = \sum_{\mathbf{v}} p(\mathbf{v}) \bar{P}_{e,\mathbf{v}}$$
$$\leq 2^{-nE_0(\rho,\mathbf{Q})} \sum_{\mathbf{v}} p(\mathbf{v}) \left(\sum_{\mathbf{w}} p(\mathbf{w}|\mathbf{v})^{1/(1+\rho)} \right)^{1+\rho}.$$

Then $E_s(\rho)$ can be derived from above. □

Example 1. Assume that the source $\mathbf{w}$ is a binary memoryless source (BMS) with length l, and the channel code has length n. Each symbol in the source is i.i.d. Assume that the channel to transmit the side information $\mathbf{v}$ is a BEC with erasure probability β. Then, $E_s(\rho)$ can be computed as

$$E_s(\rho) = l \log \sum_v p(v) \left(\sum_w Q(w|v)^{1/(1+\rho)} \right)^{1+\rho}$$
$$= l \log \left((1-\beta) + \beta 2^\rho \right)$$

Without loss of generality assume $R = \frac{l}{n}$, we have

$$E_s(\rho) = Rn \log \left((1-\beta) + \beta 2^\rho \right). \tag{2}$$

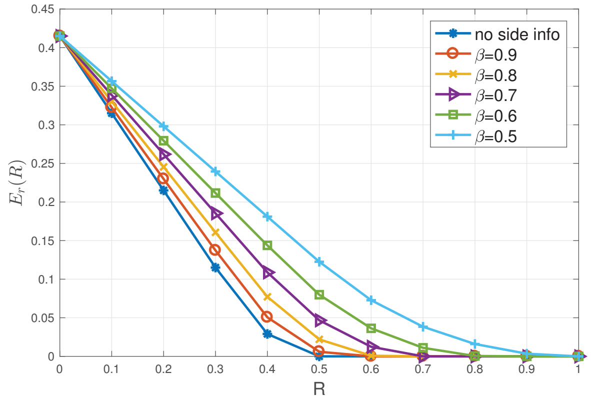

Fig. 2: Random error exponent for BEC with side information.

Assume the channel is BEC(ϵ) and the channel for side information is BEC(β). $E_0(\rho)$ is derived as $E_0(\rho) = -\log\left((1-\epsilon)2^{-\rho}+\epsilon\right)$, and $E_s(\rho)$ is given in (2). Let $E_r(R) = \max_{\mathbf{Q}} E_0(\rho,\mathbf{Q}) - E_s(\rho)/n$ denote the random exponent. We plot $E_r(R)$ in Fig. 2. It can be observed that the random exponent with $\beta \in [0,1)$ is larger than that without side information. The maximum rate for which $E_r(R) > 0$ is also improved by the side information.

The preceeding discussion shows that for the random code ensemble, the rate can be improved by side information available at decoder. In the following sections, we consider a practical joint decoding scheme with polar codes. We study a particular model for the side information and analyze the improved rate of polar codes for that model.

IV. Theoretical Model for Natural Redundancy in Languages

Let $[n_0, t]$ denote a block code of length n_0 that can correct t erasures. The natural redundancy in the source is modeled such that each block of n_0 consecutive output bits from the side information channel corresponds to a word in the dictionary or a longer text and is assumed to be a codeword of a $[n_0, t]$ outer code. Let $B = (b_0, b_1, b_2 \cdots)$ be the bits in a compressed text. We assume that B is a sequence of $[n_0, t]$ codes. The model is suitable for many compression algorithms, such as Huffman coding and LZW coding with a fixed dictionary of patterns, where every binary codeword in the compressed text represents a sequence of characters in the original text, and the natural redundancy in multiple adjacent codewords (e.g., the sparsity of valid words, phrases or grammatically correct sentences) can be used to correct erasures. Theoretically, the greater n_0 is, the better this model is for languages.

V. Bit-Allocation Algorithm for Optimal Rate Improvement of Polar Codes

The natural redundancy in compressed texts can be used to improve the rate of a polar code. Let $\mathcal{C}$ be an (n, k) polar code. Its encoder normally encodes n input bits $U = (u_0, u_1, \cdots, u_{n-1})$ – k of which are information bits, and $n-k$ of which are frozen bits – into a polar codeword $X = (x_0, x_1, \cdots, x_{n-1})$. Let $\mathcal{F} \subset \{0, 1, \cdots, n-1\}$ denote the indices of the frozen bits. Here, with natural redundancy in information bits, we can improve the code rate by *unfreezing some of the frozen bits*, namely, to use some frozen bits to also store information bits. Let $\mathcal{E} \subset \mathcal{F}$ denote the indices of the bits we unfreeze. We use the bits $\{u_i \mid i \in \{0, 1, \cdots, n-1\} - (\mathcal{F} - \mathcal{E})\}$ to store the bits in B, and use the sc decoding for error correction. Since B is a sequence of $[n_0, t]$ codes, to make decoding successful, the requirement is that for every $[n_0, t]$ code, its first $n_0 - t$ bits should be stored in bits (of U) whose indices are not in $\mathcal{F}$. The code-rate improvement is $\Delta R = |\mathcal{E}|/n$.

We have derived a lower bound to ΔR in [4]. Here we present a new algorithm for mapping the information bits $b_0, b_1, b_2 \cdots$ to the input bits of the polar code's encoder, which maximizes ΔR: (1) As initialization, let $\alpha \leftarrow -1$ and $\beta \leftarrow -1$. (2) For $j = 0, 1, 2 \cdots$, if $(j \mod n_0) < n_0 - t$ (which means b_j is one of the first $n_0 - t$ bits of an $[n_0, t]$ code), let $\alpha \leftarrow \min\{i \mid \alpha < i < n, i \notin \mathcal{F}\}$ and then assign b_j to u_α. Otherwise (which means b_j is one of the last t bits of an $[n_0, t]$ code), assign b_j as follows:

1) If $\{i \mid i > \alpha, i > \beta, i \in \mathcal{F}\} \neq \emptyset$, let $\beta \leftarrow \min\{i \mid i > \alpha, i > \beta, i \in \mathcal{F}\}$, and assign b_j to u_β.
2) If $\{i \mid i > \alpha, i > \beta, i \in \mathcal{F}\} = \emptyset$, let $\alpha \leftarrow \min\{i \mid \alpha < i < n, i \notin \mathcal{F}\}$ and then assign b_j to u_α.

The above algorithm ends when there is no more value to assign to α (namely, $\{i \mid \alpha < i < n, i \notin \mathcal{F}\} = \emptyset$). $\mathcal{E}$ includes all bits denoted by u_β. The algorithm maximizes $|\mathcal{E}|$ and therefore ΔR. It is a greedy algorithm, and has linear time complexity. We present a sketch of proof for its optimality: for every $[n_0, t]$ outer-codeword, call its first $n_0 - t$ bits pseudo-info bits, and call its last t bits pseudo-check bits. Since SC decoding decodes the bits of outer-codewords in the order of their increasing indices, in encoding, it is optimal to allocate the pseudo-info bits to non-frozen bits *codeword by codeword*, without interleaving them. For pseudo-check bits, it is optimal to allocate them to as many frozen bits as possible; and each time we unfreeze a frozen bit to store a pseudo-check bit, it is optimal to choose the feasible frozen bit of the minimum index. Due to space limitation, we skip the details of the proof.

VI. Convergence of Frozen-Bit Distribution in Polar Codes

The improvement in code rate depends on the distribution of frozen bits in the polar code. (Generally speaking, the further "behind" the frozen bits' positions are in the polar code, the more code-rate improvement can be achieved because more frozen bits can be used in the $[n_0, t]$ codes.) We characterize the distribution of frozen bits as follows. Given a real number y, let y^- be $y-1$ if y is an integer, and be $\lfloor y \rfloor$ otherwise. $\forall x \in [0, 1]$, let $\mathcal{F}_n(x)$ be the frozen set (i.e., indices of frozen bits) among the first $(nx)^- + 1$ bits of the polar code. Formally, given an arbitrary small $\epsilon \in [0, 1)$,

$$\mathcal{F}_n(x) \triangleq \{i \in [0, (nx)^-] : I(W_m^{(i)}) \in [0, 1-\epsilon)\}.$$

Let $f_n(x) \triangleq \frac{|\mathcal{F}_n(x)|}{nx} \in [0,1]$, which describes the distribution of frozen bits (or more specifically, the proportion of frozen bits among the first nx bits). We show the convergence of $f_n(x)$ in the following theorem.

Theorem 2. For a polar code with length $n = 2^m$ and rate $R = I(W) - \delta$ where δ is an arbitrary small positive number, the function $f_n(x)$ converges as $n \to \infty$.

Proof. Assume a polar code has length $n = 2^m$ and rate R. Let $\mathcal{B}_n(x)$, $\mathcal{M}_n(x)$ and $\mathcal{A}_n(x)$ be defined as follows: for any arbitrary small $\epsilon \in [0,1)$,

$$\begin{aligned}\mathcal{B}_n(x) &\triangleq \{i \in [0,(nx)^-] : I(W_m^{(i)}) \in [0,\epsilon]\},\\ \mathcal{M}_n(x) &\triangleq \{i \in [0,(nx)^-] : I(W_m^{(i)}) \in (\epsilon,1-\epsilon)\},\\ \mathcal{A}_n(x) &\triangleq \{i \in [0,(nx)^-] : I(W_m^{(i)}) \in [1-\epsilon,1]\}.\end{aligned}$$

We have $\mathcal{F}_n(x) = \mathcal{B}_n(x) \cup \mathcal{M}_n(x)$. Define $g_n(x)$, $t_n(x)$ and $h_n(x)$ as

$$h_n(x) = \frac{|\mathcal{B}_n(x)|}{nx},\ t_n(x) = \frac{|\mathcal{M}_n(x)|}{nx},\ \text{and}\ g_n(x) = \frac{|\mathcal{A}_n(x)|}{nx}.$$

From the definition of frozen set, we have $f_n(x) = h_n(x) + t_n(x) = 1 - g_n(x)$. As $n \to \infty$, almost all channels polarize in the sense that $g_n(1) \to I(W)$, $h_n(1) \to 1 - I(W)$ and $t_n(1) \to 0$. For polar code of length $n = 2^m$, if we do one more step of transformation, the length of the code is increased to $2n$. The ith bit channel $W_m^{(i)}$ is converted to $W_{m+1}^{(2i)}$ and $W_{m+1}^{(2i+1)}$, where $I(W_{m+1}^{(2i)}) \le I(W_m^{(i)}) \le I(W_{m+1}^{(2i+1)})$. Define $\Delta_I = I(W_m^{(i)}) - I(W_{m+1}^{(2i)})$. We know that $\Delta_I = I(W_{m+1}^{(2i+1)}) - I(W_m^{(i)})$. By computing Δ_I for the two extreme channels (BEC and BSC), upper and lower bounds of Δ_I can be derived. The maximum value of Δ_I is achieved for BEC:

$$\Delta_I^{max} = I(W_m^{(i)}) - I^2(W_m^{(i)}),$$

and the minimum value is achieved for BSC:

$$\Delta_I^{min} = H(2p(1-p)) - H(p)$$

where $I(W_m^{(i)}) = 1 - H(p)$, and $H(p)$ is the entropy of a BSC with crossover probability p.

Consider the ith bit channel, if $i \in \mathcal{A}_n(x)$, $I(W_m^{(i)})$ is bounded by $1 - \epsilon \le I(W_m^{(i)}) \le 1$. It is easily seen that $2i+1 \in \mathcal{A}_{2n}(x)$ as $I(W_m^{(i)}) \le I(W_{m+1}^{(2i+1)})$. Since $I(W_{m+1}^{(2i)}) \ge I(W_m^{(i)})^2$, we have $I(W_{m+1}^{(2i)}) \ge (1-\epsilon)^2$. Since $(1-\epsilon)^2 > \epsilon$ for $\epsilon < \frac{3-\sqrt{5}}{2} = 0.382$, if $\epsilon < 0.382$, it is guaranteed that $I(W_{m+1}^{(2i)}) > \epsilon$, which indicates $2i \notin \mathcal{B}_{n+1}(x)$. Similarly, if $i \in \mathcal{B}_n(x)$, $I(W_m^{(i)})$ satisfies $0 \le I(W_m^{(i)}) \le \epsilon$. Since $I(W_{m+1}^{(2i+1)}) \le 2I(W_m^{(i)}) - I(W_m^{(i)})^2$, we have $I(W_{m+1}^{(2i+1)}) \le 2\epsilon - \epsilon^2$. Since $2\epsilon - \epsilon^2 < 1 - \epsilon$ for $\epsilon < \frac{3-\sqrt{5}}{2} = 0.382$, if $\epsilon < 0.382$, it is guaranteed that $I(W_{m+1}^{(2i+1)}) < 1 - \epsilon$, indicating $2i+1 \notin \mathcal{A}_{2n}(x)$. In summary, if $\epsilon < 0.382$, the following constraints are satisfied: if $i \in \mathcal{A}_n(x), 2i \notin \mathcal{B}_{2n}(x)$, and if $i \in \mathcal{B}_n(x), 2i+1 \notin \mathcal{A}_{2n}(x)$. We can conclude that for n large enough the following hold:

1) If $i \in \mathcal{B}_n(x)$, $2i \in \mathcal{B}_{2n}(x)$, and $2i+1$ can be either in $\mathcal{B}_{2n}(x)$ or $\mathcal{M}_{2n}(x)$;
2) If $i \in \mathcal{A}_n(x)$, $2i+1 \in \mathcal{A}_{2n}(x)$, and $2i$ can be either in $\mathcal{A}_{2n}(x)$ or $\mathcal{M}_{2n}(x)$;
3) If $i \in \mathcal{M}_n(x)$, $2i$ can be in either $\mathcal{M}_{2n}(x)$ or $\mathcal{B}_{2n}(x)$; $2i+1$ can be in either $\mathcal{M}_{2n}(x)$ or $\mathcal{A}_{2n}(x)$.

From above facts, $|\mathcal{A}_{2n}(x)|$ can be bounded by

$$|\mathcal{A}_{2n}(x)| \le 2|\mathcal{A}_n(x)| + |\mathcal{M}_n(x)|.$$

Thus $g_{2n}(x)$ is bounded by

$$\begin{aligned}g_{2n}(x) = \frac{|\mathcal{A}_{2n}(x)|}{2nx} &\le \frac{|\mathcal{A}_n(x)|}{nx} + \frac{|\mathcal{M}_n(x)|}{2nx}\\ &= g_n(x) + \frac{1}{2}t_n(x).\end{aligned}$$

We have a lower bound for $f_{2n}(x)$

$$f_{2n}(x) = 1 - g_{2n}(x) \ge f_n(x) - \frac{1}{2}t_n(x) \tag{3}$$

On the other hand, $|\mathcal{B}_{2n}(x)|$ is upper bounded by

$$|\mathcal{B}_{2n}(x)| \le 2|\mathcal{B}_n(x)| + |\mathcal{M}_n(x)|.$$

Thus $h_{2n}(x)$ is bounded by

$$\begin{aligned}h_{2n}(x) = \frac{|\mathcal{B}_{2n}(x)|}{2nx} &\le h_n(x) + \frac{1}{2}t_n(x)\\ &= f_n(x) - \frac{1}{2}t_n(x),\end{aligned}$$

and we get an upper bound of $f_{2n}(x)$

$$f_{2n}(x) = h_{2n}(x) + t_{2n}(x) \le f_n(x) - \frac{1}{2}t_n(x) + t_{2n}(x). \tag{4}$$

Combining (3) and (4), $f_{2n}(x)$ is bounded by

$$f_n(x) - \frac{1}{2}t_n(x) \le f_{2n}(x) \le f_n(x) - \frac{1}{2}t_n(x) + t_{2n}(x).$$

By induction, if taking s steps of transformation for any $s > 0$, $f_{2^s n}(x)$ is bounded by

$$\begin{aligned}&f_n(x) - \frac{1}{2}\sum_{i=1}^{s} t_{2^{i-1}n}(x) \le f_{2^s n}(x)\\ &\le f_n(x) + \frac{1}{2}\left(\sum_{i=1}^{s} t_{2^{i-1}n}(x)\right) - t_n(x) + t_{2^s n}(x)\end{aligned} \tag{5}$$

By taking limit of both sides of (5), we get

$$\begin{aligned}&\lim_{n\to\infty} f_n(x) - \frac{1}{2}\lim_{n\to\infty}\sum_{i=1}^{s} t_{2^{i-1}n}(x) \le \lim_{n\to\infty} f_{2^s n}(x)\\ &\le \lim_{n\to\infty} f_n(x) + \frac{1}{2}\lim_{n\to\infty}\left(\sum_{i=1}^{s} t_{2^{i-1}n}(x)\right)\\ &- \lim_{n\to\infty} t_n(x) + \lim_{n\to\infty} t_{2^s n}(x)\end{aligned}$$

Since almost all bit channels will polarize in the limit of blocklength, $\lim_{n\to\infty} t_n(x) = 0$. Finally it can be derived

$$\lim_{n\to\infty} f_{2^s n}(x) = \lim_{n\to\infty} f_n(x).$$

□

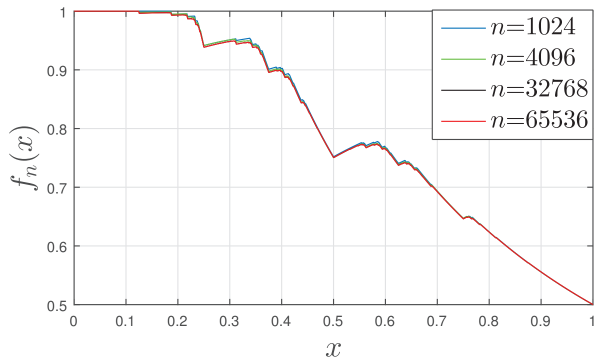

Fig. 3: Distributions of frozen bits with different code lengths.

We show $f_n(x)$ for rate 1/2 polar codes with different code lengths in Fig. 3. It can be seen that the distribution is very close to each other, which validates Theorem 2.

VII. Rate Improvement

In this section, we show the convergence of the improved rate of polar codes.

Theorem 3. Assume a polar code has length n and rate $R = I(W) - \delta$, where δ is an arbitrary small positive number. Let the polar code be outer coded by a set of $[n_0, t]$ block codes, where n_0 and t scales linearly with n. If the frozen bit distribution $f_n(x)$ converges to some function $f(x)$, the improved rate of polar codes ΔR_n converges as $n \to \infty$.

Sketch of proof. Consider the optimal bit allocation scheme in Section V. The bits are grouped into $[n_0, t]$ outer codes sequentially, where the first $n_0 - t$ bits are information bits and last t bits are what can be unfrozen. Let nx_i and $ny_i, i \in [1, s]$ be the last information bit and last bit in the ith outer code, respectively. Given the convergence of $f_n(x)$ and linearity of n_0 and t with n, we can show the convergence of x_i and $y_i, i \in [1, s]$, and thus the convergence of s. Since $\Delta R_n = \frac{st+c}{n}$ where $c < t$, the convergence of ΔR_n follows. □

Let ΔR_{max} denote the optimal code-rate improvement achieved by the bit allocation algorithm. We illustrate examples of ΔR_{max} for rate 1/2 polar codes of different lengths in Fig. 4, for three different $[n_0, t]$ outer codes: $[2, 1]$, $[7, 2]$ and $[8, 3]$ codes. It can be seen that ΔR_{max} converges as n increases. The convergence of the code-rate improvement (for both the algorithm here and the lower bound result in [4]) depends on the convergence of $f_n(x)$.

TABLE I: Comparison of improved rates

	ΔR_{max}	ΔR_l
$[2, 1]$ outer codes	0.181	0.125
$[7, 2]$ outer codes	0.090	0.060
$[8, 3]$ outer codes	0.124	0.076

The code-rate improvement by the optimal bit allocation algorithm exceeds the lower bound to ΔR (denoted by ΔR_l) in [4], which was derived constructively using a piecewise linear approximation of the distribution function $f_n(x)$. A comparison of the two code-rate improvements is shown in Table I, for sufficiently large polar-code length n.

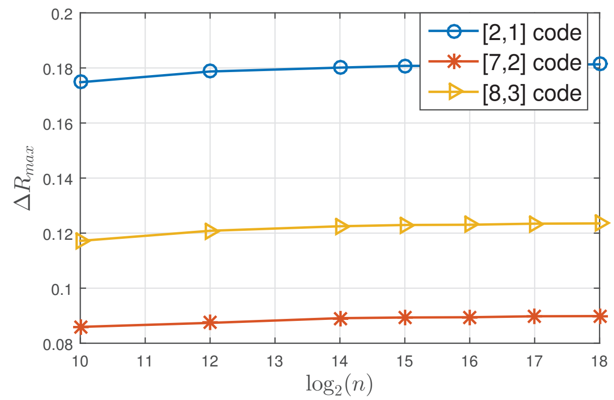

Fig. 4: Improved rates with different outer $[n_0, t]$ codes. (The $[2, 1]$, $[7, 2]$, $[8, 3]$ outer codes can correct 1, 2, 3 erasures.)

VIII. Conclusion

Source redundancy is exploited to improve the decoding of channel codes. We model the source redundancy as side information, and show the improved random coding exponent and improved rate. Practically, polar codes are explored to do joint decoding with natural redundancy. We show the improved rate of polar codes. We give a particular model for natural redundancy in languages, and propose an optimal information-bit allocation algorithm to achieve the maximum improved rate. A proof of the convergence of frozen bit distribution is given, and we show the convergence of maximum improved rate. We can extend this work to model the source redundancy as block codes with different lengths and erasure correcting abilities. By reordering the information bits in the word level before encoding based on the erasure correcting ability of the outer codes, better performance can be achieved.

References

[1] A. Jiang, P. Upadhyaya, E. Haratsch, and J. Bruck, "Error correction by natural redundancy for long term storage," in *Proc. NVMW*, 2017.
[2] A. Jiang, Y. Li, and J. Bruck, "Error correction through language processing," in *Proc. IEEE ITW*, 2015, pp. 1–5.
[3] ——, "Enhanced error correction via language processing," in *Proc. NVMW*, 2015.
[4] Y. Wang, M. Qin, K. R. Narayanan, A. Jiang, and Z. Bandic, "Joint source-channel decoding of polar codes for language-based sources," in *Proc. IEEE Globecom*, 2016.
[5] Y. Wang, A. Jiang, and K. R. Narayanan, "Modeling and analysis of joint decoding of language-based sources with polar codes," in *Proc. NVMW*, 2017.
[6] P. Upadhyaya and A. Jiang, "LDPC decoding with natural redundancy," in *Proc. NVMW*, 2017.
[7] Y. Li, Y. Wang, A. Jiang, and J. Bruck, "Content-assisted file decoding for nonvolatile memories," in *Proc. 46th Asilomar Conference on Signals, Systems and Computers*, pp. 937–941, 2012.
[8] E. Arıkan, "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, 2009.
[9] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.