On the Construction of Polar Codes for Achieving the Capacity of Marginal Channels

Amisina Torfi*, Sobhan Soleymani*, Student Member, IEEE, and Vahid Tabataba Vakili†, Member, IEEE
*School of Computer Science and Electrical Engineering, West Virginia University
†School of Electrical Engineering, Iran University of Science and Technology

Abstract—Achieving security against adversaries with unlimited computational power is of great interest in a communication scenario. Since polar codes are capacity achieving codes with low encoding-decoding complexity and they can approach perfect secrecy rates for binary-input degraded wiretap channels in symmetric settings, they are investigated extensively in the literature recently. In this paper a polar coding scheme to achieve secrecy capacity in non-symmetric binary input channels is proposed. The proposed scheme satisfies security and reliability conditions. The wiretap channel is assumed to be stochastically degraded with respect to the main channel and message distribution is uniform. Information set is sent over channels that are good for Bob and bad for Eve. Random bits are sent over channels that are good for both Bob and Eve. A frozen vector is chosen randomly, and is sent over channels bad for both. We prove that there exists a frozen vector for which the coding scheme satisfies reliability and security conditions and approaches the secrecy capacity. We further empirically show that in the proposed scheme for non-symmetric binary-input discrete memoryless channels, the equivocation rate achieves its upper bound in the whole capacity-equivocation region.

I. INTRODUCTION

A cryptosystem is information-theoretically secure if it has no information leakage. It means that Eve with unlimited computational power could not break the system. An encryption protocol in which information-theoretic security is considered is not vulnerable to developments in computational power.

An example of an information-theoretically secure system is the one-time pad. Information-theoretically secure communication was introduced in 1949 by Shannon. He proved that one-time pad cryptosystem is secure [1].

An encryption algorithm is perfectly secure if its output cipher-text provides no information about the plain-text, when the key does not known. If $E$ is a perfectly secure encryption function, for any fixed message $M$ there must exist at least one key for each cipher-text $c$, such that $c = E_k(M)$. It has been proven that any encryption algorithm with perfect secrecy must use keys similar to one-time pad keys [1].

There are wide variety of cryptographic tasks that implement information-theoretic security as a useful and meaningful concept. Some of them are mentioned below:

- Shamir’s secret sharing algorithm is information-theoretically secure, which splits the secret into pieces and gives each piece to a specific person. To regenerate the secret at least a portion of pieces is required, otherwise reconstruction of the secret is impossible.
- Quantum cryptography is a very important part of information-theoretic cryptography.

A weaker notion of security named physical layer encryption defined by Wyner established a large area of research. It uses the physical channel for its security by signal processing and coding techniques. This notion of security is provable, unbreakable, and quantifiable.

Wyner’s initial physical layer encryption work assumes that Alice wants to send a secret message to Bob without Eve being able to decode it. It was shown that if the channel from Alice to Bob is statistically better than the channel from Alice to Eve, secure communication is possible [2]. In this setting, Alice wishes to send messages to Bob through a communication channel called the main channel, but her transmissions also reach an adversary Eve through another channel, called the wiretap channel.

Encoder maps $k$-bit message $M$ to codeword $X$ and sends it on the channel. Bob receives $Y$ on the main channel $W_m$, while on the wiretap channel $W_w$ Eve receives $Z$. Decoder maps $Y$ to $M$. A reliable and secure system should be achieved when message length tends to infinity.

Reliability: $\lim_{k \to \infty} \Pr(M \neq \hat{M}) = 0$ (1)

Security is defined as the normalized mutual information between $M$ and $Z$:

Weak Security: $\lim_{k \to \infty} \frac{I(M; Z)}{k} = 0$ (2)

Secrecy assures that observing $Z$ does not provide sufficient information about $M$. Maurer in [3], [4] proves that the conventional notion of security (Eq. 2) is a weak notion, since it is possible to construct examples where $k^{1-\varepsilon}$ out of the $k$ message bits are disclosed to Eve while still satisfying Eq. 2. Maurer defined strong security condition in [3]:

Strong Security: $\lim_{k \to \infty} I(M; Z) = 0$ (3)

Both weak and strong security conditions are information-theoretic and not computational, there is no limitation for computational power of adversary, and security does not depend on computational complexity of algorithm.

A. Prior Works

In 1975, Wyner considered the special setting where both $W_m$ and $W_w$ are discrete memoryless channels (DMCs) and $W_w$ is degraded with respect to $W_m$. He proved that such a system is characterized by a single constant $C_s$, called the secrecy capacity which has the following meaning: For $\forall \varepsilon >$
0, there exists a coding scheme of information rate $R > C_s - \varepsilon$ that satisfy (1) and (2); conversely, it is not possible to satisfy both (1) and (2) at rates greater than $C_s$ [2].

Since 1975, Wyner’s results have been extended to a variety of works like Gaussian channels [5], general broadcast channels with confidential messages [6], and channels that impose a computational constraint on the adversary [7], [8].

However, a large number of these works are based on nonconstructive random-coding to establish the main results. Such results prove the existence of a code that achieves secrecy capacity, but a few of them design specific polynomial-time encoding-decoding algorithms.

To the best of our knowledge, constructive solutions are available only in two special cases. The first special case is when the main channel is noiseless and the wiretap channel is the binary erasure channel (BEC). A coding scheme for this case, using LDPC codes for the BEC, is presented in [9] and [10] and proved to achieve secrecy capacity.

The other special case is when the adversary is constrained computationally: Eve can select to observe some $t$ out of the $n$ transmitted symbols, while the remaining $n-t$ symbols are erased. This situation is studied by Ozarow and Wyner in [8]. It is observed that even for the simple situation where $W_m$ is noiseless and $W_w$ is a binary symmetric channel, it is not known how to explicitly construct codes that achieve secrecy capacity.

### B. Assumptions and Settings

In this paper, we present a coding scheme that achieves the secrecy capacity of wiretap channels whenever $W_m$ and $W_w$ are binary-input nonsymmetric DMC and $W_w$ is degraded with respect to $W_m$. This is the situation originally studied by Wyner. Reliability and security conditions (1) and (2) could be satisfied using an encoding/decoding algorithms by polar coding. The number of operations required for encoding and decoding is $O(N \log N)$. Our construction is based upon key results in polar codes recently presented by Arkan [11].

It is proved in [11] that polar codes achieve the capacity of arbitrary binary-input channels DMC with low encoding and decoding complexity. This proof is based on channel polarization. Arikan in [11] considered the channels seen by each of the individual bits during transformation. The channels seen by individual bits are called bit-channels. It is shown in [11] that by increasing the block length the bit-channels start polarizing. They approach either a noiseless channel or a complete noisy one. The bit-channels related to noiseless channels are called good channels and the other ones are called bad channels. One of the key results of [11] is that the fraction of good channels approaches the capacity of $W$ as $N$ goes to infinity.

According to the channel polarization phenomenon, main idea of our construction is as follows: Random bits are transmitted over those bit-channels that are good for both Eve and Bob. Information bits are transmitted over those bit-channels that are good for Bob but bad for Eve, and a fixed vector is transmitted over those bit-channels that are bad for both Bob and Eve. We prove that there exist a sequence of frozen vector that our coding scheme satisfies the reliability and security condition.

### Organization

In Section II, we represent relevant concepts related to wiretap channels, in order to show an expression for the secrecy capacity in the setting where $W_m$ and $W_w$ are DMC and $W_w$ is degraded with respect to $W_m$. Also the notion of symmetric channels and secrecy capacity are defined. Section III is devoted to polar codes and theorems necessary for our proofs. We represent The proposed coding scheme in Section IV and we prove the security and reliability of proposed scheme. We also prove that the code rate approaches to secrecy capacity. In Section V, we prove that the equivocation rate for proposed coding scheme approaches to its upper bound in the whole capacity-equivocation region. In section VI the simulation results are presented. Simulation results show that equivocation rate achieves its upper bound in the whole capacity-equivocation region.

### Notations

Random variables are denoted by upper case letters, the samples by the corresponding lower case letters. Calligraphic font represents the alphabet set of related random variable. $|X|$ is the alphabet size. Notation $A^n$ is vector $A$ of length $N$. $P_X$ is the distribution of $X$. If $f(x)$ and $g(x)$ are defined on a subset of real numbers then $f(x) = O(g(x))$, if for large $x$ there exists a constant number $M$ for which the inequality $f(x) \leq M g(x)$ holds. $a_i^n$ is vector $(a_1, a_2, \ldots, a_N)$ and notation $a_A$ represents the sub vector $(a_i : i \in A)$. $C_W$ represents the capacity of the channel $W$:

$$C = \max_{P_X} I(X; Y). \tag{4}$$

$I_W$ is the symmetric capacity of the channel $W$ and for general channels it is defined as:

$$I(W) = \sum_{y \in Y} \sum_{x \in X} \frac{1}{|X|} W(y|x) \log \frac{W(y|x)}{\sum_{x' \in X} W(y|x')} \tag{5}$$

This is the maximum achievable rate when all channel inputs $x$ are used with the same probability. If the maximizing distribution $P_X$ in (5) is the uniform distribution then the symmetric capacity is equal to the capacity. $\log(\cdot)$ is based on 2 in the rest of the paper.

### C. Related works

The works by Hof and Shamai [13] and Mahdavifar and Vardy [12] on polar coding for wiretap channels assume binary-input channels for symmetric setting. In this work we assume non-symmetric binary-input channels. Work in [12] considers only achieving secrecy capacity, but we prove that the proposed scheme for non-symmetric setting achieves all capacity-equivocation region. In [12] there is no assumption on the distribution of message for proving security condition which is a fair assumption on message $M$. But we consider uniform distribution since it is the necessary condition for
II. Symmetric concept and secrecy capacity

In this section we review the works in [10] and [15] to provide a simple expression for the secrecy capacity $C_s$. Our discussion is limited to binary discrete memoryless channel with finite input-output alphabet size. Such channel is a triple $\left(\mathcal{X}, \mathcal{Y}, W(y|x)\right)$ in which $\mathcal{X}$ and $\mathcal{Y}$ are finite input-output alphabet of channel and $W$ is a $|\mathcal{X}| \times |\mathcal{Y}|$ matrix with $W(y|x)$ as entries. $W(y|x)$ is the probability of receiving $y \in \mathcal{Y}$ if $x \in \mathcal{X}$ is sent.

A matrix $W$ is symmetric if the rows of $W$ are permutations of each other and the columns of $W$ are also permutations of each other. A channel $W$ is symmetric if the rows of $W$ is symmetric, $W$ is weakly symmetric if the columns of $W$ can be partitioned into subsets such that each subset forms a symmetric matrix.

For channel $W_b$ with input alphabet $\mathcal{Y}$ the wiretapper’s channel is stochastically degraded with respect to the main channel, if it holds the following equation:

$$W_w(z|x) = \sum_{y \in \mathcal{Y}} W_m(y'|x)W_b(z|y') \forall x, z \in \mathcal{X}$$

(6)

If the channel transition probability factorizes as:

$$W(y,z|x) = W(y|x)W(z|y)$$

(7)

the wiretapper’s channel is called physically degraded with respect to the main channel. Since the capacity-equivocation region only depends on the marginal probabilities, the capacity-equivocation region for physically degraded and stochastically degraded wiretap channels is the same:

$$U_{\mathcal{X} \mathcal{Y} \mathcal{Z}|\mathcal{X}} = \begin{cases} (R_x, R_y) : \\ 0 \leq R \leq I(X; Y) \\ 0 \leq R_c \leq R \\ R_c \leq I(X; Y) - I(X; Z) \end{cases}$$

(8)

in which $R_c$ is equivocation rate and define by $\frac{1}{N}H(M|Z)$ when $N$ goes to infinity. In this case the secrecy capacity is:

$$C_s = \max_{P_x} \{I(X; Y) - I(X; Z)\}$$

(9)

If the wiretap channel is physically degraded to main channel then $X \rightarrow Y \rightarrow Z$ and $I(X; Z) \leq I(X; Y)$. In this case if the same input distribution $P_X$ maximizes both $I(X; Z)$ and $I(X; Y)$, for instance when both $W_m$ and $W_w$ are symmetric channels, the capacity-equivocation region is given by:

$$R_c \leq R \leq C_{W_m}, 0 \leq R_c \leq C_{W_m} - C_{W_w}$$

(10)

If $W$ is nonsymmetric, $C_{W_m}$ and $C_{W_w}$ are equal to $I_{W_m}$ and $I_{W_w}$. The secrecy capacity is:

$$C_s = C_{W_m} - C_{W_w}$$

(11)

for the rest of the paper, degraded means stochastically degraded.

III. Polar Coding

In this section important notions of polar coding are defined which are used in our designs and proofs.

A. Primitive Definitions

Consider a binary-input discrete memoryless channel (B-DMC) given by $W(y|x)$ where $x \in \mathcal{X} = \{0,1\}$, $y \in \mathcal{Y}$ for finite set $\mathcal{Y}$. The $N$ uses of $W$ is denoted by $W^N(y^N|x^N)$. The symmetric capacity of a B-DMC is given by:

$$I(W) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{2 \times W(y|x)}{W(y|0) + W(y|1)}$$

(12)

that is a special case of (5). The Bhattacharyya parameter of $W$ is given by:

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

(13)

which measures the reliability of $W$ since it is an upper bound on the probability of ML decision error on a single use of the channel.

Polar coding is introduced by Arikan [11]. The channel polarization phenomenon is used to construct codes that achieve the symmetric capacity $I(W)$ for B-DMC $W$ with encoding and decoding complexity that scales as $O(N \log N)$ with the block length $N$. Channel polarization consists of two operations: channel combining and channel splitting. Let $u^N_i$ be the vector that supposed to be formed. The combined channel $W_N$ is represented by:

$$W_N(y^N_1|u^N_1) = W^N(y_1^N|u_1^N)B_N F^\otimes N$$

(14)

Let $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, and let $F^\otimes n$ denote the $n$-th Kronecker power of $F$. Let $W$ be arbitrary-input DMC, and the vector $U = (U_1, U_2, \ldots, U_N)$ be a block of $N = 2^n$ bits chosen uniformly at random. Suppose $U$ is encoded as $X = UR_N F^\otimes n$, where $R_N$ is the bit-reversal permutation matrix. $X$ is transmitted through $N$ independent copies of $W$.

The channel splitting phase constructs $N$ binary-input channels from $W_N$, where the transformation is given by:

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

(15)

Polar coding utilizes the polarization effect. It transmits data through the bichannels for which $Z(W^{(i)}_N)$ is close to 0. In [20] the polar code $(N, K, A, u_{AC})$ for B-DMC $W$ is defined by $x^N = u^N_{AC} B_N F^\otimes n$ where $u_{AC}$ is a given frozen vector and the information set $A$ is chosen such that $|A| = K$ and $Z(W^{(i)}_N) < Z(W^{(j)}_N)$ for all $i \in A, j \in A^C$. The frozen vector $u_{AC}$ is given to the decoder. Successive cancellation (SC) decoder estimates the input as follows: for the frozen indices $u_{AC}$ is $u_{AC}$. For the remaining indices satisfying $i \in A, u_i = 0$, if $W_N^{(i)}(y_1^N, u_1^{i-1} | 0) \geq W_N^{(i)}(y_1^N, u_1^{i-1} | 1)$ and $u_i = 1$, otherwise.
B. Polar coding ensemble

According to polar coding, having a set of channels between encoder and decoder and sending information set on the good channels, is a coding scheme. Inputs to the other channels remain fix and are declared to the decoder. As the fraction of good channels is \( I(W) \), the rate of \( I(W) \) is achievable.

Mutual information \( I(U_1; Y_1^N, U_1^{-1}) \) corresponds to decoding \( U_1 \) with respect to the knowledge of \( U_1^{-1} \) and output \( Y_1^N \). In this case decoder should have the knowledge of \( U_1^{-1} \) for decoding \( U_1 \). However, the decoder only knows \( U_j \) where \( j \) belongs to the indices of bad channels so for the other indices decoder should use the estimate \( \hat{U}_j \), which could be incorrect. Successive cancellation decoder,decodes bits in clean order \( U_1, ..., U_N \). For the moment polar code notation is represented which is used for defining polar coding ensemble and the rest of the paper.

Definition 3.1 (Polar Coding)\(^{[17]}\): Polar code \( \mathcal{P}(N,A,u_F) \) for every \( A \subseteq \{1, ..., N\} \) and \( u_F \in \mathcal{X}^{|F|} \) is a linear code according to the following notation:

\[
\mathcal{P}(N,A,u_F) = \{ x_1^N = u_1^N G_N : u_F \in \mathcal{X}^{|F|} \} \quad (16)
\]

Code \( \mathcal{P}(N,A,u_F) \) is constructed using a fixed vector \( u_F \) with indices set \( F \) and choosing from all possible vectors in indices \( F^c \) or \( A \). Notations \( \mathcal{P}(N,A,u_F) \) and \( \mathcal{P}(N,A,u_{A^c}) \) are equivalent. \( F \) is frozen set and its indices are called frozen indices. Also \( A \) is information set and its indices are called information indices. Using code \( \mathcal{P}(N,A,u_{A^c}) \) is corresponds to sending \( U_1^N \) on channel \( W_N \) with a fixed \( u_F \) on the indices \( F \).

Definition 3.2 (Polar Coding Ensemble)\(^{[17]}\): Polar code Ensemble \( \mathcal{P}(N,A) \) for every \( A \subseteq \{1, ..., N\} \) represents the Ensemble below:

\[
\mathcal{P}(N,A) = \{ \mathcal{P}(N,A,u_F) : \forall u_F \in \mathcal{X}^{|F|} \} \quad (17)
\]

\( P_{B,e}(A,u_F) \) represents the error probability of code block \( \mathcal{P}(N,A,u_F) \) with uniform distribution assumption on all codewords. \( P_{B,e}(A) \) is the average block error probability of ensemble \( \mathcal{P}(N,A) \) which is averaging \( P_{B,e}(A,u_F) \) on all possible choices of \( u_F \in \mathcal{X}^{|F|} \) with equal probability.

Lemma 3.1 (Average Block Error Probability Upper Bound)\(^{[16]}\): For a B-DMC \( W \) and information set \( A \), average block error probability over all possible choices of frozen bits can be bounded as follows:

\[
P_{B,e}(A) \leq \sum_{i \in A} Z(W_N^{(i)}) \quad (18)
\]

C. Rate of polarization and achieving symmetric capacity

Theorem 3.1 (Rate of Convergence)\(^{[17]}\): for any B-DMC \( W \) for \( N = 2^n \) and \( \delta \in (0,1) \):

\[
\lim_{N \to \infty} \frac{\sum_{i \in \{1, ..., N\} : I(W_N^{(i)}) \in (1-\delta,1]} N = I(W) \quad (19)
\]

\[
\lim_{N \to \infty} \frac{\sum_{i \in \{1, ..., N\} : I(W_N^{(i)}) \in (0,\delta]} N = 1 - I(W) \quad (20)
\]

in order to derive the rate of channel polarization the random process \( Z_n \) defined in \([11]\) and \([17]\):\(^{[17]}\)

\[
\Pr\{Z_n \in (a,b) = \frac{|i \in \{1, ..., N\} : Z(W_2^{(i)}) \in (a,b)|}{2^n} \quad (21)
\]

Theorem 3.2 (Polarization rate)\(^{[17]}\), Theorem 1): For any B-DMC \( W \) and \( 0 < \beta < 1/2 \), \( \lim_{n \to \infty} \Pr\{Z_n < 2^{-2^n\beta}\} = I(W) \).

Theorem 3.3 (Polarization rate)\(^{[17]}\), Theorem 2): For any B-DMC \( W \) in which \( I(W) > 0 \) and \( R < I(W) \), parameter \( \beta \in (0,1/2) \) is considered to be fixed. Block error probability of polar coding averaged over all possible choices of frozen bits satisfy the following equality:

\[
P_{B,e}(A) = O(2^{-N^\beta}) \quad (22)
\]

A lemma from \([16]\) is used for realizing good channels and bad channels from each other.

Lemma 3.2 (Polar Coding Ensemble)\(^{[16]}\), Lemma 4.7): If \( W : X \to Y \) and \( W_d : X \to Y_d \) are two B-DMC W and \( W_d \) is degraded with respect to \( W \) then there exists a channel like \( W_d : Y \to Y_d \) that \( W_d(y_d|x) = \sum_{y \in Y} W(y|x)W(y_d|y) \). In this condition \( W^{(i)}_d \) is degraded with respect to \( W^{(i)}_N \) and \( Z(W^{(i)}_d) \geq Z(W^{(i)}_N) \).

This lemma implies that with assumption of degradation of wiretap channel with respect to main channel if a channel is good to Eve, it is good for Bob. Conversely if a channel is bad for Bob, it is bad for Eve too.

D. Nested polar code

We consider binary polar codes of block length \( N = 2^n \). Let \( A \) and \( B \) be two index sets such that \( B \subset A \subseteq \{1, ..., N\} \). Nested structure of polar codes comes from the cosets of a smaller subcodes. Consider the polar codes \( \mathcal{P}(N,A,u_{A^c}) \) and \( \mathcal{P}(N,B,[s,u_{A^c}]) \). Here \([s,u_{A^c}]\) is a binary vector whose elements are equal to vector \( s \) for the indices \( i \in A \cap B \) and otherwise they equal the corresponding elements in \( u_{A^c} \). \( A^c \) is a frozen set for both codes, but \( B^c \) is frozen only for \( \mathcal{P}(N,B,[s,u_{A^c}]) \).

Definition 3.3 (Nested Polar Code)\(^{[18]}\): Let \( G_N \) be the generator matrix of polar code and let \( G_N(I) \) be the submatrix composed of the rows of \( G_N \) whose indices belong to the set \( I \). The nested polar code \( \mathcal{P}(N,A,B,u_{A^c}) \) is the set of codewords \( x_N \) of the following form:

\[
x_N = u_B G_N(B) \oplus u_A \setminus B G_N(A \setminus B) \oplus u_{A^c} G_N(A^c) \quad (23)
\]

The rate of the sub-codes \( \mathcal{P}(N,B,[u_{A\setminus B},u_{A^c}]) \) equal \( \frac{|B|}{N} \), and the rate of the overall code \( \mathcal{P}(N,A,u_{A^c}) \) equals \( \frac{|A|}{N} \).

IV. CODING SCHEME FOR ACHIEVING SECRECY CAPACITY

In this section we represent a coding scheme and prove its security and reliability. Also we show that it achieves the rate of secrecy capacity.
A. Secret Codebook

A discrete memoryless wiretap channel is denoted by the following notation:

\[
(\mathcal{X}, W(y, z|x), \mathcal{Y} \times \mathcal{Z})
\]

finite sets \(\mathcal{X}, \mathcal{Y}\) and \(\mathcal{Z}\) are input alphabet, main channel and wiretap channel alphabet in the corresponding order. The channel is assumed to be memoryless and time-invariant:

\[
W(y_i, z_i|x_i^i, y_i^{-1}, z_i^{-1}) = W(y_i, z_i|x_i)
\]

Assume that the transmitter has a confidential message \(M\) which is to be transmitted to the receiver and to be hidden from the Eve. The secret codebook is defined as below:

1) secret message \(M\). The transmitted messages are assumed to be uniformly distributed over message set \(\mathcal{M}\).
2) encoding function \(\text{encod}()\) at the transmitter which maps the secret messages to the transmitted symbols: for each \(m \in \mathcal{M}\) \(\text{encod}: m \rightarrow x_1^N\)
3) Decoding function \(\text{decod}()\) at which receiver maps the received symbols to estimate of the message: \(\text{decod}(y_i^N) = \hat{m}\)

Reliability is measured by block error probability:

\[
P_e = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr(\text{decod}(y_i^N) \neq m | m \text{ is sent})
\]

Security is measured by the mutual information leakage rate to the eavesdropper:

\[
\frac{1}{N} I(M; Z_Y^N)
\]

The rate \(R\) is called achievable secrecy rate, if for any given \(\varepsilon > 0\), there exists a secret codebook such that:

\[
\text{Rate} : \frac{1}{N} \log(|\mathcal{M}|) = R
\]

\[
\text{Reliability} : P_{B,e}(A_m) \leq \varepsilon
\]

\[
\text{Security} : \frac{1}{N} \log(|\mathcal{M}|; Z_Y^N) \leq \varepsilon
\]

Wiretap channel is degraded with respect to the main channel. Relation between the input and output of the main and wiretap channels depicted as following Markov chain: \(U \rightarrow X \rightarrow (Y, Z)\)

For sufficiently large \(N\) and \(0 < \beta < 1/2\) following sets are defined:

\[
A_m = \{i \in \{1, ..., N\} : Z(W_m(i)) \leq \frac{1}{N} 2^{-N^\beta}\}
\]

\[
A_w = \{i \in \{1, ..., N\} : Z(W_w(i)) \leq \frac{1}{N} 2^{-N^\beta}\}
\]

According to the defined sets, \(A_m\) is the good channel indices for the main channel and \(A_w\) is the good channel indices for the wiretap channel. According to polar coding definition and lemma 3.2 it is concluded that \(\mathcal{F}_m \subseteq \mathcal{F}_w, A_w \subseteq A_m\). We consider nested polar code \(\mathcal{P}(N, A_m, A_w, u_{F_m})\) defined in Section IV.D. The main objective is to form \(u_1^N\) based on defined indices sets. \(u_1^N\) is the vector that multiplied by generator matrix and forms transmitted codeword. Figure.1 demonstrates the relation between defined indices.

B. Encoding Algorithm

Secret message is mapped on the vector \(V_m\) and random vector \(V_e\) is generated with uniformly random distribution. The vector \(u_1^N\) is formed as below:

1) Information bits are sent over the indices which are good for the main channel and bad for the wiretap channel. This concept could be shown by \(u_{A_m \setminus A_w} = u_{F_w \setminus F_m} = V_m\) message length is \(k\) and \(|V_m| = |A_m| - |A_w| = k\).

Message distribution is uniform and therefore \(\log |M| = k\).

2) We send random bits over indices that belong to good channel for both main and wiretap channels as \(u_{A_w} = V_e\).

3) Over bad channels for both \(u_{F_m}\), we send a frozen vector assumed to be chosen from all possible choices \(u_{F_m} \in \chi^{[F_m]}\) and given to decoder of Bob and Eve. Coding scheme is built over \(\forall u_{F_m} \in \chi^{[F_m]}\) and we should prove that there exists a specific frozen vector like \(u_{F_m}\) for which our coding scheme satisfies the reliability and security condition and additionally achieves the secrecy capacity.

C. Decoding

Decoding should satisfy reliability and coding rate.

1) Reliability: \(V_m\) and \(V_e\) are defined over good indices of main channel, that according to theorem 3.3, both could be decoded using SC decoding with probability of error \(P_{B,e}(A_m) = O(2^{-N^\beta})\) (averaging over all possible choices of \(u_{F_m}\)). Therefore, the reliability condition is satisfied.

2) Rate: Message distribution is uniformly at random. That leads to \(\log |M| = |V_m| = k\), consequently for sufficiently large \(N\):

\[
R = \frac{|V_m|}{N} = \frac{|A_m| - |A_w|}{N} = I(W_m) - I(W_w) = C_s
\]

Thus, the coding scheme achieves the secrecy capacity.

D. Security Proof

Since the scheme is formed over all possible choices of frozen bits (polar coding ensemble), mutual information between message and Eve evaluated using one random chosen vector \(u_{F_m}\) over the whole set \(u_{F_m} \in \chi^{[F_m]}\). After choosing \(u_{F_m}\), we fix it and ultimately we prove that there exists such \(u_{F_m}\). The decoding error probability of Eve has been evaluated over the ensemble in average sense.

\[
I(M; Z_1^N | U_{F_m}) = I(V_m; Z_1^N | U_{F_m})
\]
\[ I(V_m, V_r; Z_1^N | U_{F_m}) = I(V_r; Z_1^N | V_m, U_{F_m}) \]  
(35)

\[ = I(U_1^N; Z_1^N) - I(V_r; Z_1^N | V_m, U_{F_m}) \]  
(36)

\[ = I(U_1^N; Z_1^N) - H(V_r) + H(V_r | V_m, U_{F_m}, Z_1^N) \]  
(37)

\[ \leq I(X_1^N; Z_1^N) - H(V_r) + H(V_r | V_m, U_{F_m}, Z_1^N) \]  
(38)

\[ \leq N I(W_u) - |A_w| + H(V_r | V_m, U_{F_m}, Z_1^N) \]  
(39)

Equation (35) follows from the chain rule of mutual information and (36) is a consequence of the following:

\[ I(U_1^N; Z_1^N) = I(U_1^N; V_m, U_{F_m}, Z_1^N) \]
\[ = I(U_1^N; Z_1^N) + I(V_m, V_r; Z_1^N | U_{F_m}) \]
\[ = I(V_m, V_r; Z_1^N | U_{F_m}) \]  
(40)

The last equality in (40) is derived from \( I(U_{F_m}; Z_1^N) \) being equal to zero, since \( u_{F_m} \) is sent over bad channels for both main and wiretap channels. Equation (37) follows from the independence of \( V_r, V_m, U_{F_m} \) and the error probability in average sense is

\[ H(E) = \frac{1}{N} \sum_{i=0}^{N-1} P(E = i) H(V_r | E = i, V_m, U_{F_m}, Z_1^N) \]
\[ = P(E = 1) H(V_r | E = 1, V_m, U_{F_m}, Z_1^N) \]
\[ + (1 - P(E = 1)) \times 0 \]
\[ \rightarrow H(V_r | V_m, U_{F_m}, Z_1^N) \]
\[ = P(E = 1) H(V_r | E = 1, V_m, U_{F_m}, Z_1^N) \]
\[ \leq P(E = 1) H(V_r | E) = P_e |A_w| \]

(48)

(49)

(50)

(51)

(52)

(53)

(54)

(55)

(56)

(57)

Therefore, the security of coding scheme is proved, since \( I(M; Z) / N \) is equivalent to \( I(M; Z) / k \). Equation (54) holds since \( I(W_u) \rightarrow \infty \) as \( k \rightarrow \infty \).

V. EXTENDING PROOFS TO THE WHOLE CAPACITY-EQUIVOCATION REGION

In this section we show that for the proposed scheme the equivocation rate approaches its upper bound for rates higher than secrecy capacity \( C_s \). According to (10), for rates higher than \( C_s \), and for binary-input non-symmetric channels the upper bound for equivocation rate equals \( C_s = I_{W_m} - I_{W_u} \). We expand \( I(X_1^N, M; Z_1^N) \) in two forms:

\[ I(X_1^N, M; Z_1^N | U_{F_m}) = I(X_1^N; Z_1^N | U_{F_m}) + I(M; Z_1^N | X_1^N, U_{F_m}) \]

(55)

\[ = I(M; Z_1^N | U_{F_m}) + I(X_1^N; Z_1^N | M, U_{F_m}) \]  
(56)

In (55) \( I(M; Z_1^N | X_1^N, U_{F_m}) \) equals to zero \((M \rightarrow X \rightarrow Y \rightarrow Z)\). From (55) and (56):

\[ I(M; Z_1^N | U_{F_m}) = I(X_1^N; Z_1^N | U_{F_m}) \]
\[ - I(X_1^N; Z_1^N | M, U_{F_m}) \]

(57)
Equivocation rate $H(M|Z^N_1, U_{\mathcal{F}_m})/N$ is expanded as following equations:

$$
\frac{H(M|Z^N_1, U_{\mathcal{F}_m})}{N} = \frac{H(M|U_{\mathcal{F}_m}) - I(M;Z^N_1|U_{\mathcal{F}_m})}{N} = (58)
$$

$$
\frac{H(M|U_{\mathcal{F}_m}) + I(X^N;Z^N_1|M,U_{\mathcal{F}_m}) - I(X^N;Z^N_1|U_{\mathcal{F}_m})}{N} = (59)
$$

(59) concluded using (58) and (57). From (59):

$$
\frac{H(M|U_{\mathcal{F}_m})/N + H(X^N_1|U_{\mathcal{F}_m})/N - H(X^N_1|Z^N_1,M,U_{\mathcal{F}_m})/N - I(X^N;Z^N_1|U_{\mathcal{F}_m})/N}{N} = (60)
$$

$$
\frac{H(M,X^N_1|U_{\mathcal{F}_m})/N - H(X^N_1|Z^N_1,M,U_{\mathcal{F}_m})/N - I(X^N;Z^N_1|U_{\mathcal{F}_m})/N}{N} = (61)
$$

$$
\frac{H(X^N_1|U_{\mathcal{F}_m})/N - H(X^N_1|Z^N_1,M,U_{\mathcal{F}_m})/N - I(X^N;Z^N_1|U_{\mathcal{F}_m})/N}{N} = (62)
$$

Equation (62) is derived from the Markov chain $M \rightarrow U \rightarrow X \rightarrow (Y,Z)$. (63) follows from $I(X^N_1;Z^N_1|U_{\mathcal{F}_m})/N \leq I_{W_w}$ and $H(X^N_1|U_{\mathcal{F}_m})/N = |A_m|/N$. Inequality $I(X^N_1;Z^N_1|U_{\mathcal{F}_m})/N \leq I_{W_w}$ holds:

$$
I(X^N_1;Z^N_1|U_{\mathcal{F}_m}) = H(Z^N_1|U_{\mathcal{F}_m}) - H(Z^N_1|X^N_1,U_{\mathcal{F}_m})
$$

$$
= H(Z^N_1|U_{\mathcal{F}_m}) - \sum \limits_{i=1}^N H(Z_i;X_i,U_{\mathcal{F}_m})
$$

$$
\leq \sum \limits_{i=1}^N (H(Z_i|U_{\mathcal{F}_m}) - H(Z_i|X_i,U_{\mathcal{F}_m}))
$$

$$
= \sum \limits_{i=1}^N I(X_i;Z_i|U_{\mathcal{F}_m}) \leq NI_{W_w}
$$

Equation $H(X^N_1|Z^N_1,M,U_{\mathcal{F}_m}) = H(V_1|Z^N_1,M,U_{\mathcal{F}_m})$ and lemma 4.1 prove that there exists a sequence of frozen bits for which the following inequality holds:

$$
H(X^N_1|Z^N_1,M,U_{\mathcal{F}_m}) \leq N\varepsilon
$$

Therefore,

$$
H(M|Z^N_1,U_{\mathcal{F}_m})/N \geq |A_m|/N - \varepsilon - I_{W_w}
$$

$$
\rightarrow \frac{H(M|Z^N_1,U_{\mathcal{F}_m})}{N} \geq \frac{|A_m|}{N} - \varepsilon - I_{W_w}
$$

$$
\rightarrow \frac{H(M|Z^N_1)}{N} \geq \frac{|A_m|}{N} - \varepsilon - I_{W_w}
$$

And for sufficiently large $N$:

$$
\lim \limits_{N \rightarrow \infty} \frac{H(M|Z^N_1)}{N} = I_{W_w} - I_{W_w}
$$

$\textbf{VI. SIMULATION RESULTS}$

In this section results for calculation of equivocation at Eve are presented to support the theoretic proofs. We show that the equivocation rate achieves its upper bound for all rates. First, equivocation at Eve $H(M|Z^N)$ has been introduced. Then, results has been presented. In all the settings, message has uniform distribution and both main and wiretap channels are BEC. BEC channels are of great interest since there are recursive equations to calculate the Bhattacharrya parameter [11].

$\textbf{A. Equivocation at Eve}$

The measure for security is $I(M;Z^N)/N$. It is expanded to:

$$
\frac{I(M;Z^N)}{N} = \frac{H(M)}{N} - \frac{H(M|Z^N)}{N}
$$

$$
= \frac{k}{N} - \frac{H(M|Z^N)}{N} = R - \frac{H(M|Z^N)}{N}
$$

To derive $I(M;Z^N)/N$, calculating $H(M|Z^N)$ is sufficient. We propose the following lemma which is an extension to [18] Lemma 4.1.

$\textbf{Lemma 6.1:}$ Assume that the nested polar code is considered $P(N,A,B,\mathcal{F}_T)$ ($u_\mathcal{F}_T$ is a randomly chosen vector and fixed after selecting), and $B \subseteq A$. $H_T$ is the parity check matrix for overall code ($P(N,A)$ in polar coding) and $H_S$ is the parity check matrix for sub-code $P(N,B)$, and channel is BEC. Then, equivocation at Eve is calculated as following:

$$
H(M|Z^N) = \text{Rank}(\tilde{H}_S(\varepsilon)) - \text{Rank}(\tilde{H}_T(\varepsilon))
$$

$\tilde{H}(\varepsilon)$ is the matrix formed by columns of $H$ that belong to erased positions.

$\textbf{Proof}$

$$
I(M;X^N|Z^N) = H(M|Z^N) - H(M|X^N,Z^N)
$$

$$
= H(X^N|Z^N) - H(X^N|M,Z^N)
$$

$$
\rightarrow H(M|Z^N) = H(X^N|Z^N) - H(X^N|M,Z^N)
$$

$$
H(M|X^N,Z^N) \text{ is equal to since knowing transmitted codeword results in the message to be realized. Channels are BEC. Therefore, for the received } Z, X \text{ is explicit with some erased symbols. Transmitted vector is built as } x^N = u_AG_N(A) \oplus u_\mathcal{F}_NG_N(\mathcal{F}). G_T = G_N(A\cup\mathcal{F}_{non-zero}) \text{ is the generator matrix of polar code formed from the rows of the mother generator matrix } G_N \text{ that belongs to information indices and nonzero positions of frozen vector. } H(X^N|Z^N) \text{ corresponds to overall code } P(N,A), \text{ consequently for a received } Z \text{ parity check equation hold: } \tilde{H}_T (x^N_T) = 0. \text{ Therefore:}
$$

$$
\tilde{H}_T(\varepsilon)x^N_T + \tilde{H}_T(\varepsilon)x^N_T = 0
$$

Equation (75) holds for unknown $x^N_T$. It has $2^{\text{rank}(\tilde{H}_T(\varepsilon))}$ solutions with equal probabilities, since codewords are equal likely. $|\varepsilon|$ is the number of erased position. Consequently $H(X|Z^N) = |\varepsilon| - \text{Rank}(\tilde{H}_T(\varepsilon))$.

It can also be proven that $H(X|M,Z^N) = |\varepsilon| - \text{Rank}(\tilde{H}_S(\varepsilon))$. Therefore, it concludes (73).
Then, the parity check matrix could be calculated as

$$G_T = G_N (A \cup \mathcal{F}_{\text{non-zero}}) = [I_{|A|} | \mathcal{F}_{\text{non-zero}}] P_1.$$  

Then, the parity check matrix could be calculated as $H_T = [P_1^{-1} I_{N-|A|} | \mathcal{F}_{\text{non-zero}}]$. 

### Table I | EQUIVOCATION RATE

| Rate   | $R_e = H(M; Z^N)/N$ | $I(M; Z^N)/N$ |
|--------|---------------------|---------------|
| 0.05   | 0.012               |               |
| 0.1    | 0.0053              |               |
| 0.15   | 0.025               |               |
| 0.2    | 0.0008              |               |
| 0.25   | 0.0075              |               |
| 0.3    | 0.052               |               |
| 0.4    | 0.151               |               |
| 0.5    | 0.251               |               |
| 0.6    | 0.350               |               |

### VII. CONCLUSION AND DISCUSSION

In this paper we considered binary-input non-symmetric wiretap channels. We proved that there exists a frozen vector for which coding scheme satisfies reliability and security conditions and also code rate achieves secrecy capacity. We proved that the equivocation rate achieves its upper bound for all rates in non-symmetric channels. Our results extend to discrete memoryless channels with non-binary input. It is proved in [19] that channels with an input alphabet of prime size are polarized by the same transformation. If the alphabet is not of prime size, then splitting the input alphabet into prime subsets can solve the problem.

All the constructions in this paper are as explicit as the polar codes, since only existence of a suitable frozen vector is proved and the method to choose it is not explored. On the other hand, it is proved by Maurer-Wolf [4] that coding schemes that satisfy the weak security condition can be converted to schemes that satisfy the strong security. This is accomplished using information reconciliation and privacy amplification protocol [21]. Therefore, the proposed scheme could be extended to strong security using privacy amplification protocol.

Another possible further problem to explore is to construct codes when the wiretap channel is not degraded, since degradation of wiretap channel is a sufficient but not necessary condition. Also, the proposed algorithm benefits from successive cancellation decoding, which depends on the past estimates. Therefore, if the estimates are incorrect, the error will propagate. A decoder can be implemented to overcome this set-back. In addition, it worth investigating the possibility of the other decoding methods, such as belief propagation [23] and recursive-list decoding [22] being able to eliminate this deficiency.

Recently, [20] has suggested an algorithm to overcome the polarization limitation. The method defines pseudo-random frozen bits and extent to all channels. This method can be utilized to generalize the proposed scheme for arbitrary discrete memoryless channels.

### REFERENCES

[1] C. E. Shannon, “Communication Theory of Secrecy Systems,” Bell System Technical Journal (USA: AT&T Corporation), vol. 28, pp. 656-715, Oct. 1949.
[2] A. D. Wyner, “The wire-tap channel,” Bell Syst. Tech. J., vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
[3] U. M. Maurer, “The strong secret key rate of discrete random triples,” in Communication and Cryptography: Two Sides of One Tapestry, et al., Boston, MA, 1994, pp. 271–285.
[4] U. M. Maurer and S. Wolf, “Information-theoretic key agreement: From weak to strong secrecy for free,” in Lecture Notes in Computer Science, Berlin, Germany: Springer, 2000, vol. 1807, pp. 351–368.
[5] S. L. Y. Cheong and M. E. Hellman, “The Gaussian wiretap channel,” IEEE Trans. Inf. Theory, vol. IT-24, no. 4, pp. 451–456, Jul. 1978.
[6] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” IEEE Trans. Inf. Theory, vol. IT-24, no. 3, pp. 339–348, May 1978.
[7] M. Cheraghchi, F. Didier, and A. Shokrollahi, “Invertible extractors and wiretap protocols,” in Proc. IEEE Int. Symp. Information Theory, Seoul, Korea, 2009, pp. 1934–1938.
[8] L. H. Ozarow and A. D. Wyner, “Wiretap channel II,” Bell Lab. Tech. J., vol. 65, no. 10, pp. 2132–2157, Dec. 1984.
[9] A. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin, “Strong secrecy for erasure wiretap channels,” in Proc. IEEE Information Theory Workshop, Dublin, Ireland, Sep. 2010.
[10] A. Thangaraj, S. Dhiidar, A. R. Calderbank, S. W. McLaughlin, and J. Merolla, “Applications of LDPC codes to the wiretap channel,” IEEE Trans. Inf. Theory, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.
[11] E. Arikan, “Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. on Information Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.
[12] H. Mahdavifar, A. Vardy, “Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes,” IEEE Trans. Inf. Theory, Vol. 57, No. 15, pp. 6428–6443, October 2011.
[13] E. Hof and S. Shamai, “Secrecy-achieving polar-coding,” in Proc. IEEE Information Theory Workshop, Dublin, Ireland, pp. 15 Aug.-Sep. 2010.
[14] A. Torfi, S. Soleymani, S. M. Iranmanshe, H. Kazemi, R. A. Shirvani, and V. T. Vaithi, “Polar coding for achieving the capacity of marginal channels in nonbinary-input setting,” In 51st Annual Conference on Information Sciences and Systems (CISS), 2017, IEEE, 2017.
[15] S. Leung-Yan-Cheong, “On a special class of wire-tap channels,” IEEE Trans. Inf. Theory, vol. IT-23, no. 5, pp. 625–627, Sep. 1977.
[16] S. B. Korada, “Polar Codes for Channel and Source Coding,” Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2009.
[17] E. Arikan and E. Telatar, “On the rate of channel polarization,” in Proc IEEE International Symposium on Information Theory (ISIT 2009), Seoul, Korea, Jun. 2009.
[18] M. Anderson, V. Rathi, R. Thobaben, J. Kliwer, and M. Skoglund, “Nested Polar Codes for Wiretap and Relay Channels,” IEEE Communications Letters, Vol. 14, No. 8, pp. 752–754, August 2010.
[19] E. Sasoglu, E. Telatar, and E. Arikan, “Polarization for arbitrary discrete memoryless channels,” in Proc. IEEE Information Theory Workshop, Taormina, Italy, pp. 144–148, Oct. 2009.
[20] J. Honda and H. Yamamoto, “Polar Coding Without Alphabet Extension for Asymmetric Models,” IEEE Transactions on Information Theory (IEEE), Vol. 59, No. 12, pp. 7829–7838, December 2013.
[21] C. Bennett, G. Brassard, C. Cérémon, and U. M. Maurer, “General- ized privacy amplification,” IEEE Trans. Inf. Theory, vol. 41, no. 6, pp. 1915–1923, Nov. 1995.
[22] I. Damir and K. Shabunov, “Soft-decision decoding of Reed-Muller codes: Recursive lists,” IEEE Trans. Inf. Theory, vol. 52, no. 3, pp. 1260–1266, Mar. 2006.
[23] N. Hussami, R. L. Urbanke, and S. B. Korada, “Performance of polar codes for channel and source coding,” in Proc. IEEE Int. Symp. Information Theory, Seoul, Korea, Jun. 2009, pp. 1488–1492.