POLAR CODES

BRICE HUANG

Abstract. Shannon's celebrated Channel Coding Theorem proves that a channel can support codes with arbitrarily small error and rate arbitrarily close to its channel capacity. However, Shannon's proof is nonconstructive, and does not specify a family of codes achieving these error and rate properties. Recently, Arikan provided the first constructive proof of the Channel Coding Theorem, using so-called polar codes. These are the first explicit family of codes to satisfy the error and rate properties specified in the Channel Coding Theorem. This paper surveys Arikan's construction.

1. Introduction

The field of coding theory addresses how to efficiently send information over a channel. In this model, two parties are communicating over a noisy channel; since the channel is noisy, each symbol the channel outputs is a probabilistic function of the corresponding input. A natural goal is to use this channel to transmit information in a reliable way.

Formally, let the channel W have input alphabet $\mathscr{X}$ and output alphabet $\mathscr{Y}$, and let the channel input and output be random variables X, Y. The channel's behavior is characterized by the conditional distribution $W(y|x) = \mathbb{P}[\mathsf{Y} = y|\mathsf{X} = x]$. We assume the channel is memoryless, so each output symbol depends only on its corresponding input symbol. Recall that the *channel capacity* is $C(W) = \max_{\mathsf{X}\sim p(x)} I(\mathsf{X};\mathsf{Y})$.

Recall that a (M, n)-*code* consists of an encoder $X^n : \{1, \dots, M\} \to \mathscr{X}^n$ and a decoder $g : \mathscr{Y}^n \to \{1, \dots, M\}$. Codewords are blocks of n symbols, transmitted one at a time. To transmit a message $m \in \{1, \dots, M\}$, the sender transmits $X^n(m)$ over the channel; the receiver receives y^n from the channel, and decodes the message as $\hat{m} = g(y^n)$. The *average* and *maximum rates of error* are $P_e^{(n)} = \sum_m \mathbb{P}[\hat{m} \neq m]$ and $\lambda^{(n)} = \max_m \mathbb{P}[\hat{m} \neq m]$, and the code *rate* is $R = \frac{\log_2 M}{n}$, the number of bits transmitted per symbol.

The Channel Coding Theorem, proved by Shannon in 1948, states that there are codes with arbitrarily low error probability and rate arbitrarily close to the channel capacity.

Theorem 1.1 (Channel Coding Theorem)**.** [2] *For every rate $R < C(W)$, there exists a sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$.*

Shannon's original proof of this theorem relies on a probabilistic argument, that a randomly-generated $(2^{nR}, n)$ code, decoded with jointly-typical decoding, has low error with positive probability. This proof does not construct an explicit code family with rate R and error probability limiting to 0, and no such code family was known for over a half century.

Date: May 16, 2019.
Key words and phrases. Channel Coding, Polar Codes.

In 2009, Arikan constructively proved the Channel Coding Theorem, under mild assumptions. Arikan's construction is specific to binary codes, where $\mathscr{X} = \{0,1\}$ (but $\mathscr{Y}$ is still arbitrary); henceforth we assume $\mathscr{X} = \{0,1\}$. More critically, Arikan's construction constructs codes with rate up to the *symmetric channel capacity* $I(W)$, defined to be $I(\mathsf{X};\mathsf{Y})$ where X is uniform on $\{0,1\}$. Explicitly,

$$I(W) = \sum_{x\in\mathscr{X},y\in\mathscr{Y}} p(x,y)\log_2\frac{p(y|x)}{p(y)} = \sum_{x\in\mathscr{X},y\in\mathscr{Y}} \frac{1}{2}W(y|x)\log_2\frac{W(y|x)}{\frac{1}{2}\left(W(y|0)+W(y|1)\right)}. \tag{1.1}$$

Note that $I(W) \le C(W)$ in general; however, $I(W) = C(W)$ in the important case that the channel W is *symmetric* [1]. We say W is symmetric if there is an involution π on the output symbols such that $W(y|1) = W(\pi(y)|0)$ for all $y \in \mathscr{Y}$; intuitively, if we swap 0 and 1 and relabel the output symbols by π, we get the same channel.

For rate $R < I(W)$ and block size n, Arikan's construction produces an ensemble of *polar codes*, each specified by a setting of so-called *frozen bits*. Let $P_e(n,R)$ denote the mean value of P_e over this ensemble.

Theorem 1.2. [1] *For any rate $R < I(W)$, the average error probability of the polar code ensemble, under successive cancellation decoding, satisfies $P_e(n,R) = O(n^{-1/4})$.*

Remark 1.1. This proof is slightly nonconstructive, as it only shows that some code in the polar code ensemble has $P_e^{(n)} = O(n^{-1/4})$. However, it can be shown that for a symmetric channel, the values of $P_e^{(n)}$ for all codes in the polar code ensemble are the same. Thus for symmetric channels, taking any member of the polar code ensemble gives a fully explicit construction.

Remark 1.2. When W is symmetric, $I(W) = C(W)$; in this case, we have a fully explicit construction of a code family with any rate $R < C(W)$ and $P_e^{(n)} = O(n^{-1/4})$. By throwing out the half of the codewords with highest error probabilities, we can in fact convert this code family to one with rate R and maximum error probability $\lambda^{(n)} = O(n^{-1/4})$.

The rest of this paper is structured as follows. In section 2 we introduce the channel polarization operation and discuss its properties; these properties underpin polar coding's effectiveness. In section 3 we construct the polar code ensemble, and in section 4 we prove Theorem 1.2. Finally in section 5 we discuss this result in the context of coding theory.

2. Channel Polarization

In this section we will discuss *channel polarization*, the operation enabling the construction of polar codes. At a high level, this operation takes n independent copies of the channel W and synthesizes n artificial channels $W_n^1, \dots, W_n^n$ such that, for all but a vanishing fraction of these channels as n becomes large, the symmetric capacities $I(W_n^i)$ tend to 0 or 1. This operation concentrates the information capacity of the n copies of W into about $NI(W)$ artificial channels, each with information capacity close to 1. Polar coding then sends information along these high-capacity artificial channels.

2.1. The Polarization Operation.

The basic polarization operation f takes two independent copies of a channel $W : \mathscr{X} \to \mathscr{Y}$ and synthesizes a low-information channel

$W_0 : \mathscr{X} \to \mathscr{Y}^2$ and a high-information channel $W_1 : \mathscr{X} \to \mathscr{Y}^2 \times \mathscr{X}$. First, this operation combines the two copies of W into a merged channel $W' : \mathscr{X}^2 \to \mathscr{Y}^2$, defined by

$$W'(y_0, y_1|u_0, u_1) = W(y_0|u_0 \oplus u_1)W(y_1|u_1). \tag{2.1}$$

In other words, the merged channel W' encodes $(u_0, u_1) \in \mathscr{X}^2$ by $(u_0 \oplus u_1, u_1)$, and sends this encoding along the two copies of W. We then split W' into the channels W_0, W_1, given by

$$W_0(y_0, y_1|u_0) = \sum_{u_1} \frac{1}{2} W'(y_0, y_1|u_0, u_1) \tag{2.2}$$

$$W_1(y_0, y_1, u_0|u_1) = \frac{1}{2} W'(y_0, y_1|u_0, u_1). \tag{2.3}$$

We denote this operation with the notation $f(W, W) = (W_0, W_1)$.

Intuitively, W_0 represents the conditional distribution the decoder observes when he tries to decode u_0 given the channel output (y_0, y_1), and W_1 represents the conditional distribution the decoder observes when he tries to decode u_1 given (y_0, y_1) and oracle access to u_0.

Let us motivate this formulation of W_1. Since a decoding error in *any* position makes the decoding incorrect, a decoder trying to decode (u_1, u_2) from (y_1, y_2) does not need the oracle access to u_1. Instead, he first decodes u_0 using $W_0(y_0, y_1|u_0)$, and then, assuming his deocding of u_0 is correct, decodes u_1 using $W_1(y_0, y_1, u_0|u_1)$.

2.2. **Properties of Polarization.** We introduce the *Bhattacharyya parameter* of a channel W,

$$Z(W) = \sum_{y \in \mathscr{Y}} \sqrt{W(y|0)W(y|1)}. \tag{2.4}$$

This is a measure of the reliability of W, in the sense that

$$\frac{1}{2} Z(W) \geq \frac{1}{2} \sum_{y \in \mathscr{Y}} \min(W(y|0), W(y|1)) \tag{2.5}$$

and the right-hand side is the maximum-likelihood decision error when W transmits a single bit.

Both $Z(W)$ and $I(W)$ are real values in $[0, 1]$. Intuitively, we expect $I(W)$ to be close to 1 when $Z(W)$ is close to 0, and vice versa. The following lemma, whose proof we omit, formalizes this intuition.

Lemma 2.1. [1] *For all binary channels W,* $\log_2 \frac{2}{1+Z(W)} \leq I(W) \leq \sqrt{1 - Z(W)^2}$.

The following lemma characterizes how polarization affects the symmetric capacity and Bhattacharyya parameter.

Lemma 2.2. [1] *If $f(W, W) = (W_0, W_1)$, then*

$$2I(W) = I(W_0) + I(W_1) \tag{2.6}$$

$$Z(W_1) = Z(W)^2 \tag{2.7}$$

$$Z(W_0) = 2Z(W) - Z(W)^2 \tag{2.8}$$

$$I(W_0) \leq I(W) \leq I(W_1) \tag{2.9}$$

$$Z(W_0) \geq Z(W) \geq Z(W_1) \tag{2.10}$$

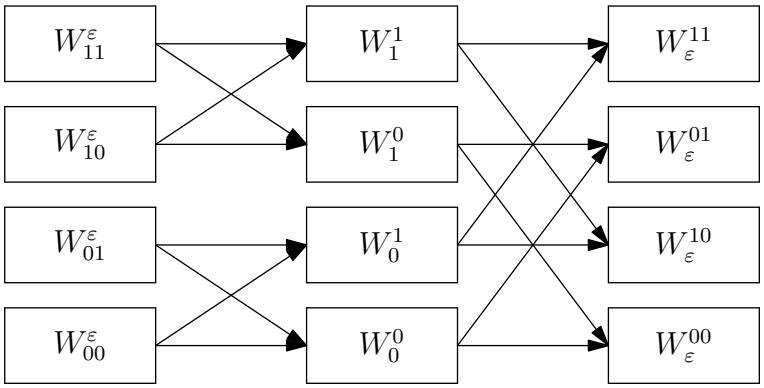

FIGURE 1. Recursive Polarization when $n = 4$

2.3. **Recursive Polarization.** The key insight of Arikan's construction is that the polarization operation can be applied recursively. Our construction begins with $n = 2^m$ independent copies of the channel W. We will construct a recursive family of channels indexed by W_s^t, indexed by all pairs of bit-strings s, t with combined length m. Throughout, let $\ell(s)$ denote the length of bit-string s and $n(s)$ denote s interpreted as an integer written in binary.

The channels W_s^{ε}, where ε is the empty string and s ranges over all m-bit strings, are the 2^m original copies of W. For all pairs of bit-strings s', t' with combined length $m-1$, we let

$$f(W_{s'0}^{t'}, W_{s'1}^{t'}) = (W_{s'}^{t'0}, W_{s'}^{t'1}). \tag{2.11}$$

Intuitively, these codes W_s^t form $m+1$ layers, classified by the lengths of s and t, each with 2^m channels. For each fixed t, the channels of the form W_s^t, as s ranges over all bit-strings of length $m - \ell(t)$, are identically distributed. Thus, the layer where $\ell(t) = k, \ell(s) = m - k$ consists of 2^k different types of channels and 2^{m-k} independent copies of each type. When we advance a layer, we pair up independent copies and apply polarization to each pair, causing the number of types of channels to double and the number of independent copies of each type to halve. The recursive-polarized channels are the channels W_{ε}^t in the final layer, as t ranges over all bit-strings of length m.

Example 2.1. Let $m = 2, n = 2^m = 4$. Figure 1 shows how the four original copies of W, on the left, give rise to the four polarized copies on the right. In each layer, channels pair up and each pair yields two channels in the subsequent layer.

Here is another characterization of recursive polarization, which we can show is equivalent. For $k = 0, 1, \ldots, m$, we will define the channels W_k, which merge 2^k copies of W. The channel W_0 is just W. Let the inputs and outputs of W_k be $(u_0, \ldots, u_{2^k-1})$ and $(y_0, \ldots, y_{2^k-1})$. The channel W_k is built on two independent copies of W_{k-1}; it inputs $u_0 \oplus u_1, u_2 \oplus u_3, \ldots, u_{2^k-2} \oplus u_{2^k-1}$ into the first copy and $u_1, u_3, \ldots, u_{2^k-1}$ into the second copy. Formally,

$$\begin{aligned} &W_k(y_0, \ldots, y_{2^k-1} | u_0, \ldots, u_{2^k-1}) \\ &= W_{k-1}(y_0, \ldots, y_{2^{k-1}-1} | u_0 \oplus u_1, \ldots, u_{2^k-2} \oplus u_{2^k-1}) W_{k-1}(y_0, \ldots, y_{2^{k-1}-1} | u_1, \ldots, u_{2^k-1}). \end{aligned} \tag{2.12}$$

Example 2.2. When $n = 4$, the channel W_2 is given by

$$W_2(y_0, y_1, y_2, y_3|u_0, u_1, u_2, u_3) = W_1(y_0, y_1|u_0 \oplus u_1, u_2 \oplus u_3)W_1(y_2, y_3|u_1, u_3) \tag{2.13}$$

$$= W(y_0|u_0 \oplus u_1 \oplus u_2 \oplus u_3)W(y_1|u_2 \oplus u_3)W(y_2|u_1 \oplus u_3)W(y_3|u_3). \tag{2.14}$$

In the implementation of this code, the values $u_0 \oplus u_1 \oplus u_3 \oplus u_3, u_2 \oplus u_3, u_1 \oplus u_3, u_3$ are transmitted into the underlying four channels W. This is a $\mathbb{F}_2$-linear code, where the linear operator is $\begin{pmatrix} 1&0&0&0\\ 1&0&1&0\\ 1&1&0&0\\ 1&1&1&1 \end{pmatrix}$. In general, the channels W_k are 2^k-wide $\mathbb{F}_2$-linear block codes.

The channels W_ε^t have type $W_\varepsilon^t : \mathscr{X} \to \mathscr{Y}^{2^m} \times \mathscr{X}^{n(t)}$, and are given by

$$W_\varepsilon^t(y_0, \dots, y_{2^m-1}, u_0, \dots, u_{n(t)-1}|u_{n(t)}) = \sum_{u_{n(t)+1}, \dots, u_{2^m-1}} \frac{1}{2^{N-1}} W_m(y_1, \dots, y_{2^k-1}|u_0, \dots, u_{2^k-1}). \tag{2.15}$$

The proof that these formulations is equivalent is technical, and we only sketch it here. For fixed s of length $\ell(s) = m - k$, let the input and output of the channel W_s^t be $u_{n(t)}$ and $y_{n(t)}$. So, the collective input and ouptut of the W_s^t, as t ranges over k-bit strings, is $(u_0, \dots, u_{2^k-1})$ and $(y_0, \dots, y_{2^k-1})$. We can prove the following claim by induction on k: W_s^t has type $W_s^t : \mathscr{X} \to \mathscr{Y}^{2^k} \times \mathscr{X}^{n(t)}$, and is given by

$$W_s^t(y_0, \dots, y_{2^k-1}, u_0, \dots, u_{n(t)-1}|u_{n(t)}) = \sum_{u_{n(t)+1}, \dots, u_{2^k-1}} \frac{1}{2^{N-k-1}} W_k\left(y_0, \dots, y_{2^k-1}|u_0, \dots, u_{2^k-1}\right). \tag{2.16}$$

Equivalence follows from setting $k = m$ in this claim.

The key theorem about channel polarization is as follows.

Theorem 2.3. *Fix $\delta \in (0, 1)$. As $m \to \infty$, the fraction of channels W_ε^t with $I(W_\varepsilon^t) \in (1-\delta, 1]$ goes to $I(W)$, and the fraction of channels W_ε^t with $I(W_\varepsilon^t) \in [0, \delta)$ goes to $1 - I(W)$.*

Proof. Let us define a random infinite sequence of channels $K_0, K_1, \dots$. We let $K_0 = W_\varepsilon^\varepsilon = W$; for each $n \in \mathbb{N}$, if $K_n = W_\varepsilon^t$, then $K_{n+1} = W_\varepsilon^{t0}$ or $K_{n+1} = W_\varepsilon^{t1}$, each with probability $\frac{1}{2}$. Thus, K_n is a uniformly random channel of the form W_ε^t, where $\ell(t) = n$. Let us consider the sequences of random variables $I_n = I(K_n)$ and $Z_n = Z(K_n)$, for $n \in \mathbb{N}$.

The sequence Z_n is a supermartingale, in the sense that $\mathbb{E}[Z_n] < \infty$ and $Z_n \geq \mathbb{E}[Z_{n+1}|Z_n]$ in the appropriate probability measure. The expectation is bounded because $Z_n \in [0, 1]$, and the supermartingale bound holds because by Lemma 2.2,

$$\mathbb{E}[Z_{n+1}|Z_n] \leq \frac{1}{2}Z_n^2 + \frac{1}{2}(2Z_n - Z_n^2) = Z_n. \tag{2.17}$$

Since this supermartingale is uniformly integrable, it converges almost surely and in $\mathcal{L}^1$ to a random variable Z_∞ such that $\mathbb{E}\left[|Z_n - Z_\infty|\right] \to 0$. Thus $\mathbb{E}\left[|Z_{n+1} - Z_n|\right] \to 0$. But, by Lemma 2.2 $Z_{n+1} = Z_n^2$ with probability $\frac{1}{2}$, so

$$\mathbb{E}\left[|Z_{n+1} - Z_n|\right] \geq \frac{1}{2}\mathbb{E}\left[|Z_n^2 - Z_n|\right] \geq 0, \tag{2.18}$$

so $\mathbb{E}\left[|Z_n^2 - Z_n|\right] \to 0$, which implies $\mathbb{E}\left[|Z_\infty^2 - Z_\infty|\right] = 0$. Thus $Z_\infty \in \{0, 1\}$ almost surely.

The seuqnece I_n is a martingale, in the sense that $\mathbb{E}[I_n] < \infty$ and $I_n = \mathbb{E}[I_{n+1}|I_n]$ in the appropriate probability measure. The expectation is bounded because $I_n \in [0, 1]$, and the

martingale identity holds because by Lemma 2.2,

$$\mathbb{E}[I_{n+1}|I_n] = \frac{1}{2}(2I_n) = I_n. \tag{2.19}$$

Since this martingale is uniformly integrable, it converges almost surely to a random variable I_∞ with $\mathbb{E}[I_\infty] = I_0$.

Finally, since $Z_\infty \in \{0, 1\}$ almost surely, by Lemma 2.1 $I_\infty \in \{0, 1\}$ almost surely as well. Thus, $\mathbb{P}[I_\infty = 1] = I_0$ and $\mathbb{P}[I_\infty = 0] = 1 - I_0$. Since $I_0 = I(W)$, we are done. □

3. Construction of the Polar Code

We will use the polarization effect discussed in Section 2 to construct the polar code. Intuitively, recursive polarization transforms N channels, each with symmetric capacity $I(W)$, into $I(W)N$ high-information channels with capacity near 1 and $(i-I(W))N$ low-information channels with capacity near 0. The polar code sends code bits in the high-information channels and predetermined frozen bits in the low-information channels.

As before, let $n = 2^m$, and use recursive polarization to construct the channels W_ε^t for m-bit strings t; let us rename these channels $W_n^0, W_n^1, \ldots, W_n^{n-1}$, where W_ε^t is renamed $W_n^{n(t)}$. Thus W_n^i encodes the distribution $W_n^i(y_0, \ldots, y_{n-1}, u_0, \ldots, u_{i-1}|u_i)$.

Let us first develop the so-called *coset codes*.

Definition 3.1. Fix a set $\mathcal{A} \subset [n] = \{0, \ldots, n-1\}$. In a coset code, we send *information bits*, the bits we would like to trasnmit, along the channels $\{W_n^i | i \in \mathcal{A}\}$, and predetermined *frozen bits* along the channels $\{W_n^i | i \in \overline{\mathcal{A}}\}$, where $\overline{\mathcal{A}} = [n] \setminus \mathcal{A}$.

This is called a coset code because in the recursive polarization construction, the bits transmitted through the n underlying channels are a linear function of the input, forming a linear block code, and this code's codewords are a coset of that linear block code.

Since we have choice in which frozen bits to send, each coset code is parametrized by a tuple $(n, \mathcal{A}, \{u_i | i \in \overline{\mathcal{A}}\})$. Note that this code has rate $\frac{|\mathcal{A}|}{n}$.

Example 3.1. The $(4, \{1, 3\}, (1, 0))$ code encodes two bits (b_1, b_2) by sending b_1 along W_4^0, 1 along W_4^1, b_2 along W_4^2, and 0 along W_4^3.

We devise a *successive cancellation* decoder to decode a $(n, \mathcal{A}, \{u_i | i \in \overline{\mathcal{A}}\})$ coset code. Given an output $y_0, \ldots, y_{n-1}$ of the code, the decoder outputs a guess $\hat{u}_0, \ldots, \hat{u}_{n-1}$, a guess for the true inputs $u_0, \ldots, u_{n-1}$. Note that the frozen bits $\{u_i | i \in \overline{\mathcal{A}}\}$ are known to the decoder beforehand.

Definition 3.2. The successive cancellation decoder computes $\hat{u}_i$ for $i \in [n]$ in increasing order, according to the following rule.

- If $i \in \overline{\mathcal{A}}$, pick $\hat{u}_i = u_i$;
- If $i \in \mathcal{A}$, pick $\hat{u}_i \in \{0, 1\}$ to maximize $W_n^i(y_0, \ldots, y_{n-1}, \hat{u}_0, \ldots, \hat{u}_{i-1}|\hat{u}_i)$; in case of a tie select $\hat{u}_i = 0$.

Intuitively: if u_i is a frozen bit the decoder guesses $\hat{u}_i = u_i$; if u_i is an information bit the decoder guesses $\hat{u}_i$ by max-likelihood estimation, treating the previous guesses $\hat{u}_0, \ldots, \hat{u}_{i-1}$ as correct and the future frozen bits as random variables.

Remark 3.1. The decoder incurs no loss in accuracy for assuming the previous guesses $\hat{u}_0, \ldots, \hat{u}_{i-1}$ are correct, because if any of these guesses are wrong, decoding has already failed.

Remark 3.2. The decoder's accuracy is intentionally suboptimal; a max-likelihood decoder that takes the known values of future frozen bits into account would achieve better accuracy. However, our implementation has the advantage of being easily computable with recursive formulas. And, as we will show, the desired rate and accuracy bounds are still attainable with this decoder.

We let $P_e(n, \mathcal{A}, \{u_i | i \in \overline{\mathcal{A}}\})$ denote the average probability of error of the coset code with these parameters. We regard the n-bit coset codes with frozen positions $\mathcal{A}$ as an ensemble, with each member corresponding to a different setting of the frozen bits. Then, we can define the mean average probability of error over this ensemble

$$P_e(n, \mathcal{A}) = 2^{-|\overline{\mathcal{A}}|} \sum_{\{u_i | i \in \overline{\mathcal{A}}\}} P_e(n, \mathcal{A}, \{u_i | i \in \overline{\mathcal{A}}\}). \tag{3.1}$$

Now we can define the ensemble of polar codes.

Definition 3.3. Fix a channel W, codeword width n, and rate R. The ensemble of (n, R)-polar codes consists of the $(n, \mathcal{A}, \{u_i | i \in \overline{\mathcal{A}}\})$ coset codes, where $\mathcal{A}$ consists of the nR positions $i \in \{0, \ldots, n-1\}$ such that the Bhattacharyya parameters $Z(W_n^i)$ are minimal.

Remark 3.3. Choosing $\mathcal{A}$ to be the nR positions i with smallest Bhattacharyya parameters $Z(W_n^i)$ is an imperfect proxy for choosing it to be the nR positions with largest symmetric capacities $I(W_n^i)$. However, selecting $\mathcal{A}$ based on Bhattacharyya parameters allows us to derive an explicit bound on the code's error probability.

By slight abuse of notation, we let $P_e(n, R) = P_e(n, \mathcal{A})$, where $\mathcal{A}$ is the set of information-bit positions for the (n, R)-polar code ensemble.

We can now restate our main theorem, within this framework.

Theorem 1.2. *For any rate $R < I(W)$, the average error probability of the polar code ensemble, under successive cancellation decoding, satisfies $P_e(n, R) = O(n^{-1/4})$.*

By itself, this theorem is still nonconstructive; it only shows that for some setting of the frozen bits in the polar code ensemble, the resulting coset code has average error probability $P_e = O(n^{-1/4})$. However, it can be shown [1] that when W is a symmetric channel, any setting of the frozen bits yields the same value of P_e; thus, for symmetric W, *any* member of the polar code ensemble has average error probability $O(n^{-1/4})$. This gives, for symmetric channels, a fully explicit, deterministic construction of a code family with rate up to $I(W)$ and average error of probability tending to 0.

4. Proof of Theorem 1.2

In this section, we outline the proof of Theorem 1.2. This proof relies on two propositions that together bound the polar code's probability of error. The proofs of the two propositions are technical, and we omit them.

The first proposition is a refinement of Theorem **??**, which quantifies the rate of polarization. Fix a channel W. Let $n = 2^m$ for $m = 0, 1, 2, \ldots$, and define the channels $W_n^0, W_n^1, \ldots, W_n^{n-1}$ as in Section 3.

Proposition 4.1. [1] *For any $R < I(W)$. For $m = 0, 1, 2, \ldots$, there exists a sequence of sets $\mathcal{A}_n \subset [n]$ of size $|\mathcal{A}_n| \geq nR$, such that for all $i \in \mathcal{A}_n$, $Z(W_n^i) = O(n^{-5/4})$.*

In other words, for any $R < I(W)$, the R-quantile of the Bhattacharyya parameters of the polarized channels is $O(n^{-5/4})$.

The next proposition bounds the error probability of a coset code ensemble in terms of the Bhattacharyya parameters of the information-bit channels.

Proposition 4.2. [1] *For any $n = 2^m$ and $\mathcal{A} \subset [n]$,*

$$P_e(N, \mathcal{A}) \leq \sum_{i \in \mathcal{A}} Z(W_n^i). \tag{4.1}$$

Together these propositions prove Theorem 1.2.

Proof of Theorem 1.2. Let $\mathcal{A}$ be the information-bit positions of the (n, R)-polar code ensemble. By Proposition 4.1, there exists some c, depending only on the channel W, such that $Z(W_n^i) \leq cn^{-5/4}$ for all $i \in \mathcal{A}$. By Proposition 4.2,

$$P_e(N, R) = P_e(N, \mathcal{A}) \leq \sum_{i \in \mathcal{A}} Z(W_n^i) \leq nR \cdot cn^{-5/4} = cRn^{-1/4} = O(n^{-1/4}), \tag{4.2}$$

as desired. □

5. Conclusion

In this paper, we constructed Arikan's polar codes, which constructively prove Shannon's Channel Coding Theorem for binary symmetric channels. For binary non-symmetric channels W, this construction does not attain all the desiderata of the Channel Coding Theorem, but still achieves an impressive result: it yields a family of ensembles of codes achieving rate up to $I(W)$ (which, for these codes, is less than $C(W)$), whose mean average rate of error tends to 0.

There are two aspects of the polar coding algorithm which are very important to polar coding's success, but are out of the scope of this survey. The first is the implementation of the successive cancellation decoding algorithm. As alluded to in Section 3, this decoding algorithm was designed to be easily computed with recursive formulas; with this algorithm, the polar code can be encoded and decoded in $O(n \log n)$ time, where n is the block length. This algorithm's implementation can be found in [1]. The second is the procedure to identify the high-information channels among the polarized channels $W_n^0, \ldots, W_n^{n-1}$. Since the polar code sends information through the nR channels with lowest Bhattacharyya parameters, we must be able to identify these channels to implement polar coding. This problem turns out to be computationally hard; [1] provides a statistical algorithm to approximate this choice and an $O(n)$ time algorithm solving it if W is a binary erasure channel, but a general solution is not known.

Generalizations of this construction to non-binary channels also exist in the literature. For example, various constructions of polar codes exist, relying on the same channel-polarization phenomenon, when we replace the binary-input channel with a Gaussian channel. A comparative study of these codes can be found in [3]. In the same way that Arikan's polar codes constructively prove Shannon's Channel Coding Theorem for binary symmetric channels, these constructions prove the Gaussian Channel Coding Theorem for Gaussian channels.

References

[1] E. Arikan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, *IEEE Transactions on Information Theory* (July 2009).
[2] T. M. Cover and J. A. Thomas, Elements of Information Theory, John Wiley & Sons, 2006.
[3] H. Vangala and E. Viterbo and Y. Hong, A Comparative Study of Polar Code Constructions for the AWGN Channel, Preprint 2015, arXiv:1501.02473.