TECHNICAL UNIVERSITY OF CRETE
Department of Electronic and Computer Engineering

Coding for the Wiretap Channel

Submitted to the Department of Electronic and Computer Engineering in partial fulfillment of the requirements for the ECE Diploma Degree.

By: Antzela Kosta

Advisor: Associate Professor George Karystinos

Committee Member: Associate Professor Aggelos Bletsas

Committee Member: Professor Athanasios Liavas

Chania, January 2015

Abstract

In this work we consider the problem of achieving information-theoretic security with practical coding complexity. Specifically, we consider a transmitter, a receiver, and an eavesdropper where the communication channels between the transmitter and the receiver and between the transmitter and the eavesdropper are both binary erasure channels. The objective is the development of a practical coding scheme that attains maximum transmission rate between the transmitter and the receiver and, simultaneously, complete protection of the transmitted data from the eavesdropper.

We utilize polar codes which are known to achieve the capacity of any symmetric binary-input discrete memoryless channel. We focus on the binary erasure channel and present the pertinent polar encoding and decoding algorithms. Finally, we show how polar coding can be used for the wire-tap binary erasure channel, achieving maximum transmission rate and complete protection.

Acknowledgements

I would like to express my special appreciation and thanks to my advisor Professor George Karystinos. Without his assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished.

I would also like to thank my family for their endless love, support, encouragement and for giving me every single day the chance to follow my dreams.

Finally, I would like to thank my friends for their support and all the great moments we had together.

Contents

List of Figures

1 Introduction

Rapid advances in wireless technology are quickly taken as towards a pervasively connected word in which a vast array of wireless devices, from iPhones to biosensors, seamlessly communicate with each other. Fostered by the rapid proliferation of wireless communication devices, technologies, and applications, the need for reliable and secure data communication over wireless networks is more important than ever before. Due to its broadcast nature, wireless communication is particularly susceptible to eavesdropping.

Physical-layer security techniques have a rather long history. The notion of information theoretic security was introduced by Shannon [1], and later extended by Wyner to noisy channels [2]. Shannon provided the first truly scientific treatment of secrecy in [1], in which a secret key is considered to protect confidential messages. Wyner proposed an alternative approach to secure communication schemes in his paper [2], where he introduced the so-called wiretap channel model.

In this setting, Alice wishes to send messages to Bob through a communication channel C_1, called the main channel, but her transmissions also reach an adversary Eve through another channel C_2, called the wiretap channel. This is illustrated in Fig. 1, wherein $\mathbf{U}^k$ denotes a k-bit message that Alice wishes to communicate to Bob. We think of $\mathbf{U}^k$ as a data sequence which consists of independent copies of the binary random variable U, where $Pr\{U = 0\} = Pr\{U = 1\} = \frac{1}{2}$. The encoder maps $\mathbf{U}^k$ into a sequence $\mathbf{X}^n$ of n channel symbols. This sequence is transmitted across the main channel and the wiretap channel resulting in the corresponding channel outputs $\mathbf{Y}^n$ and $\mathbf{Z}^n$. Finally the decoder maps $\mathbf{Y}^n$ into an estimate $\hat{\mathbf{U}}^k$ of the original message.

The goal is to design a coding scheme − namely, an encoding algorithm and an decoding algorithm − that makes it possible to communicate both *reliably* and *securely*, as the message length k becomes large. Reliability is measured in terms of the *probability of error* in recovering the message. Specifically, the objective is to satisfy the following:

$$\textbf{Reliability Condition: } \lim_{k\to\infty} \Pr\{\hat{\mathbf{U}}^k \neq \mathbf{U}^k\} = 0. \tag{1}$$

where the probability is over all the relevant coin tosses in the system: in the generation of $\mathbf{U}^k$, in the encoder, and in the main channel. Security is usually measured in terms of the normalized mutual

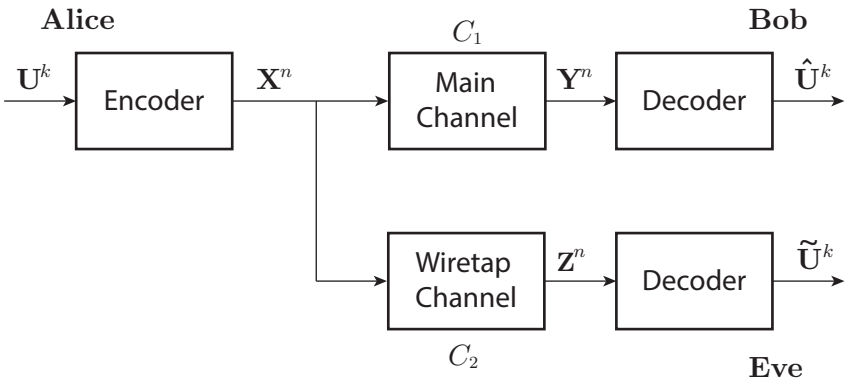

Figure 1: Wiretap channel model.

information between the message $\mathbf{U}^k$ and Eve's observations $\mathbf{Z}^n$. Specifically, we are interested in encoding algorithms that satisfy the following:

$$\textbf{Security Condition: } \lim_{k\to\infty} \frac{I(\mathbf{U}^k;\mathbf{Z}^n)}{k} = 0. \tag{2}$$

Note that $I(\mathbf{U}^k;\mathbf{Z}^n)$ is equal to the difference between the *a priori* entropy $H(\mathbf{U}^k)$ and the conditional entropy $H(\mathbf{U}^k \mid \mathbf{Z}^n)$. Intuitively, (2) means that observing $\mathbf{Z}^n$ does not provide any information about $\mathbf{U}^k$ beyond what is available *a priori*, as compared to the message length k.

Wyner demonstrated that secure communication is possible without sharing a secret key and showed that, although a wire-tapper may know the encoding scheme used at the transmitter and the decoding scheme used by the legitimate receiver, he can be kept ignorant solely by the greater noise present in his received signal. He measured confidentiality by equivocation (i.e. the level of ignorance of the eavesdropper with respect to the confidential message) and determined the rate-equivocation region. He considered a special case of the system in Fig. 1 where both C_1 and C_2 are discrete memoryless channels (DMCs) and, moreover, C_2 is degraded with respect to C_1 and determined the secrecy capacity for a wiretap channel, which has the following meaning. For all $\epsilon > 0$, there exist coding schemes of information rate $R \geq C_s - \epsilon$ that satisfy (1) and (2); conversely, it is not possible to satisfy both (1) and (2) at rates greater than C_s.

Construction of explicit and practical secure encoders and decoders whose performance is as good as promised by Wyner is still an unsolved problem in the general case. More recently, low-density parity-check (LDPC) based coding design has been studied for binary erasure wiretap channels in [3]

and [4] and type II wiretap channels in [5]. In [7] and [8], it is shown that polar codes, introduced by Arikan [6], can be constructed to achieve the secrecy capacity for the binary-input memoryless degraded wiretap channel when both the main and eavesdropper channels are arbitrary symmetric channels and the marginal channel to the eavesdropper is physically degraded with respect to the marginal channel to the legitimate user. Similar works using polar codes for the wiretap channel are provided in [9], [10].

In this work, we focus on secure coding for a class of wiretap channels, in which the main channel and the wiretap channel are binary erasure channels (BECs). We use the polar coding technique of Arikan [6] to achieve the secrecy capacity over this channel with $\mathcal{O}(n \log n)$ encoding and decoding complexity, where n is the code length.

2 Secrecy Capacity

2.1 The Wiretap Communication Model

We consider the conventional wiretap channel illustrated in Fig. 1, where the transmitter sends a confidential message to a legitimate receiver via the main channel in the presence of an eavesdropper, who listens to the message through its own channel.

The message $\mathbf{U}^k$ is chosen uniformly from a set of size k. Next, the message is encoded to a codeword $\mathbf{X}^n$ with a blocklength n over an alphabet $\mathcal{X}$. Our study specializes of the binary alphabet ($\mathcal{X} \doteq \{0, 1\}$). The ratio $\frac{k}{n}$ is the rate of the code.

The main channel is a discrete memoryless channel with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition probability $P(y \mid x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Since the channel is memoryless, the transition probability for a sequence of transmitted symbols $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ and received symbols $\mathbf{y} = [y_1, y_2, \cdots, y_n]$ is

$$P(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{n} P(y_i \mid x_i). \tag{3}$$

The wiretap channel is also a discrete memoryless channel with input alphabet $\mathcal{Y}$, finite output alphabet $\mathcal{Z}$ and and transition probability $P(z \mid x)$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$.

Given a channel $C_1 = W_M(y \mid x)$, we say that another channel $C_2 = W_{MW}(z \mid x)$ is *degraded with respect to* C_1 if there exist a third channel $C_3 = W_W(z \mid y)$ such that C_2 is the cascade of C_1 and C_3. The cascade of the main channel and the wiretap channel is another memoryless channel with transition probability

$$W_{MW}(z \mid x) = \sum_{y \in \mathcal{Y}} W_W(z \mid y) W_M(y \mid x). \tag{4}$$

The decoder is a mapping

$$f_D : \mathbf{Y}^n \to \mathbf{U}^k. \tag{5}$$

2.2 Binary-input symmetric-output memoryless channels

A DMC with a transition probability p, binary input alphabet $\mathcal{X}$, and output alphabet $\mathcal{Y}$ is said to be *symmetric* if there exists a permutation π over $\mathcal{Y}$ such that

1)The inverse permutation π^{-1} is equal to π, i.e.,

$$\pi^{-1}(y) = \pi(y)$$

for all $y \in \mathcal{Y}$.

2)The transition probability p satisfies

$$p(y \mid 0) = p(\pi(y) \mid 1)$$

for all $y \in \mathcal{Y}$.

The capacity of a symmetric channel is given by

$$C(W) = H(X) - H(X|Y) = \log_2 |X| - H(X|Y). \tag{6}$$

where the random variable X at the input of the channel is uniform over $\mathcal{X}$, and Y is the corresponding random variable at the channel output.

Gallager proved that the channel capacity of a symmetric channel can be achieved using equiprobable inputs. Therefore, the capacity-achieving input distribution of a binary-input output-symmetric channel is uniform.

Two examples of binary-input symmetric-output memoryless channels are the binary symmetric channel (BSC) and the binary erasure channel (BEC).

2.2.1 Binary Symmetric Channel (BSC)

The binary symmetric channel, shown in Fig. 2, transmits one of two symbols, the binary digits $X \in \{0, 1\}$, and returns one of two symbols $Y \in \{0, 1\}$. The channel flips a transmitted bit with probability ϵ and with probability $1 - \epsilon$ the symbol Y is the symbol that was sent. The parameter ϵ is called the crossover probability of the channel.

As stated previously, a binary input channel is *symmetric* if both input symbols are corrupted equally by the channel. The BSC channel is *symmetric* since $P(Y = 0|X = 1) = P(Y = 1|X = 0)$ and $P(Y = 0|X = 0) = P(Y = 1|X = 1)$.

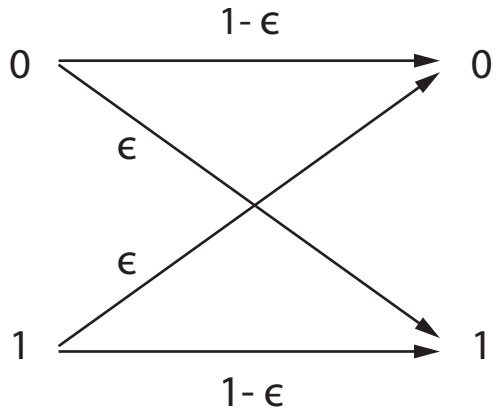

Figure 2: BSC(ϵ).

It can be proved that the capacity of BSC is:

$$C(\epsilon) = 1 - h_2(\epsilon), \tag{7}$$

where $h_2(\epsilon) = -\epsilon \log_2(\epsilon) - (1-\epsilon)\log_2(1-\epsilon)$ is the binary entropy function.

Note that, whenever $\epsilon_2 \geq \epsilon_1$, the channel $C_2 = \text{BSC}(\epsilon_2)$ is degraded with respect to $C_1 = \text{BSC}(\epsilon_1)$.

2.2.2 Binary Erasure Channel (BEC)

The binary erasure channel, shown in Fig. 3, transmits one of two symbols, usually the binary digits $X \in \{0, 1\}$. Once a bit is transmitted, the receiver will obtain either the bit correctly or a symbol that does not contain any information. In other words, BEC does not introduce incorrect information. We observe that the BEC is *symmetric* and erases a bit with erasure probability ϵ.

It can be proved that the capacity of BEC is:

$$C(\epsilon) = 1 - \epsilon. \tag{8}$$

This is a convenient form and linear to the channel's only parameter.

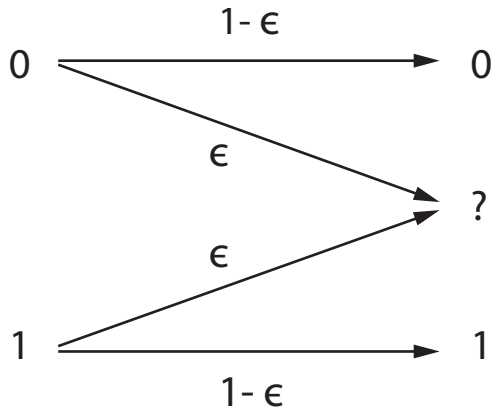

Figure 3: BEC(ϵ).

2.3 The Secrecy Capacity of the Wiretap Channel

The secrecy capacity of the wiretap-channel system in Fig. 1 is defined as follows. First, we assume that the message is uniformly random over $\{0,1\}^k$. Then C_s is the supremum over all rates $R = \frac{k}{n}$ (in bits per channel use) such that there exist coding schemes of rate R satisfying conditions (1) and (2). For the general case where C_1 and C_2 are arbitrary DMCs, computing the secrecy capacity is a difficult problem. Let X denote the single-letter input to C_1 and C_2, and Y and Z denote the corresponding single-letter outputs. The best known expression for the secrecy capacity C_s, given by Csiszár and Körner in [11], is

$$C_s = \max_U \big(I(U;Y) - I(U;Z)\big), \tag{9}$$

where the maximum is taken over all random variables U such that $U \to X \to (Y,Z)$ is a Markov chain. The problem is that this maximization is often difficult to evaluate and there is no simpler expression for the secrecy capacity even when C_1 and C_2 are both strongly symmetric, unless additional constraints are satisfied.

However, when C_1 and C_2 are symmetric and C_2 is degraded with respect to C_1, a simple expression for C_s was given by Leung-Yan-Cheong in [12]. It is shown that in this case

$$C_s = C_1 - C_2 = H(X \mid Z) - H(X \mid Y) \tag{10}$$

where X is uniform over $\mathcal{X}$. In particular, if the main channel is BSC(ϵ_1) while the wiretap channel is BSC(ϵ_2), with $\epsilon_2 \geq \epsilon_1$, then the secrecy capacity is given by $h_2(\epsilon_2) - h_2(\epsilon_1)$, where $h_2(\cdot)$ is the binary

entropy function. Similarly, if the main channel is BEC(ϵ_1) while the wiretap channel is BEC(ϵ_2), with $\epsilon_2 \geq \epsilon_1$, then the secrecy capacity is given by $\epsilon_2 - \epsilon_1$.

3 Polar Codes

3.1 Symmetric Capacity

We consider a discrete memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y}$. Then, channel polarization is a strategy to achieve the symmetric capacity of the channel, defined as

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

where W denotes a B-DMC. The symmetric capacity of the channel is equal to the Shannon capacity (i.e., the maximum achievable rate subject to arbitrarily low bit detection error probability) if we enforce the input distribution to be uniform ($\mathrm{P(X = 0) = P(X = 1) = \frac{1}{2}}$ where X is the random variable modeling channel input). Therefore, if the channel is symmetric, then polar codes achieve Shannon capacity.

3.2 Bhattacharyya Parameter

In addition to symmetric capacity which is a measure of the maximum possible reliable transmission rate through channel with uniform input, we also introduce the Bhattacharyya parameter of a B-DMC W which is defined as

$$Z(W) \doteq \sum_{y\in\mathcal{Y}} \sqrt{W(y|0)W(y|1)}, \tag{12}$$

measures the reliability of the channel, equals the sum over all output combinations of those geometric means, and constitutes an upper bound on the error probability of an uncoded bit transmission.

Symmetric capacity and Bhattacharyya parameter are related by the formulas

$$I(W) \geq \log_2 \frac{2}{1 + Z(W)}, \tag{13}$$

$$I(W) \leq \sqrt{1 - Z(W)^2}. \tag{14}$$

Proposition 1: The first corollary that can be driven by (13) and (14) is that for any W with $I(W) = 1$ or 0 we have that $Z(W) = 0$ or 1, respectively, and vice versa.

3.3 Channel Polarization

Given N independent copies of one B-DMC W, the idea behind channel polarization is to synthesize a new set of B-DMC's $\{W_N^{(i)} : 1 \leq i \leq N\}$ where some of them have absolute reliability.

From the implementation's point of view the channel polarization process consists of many levels. The root of this multilevel recursive operation is realized by the following simple scheme. We set x_1 (the bit transmitted through the top copy of W) to be a function of both u_1 and u_2 (sum mod-2 a.k.a. XOR) while x_2 equals u_2. By combining a pair of copies of W, we construct a new composite channel denoted as W_2 with two bits as input and two bits as output, as shown in Fig. 4.

At the random i^{th} level of polarization we combine two channels $W_{\frac{N}{2}}$, as it is illustrated by Fig. 5. For an array of inputs $\{u_i\}_{i=1}^N$ we obtain a new one ($\{v_i\}_{i=1}^N$) by replacing the odd indexed u's with the XOR of consecutive pairs. That is, the odd indexed v_{2i-1} will be equal to $u_i \oplus u_{i+1}$ for each $i = 1, \ldots, \frac{N}{2}$. The rest u's (that are even indexed) are passed to the reverse shuffle operator as they are. Reverse shuffling distincts the odd and even indexed elements and passes them to separate channels $W_{\frac{N}{2}}$; one channel for the odd indexed u's (top) and one for the even indexed u's (bottom).

Regarding the reverse shuffle, it is interesting to note that grouping the odd and the even indexed inputs is a permutation that can be further comprehended as re-ordering the tuple of bits that is associated with each input. This tuple is the binary expression of $i - 1$. For instance, the first input (v_1) is associated with the all-zero tuple of length $n = \log_2 N$, that is $v_1 \rightarrow v_{00\ldots0}$. Accordingly, the last input is associated with the all-one tuple, that is $v_N \rightarrow v_{11\ldots1}$. In general, $v_{b_1 b_2 \ldots b_n} \leftarrow v_i$ where $i = 1 + \sum_{j=1}^{n} b_j 2^{n-j}$. Hence, reverse shuffle changes the order of this bit string as follows: $v_{b_1 b_2 \ldots b_n} = v_{b_2 b_3 \ldots b_n b_1}$. In other words, R_N cyclically right-shifts by one the bit-indices of the elements of a left operand u_1^N.

The essential transformation after $\log_2(N)$ levels is linear. By this, we mean that before transmitting the information word u_1^N we first apply a linear transformation by multiplying it (from the right) with a matrix denoted as G_N. Each entry of the resulting codeword $x_1^N = u_1^N G_N$ is then transmitted independently though each copy of W. As we can see, transmitting a word u_1^N through the composite channel W_N is equivalent to transmitting x_1^N through N independent copies of channel W. An analysis via linear algebra results in $G_N = B_N F^{\otimes n}$ where B_N is a permutation matrix and

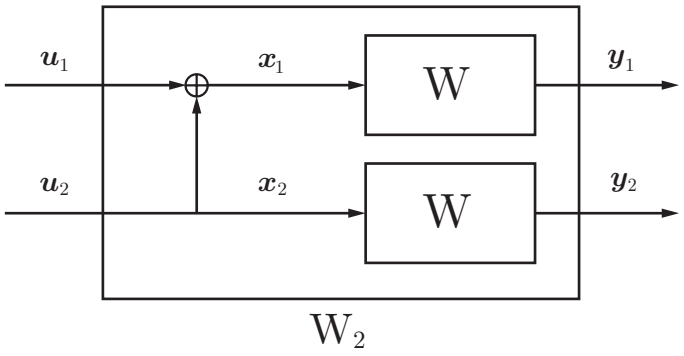

Figure 4: The channel W_2.

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

The permutation matrix B_N inverts the bits of the tuple that each input is associated with. Specifically, if $p_1^N = q_1^N B_N$, then $p_{b_1 b_2 \ldots b_n} = q_{b_n b_{n-1} \ldots b_1}$ where $n = \log_2 N$.

The transition probabilities of the two channels W_N and W^N are related by

$$W_N(y_1^N | u_1^N) = W^N(y_1^N | u_1^N G_N) \tag{15}$$

Having synthesized the vector channel W_N out of W^N, the next step of channel polarization is to split W_N back into a set of N binary-input coordinate channels $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}, 1 \leq i \leq N$, defined by the transition probabilities

$$W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \doteq \sum_{u_{i+1}^N \in \mathcal{X}^{N-1}} \frac{1}{2^{N-1}} W_N(y_1^N | u_1^N) \tag{16}$$

where (y_1^N, u_1^{i-1}) donates the output of $W_N^{(i)}$ and u_i its input, as shown in Fig. 6. Intuitively, if one considers a genie-aided decoder in which the ith decision element estimates u_i after observing y_1^N and the past channel inputs u_1^{i-1}, then (if u_1^N is *a priori* uniform) $W_N^{(i)}$ is the effective channel seen by the ith decision element.

The encoder is implemented as shown in Fig. 5. The complexity is derived with the help of Master

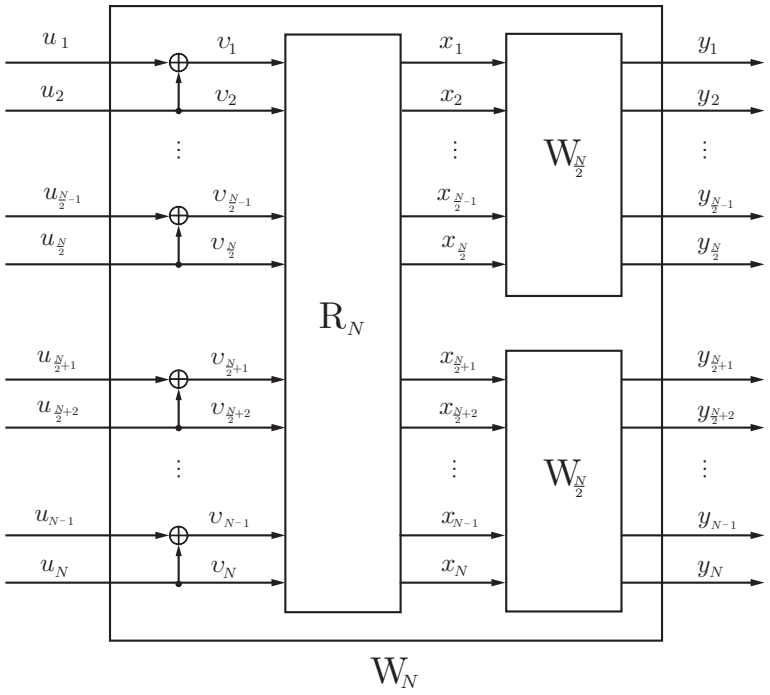

Figure 5: Recursive construction of W_N from two copies of $W_{\frac{N}{2}}$.

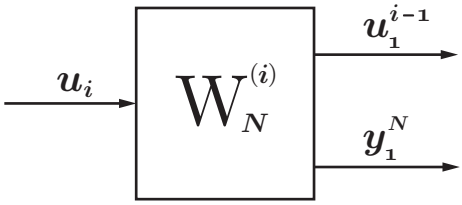

Figure 6: $W_N^{(i)}$.

Theorem

$$T(N) = \frac{N}{2} + \Theta(N) + 2T\left(\frac{N}{2}\right) \Rightarrow$$
$$T(N) = \Theta(N \log_2 N). \tag{17}$$

3.4 Code Construction

The theorem which the entire study is based on is the following: *For any B-DMC W, the channels* $\{W_N^{(i)}\}$ *polarize in the sense that, for any fixed* $\delta \in (0,1)$*, as N goes to infinity through powers of two, the fraction of indices* $i \in \{1 \ldots, N\}$ *for which* $I(W_N^{(i)}) \in (1-\delta, 1]$ *goes to I(W) and the fraction for which* $I(W_N^{(i)}) \in [0, \delta)$ *goes to* $1 - I(W)$.

The problem that we face as code designers is to determine N (the block length), A(the set of the indices to the good channels), K (the number of the good channels), and u_{A^c} (the value of the frozen bits - A_c is the compliment of A over $\{1,\ldots,N\}$). Regarding to what values frozen bits take, the resulting codeword will be a coset of $u_A G_N(A) : x_1^N = u_A G_N(A) \oplus u_{A^c} G_N(A^c)$ - where by $G_N(A)$ and $G_N(A^c)$ we denote the matrices consisting only of the rows of G_N with index in the set A or A^c, respectively, e.g. by $G_N(\{4,6,7,8\})$ we refer to the fourth, sixth, seventh, and eighth rows of G_N.

The answers to the previous questions (N?, A?, K?, u_{A^c}?) are: the larger N is the better in terms of convergence. This is based on the fact that the fraction of channels that have not yet polarized is vanishing as N grows to infinity. Hence, it is more likely that we choose channels with symmetric information close to one. K is a function of N and the rate that we choose to transmit, namely $K = \lfloor RN \rfloor$. The information set A is chosen as the K-element subset of $\{1,\ldots,N\}$ such that $Z(W_N^{(i)}) \leq Z(W_N^{(j)})$ for all $i \in A$, $j \in A^c$ (this metric will prove to be capacity achieving). Finally, u_{A^c} can be chosen arbitrarily. We have this degree of freedom due to the fact that all the distances are preserved for any coset code. That makes those coset codes equivalent to the original ($u_{A^c} = \mathbf{0}^{N-K}$). Notice that the receiver must have knowledge of u_{A^c} in order to decode appropriately.

3.5 Successive Cancellation Decoding

The decoder proposed in [6] is called successive cancellation (SC) decoder and its role is to decide with the rule of closest neighbor on i^{th} bit ($1 \leq i \leq N$) that transmitted over $W_N^{(i)}$. A SC decoder generates its decision $\hat{u}_1^N$ by computing

$$\hat{u}_i \triangleq \begin{cases} u_i & \text{, if } i \in A^c \\ h_i(y_1^N, \hat{u}_1^{i-1}) & \text{, if } i \in A \end{cases} \tag{18}$$

in the order i from 1 to N, where $h_i : \mathcal{Y}^N \times \mathcal{X}^{i-1} \to \mathcal{X}, i \in A$, are decision functions defined as

$$h_i(y_1^N, \hat{u}_1^{i-1}) \triangleq \begin{cases} 0 & \text{, if } W_N(y_1^N, u_1^{i-1}|0) \geq W_N(y_1^N|u_1^{i-1}|1) \\ 1 & \text{, otherwise} \end{cases} \tag{19}$$

for all $y_1^N \in \mathcal{Y}^N, \hat{u}_1^{i-1} \in \mathcal{X}^{i-1}$.

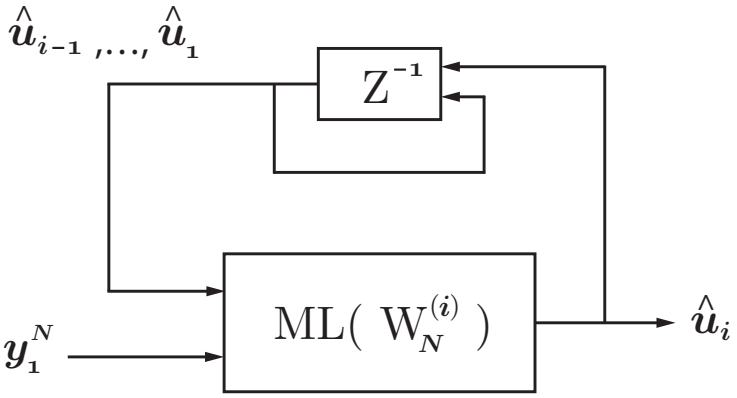

Figure 7: Decoder's scheme.

In this work we focus on the special case of BEC. For the case that W is a BEC with an erasure probability ϵ, the Bhattacharyya parameters $Z(W_N^{(i)})$ needed for the construction of the sets A and A^c can be computed efficiently through the recursion

$$\begin{aligned} Z(W_N^{(2j-1)}) &= 2Z(W_{N/2}^{(j)}) - Z\left(W_{N/2}^{(j)}\right)^2 \\ Z(W_N^{(2j)}) &= Z\left(W_{N/2}^{(j)}\right)^2 \end{aligned} \tag{20}$$

with $Z(W_1^{(1)})$.

Moreover, we can calculate the transition probabilities at (19) by the efficient recursive formulas (21) and (22) and an implementation that reduces the asymptotic complexity to $\mathcal{O}(N \log N)$ for both time and space.

$$\begin{aligned} &W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} \mid u_{2i-1}) = \\ &\sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i}) \end{aligned} \tag{21}$$

$$\begin{aligned} &W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} \mid u_{2i}) = \\ &\frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i}) \end{aligned} \tag{22}$$

We may visualize the decoder as consisting of N decision elements (DEs), for each source element u_i. $\mathcal{O}(N \log N)$ space is required to store the values that can be reused (in fact, every value will be

reused). To see where the computational saving will come from, we inspect (21) and (22) and note that each value in the pair

$$\left(W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} \mid u_{2i-1}), W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} \mid u_{2i})\right)$$

is assembled from the same pair of values

$$\left(W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}), W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i})\right)$$

The size of the matrices in which these values will be stored is $N \times (\log_2 N + 1)$. The number of the elements are exactly as many as the distinct $W_N^{(i)}$ that will be eventually calculated. Each cell is filled after $\Theta(1)$ calculations which implies that the complexity to decode each word is $\mathcal{O}(N \log N)$.

For the general case of a B-DMC W, we can calculate the likelihood ratios (LRs)

$$L_N^{(i)}(y_1^N, u_1^{i-1}) \doteq \frac{W_N(y_1^N, u_1^{i-1}|0)}{W_N(y_1^N|u_1^{i-1}|1)} \tag{23}$$

A straightforward calculation using the recursive formulas (21) and (22) gives

$$L_N^{(2i-1)}(y_1^{2N}, u_1^{2i-2}) = \frac{L_{\frac{N}{2}}^{(i)}(y_1^{\frac{N}{2}}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}) \cdot L_{\frac{N}{2}}^{(i)}(y_{\frac{N}{2}+1}^{N}, u_{1,e}^{2i-2}) + 1}{L_{\frac{N}{2}}^{(i)}(y_1^{\frac{N}{2}}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}) + L_{\frac{N}{2}}^{(i)}(y_{\frac{N}{2}+1}^{N}, u_{1,e}^{2i-2})} \tag{24}$$

$$L_N^{(2i)}(y_1^N, u_1^{2i-1}) = \left[L_{\frac{N}{2}}^{(i)}(y_1^{\frac{N}{2}}, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2})\right]^{1-2\hat{u}_{2i-1}} \cdot L_{\frac{N}{2}}^{(i)}(y_{\frac{N}{2}+1}^{N}, u_{1,e}^{2i-2}) \tag{25}$$

Thus, the calculation of an LR at length N is reduced to the calculation of two LRs at length $N/2$. This recursion can be computed down to block length 1, at which point the LRs have the form $L_1^{(1)}(y_i) = W(y_i|0)/W(y_i|1)$ and can be computed directly.

At this point it is interesting to show how the capacity of different binary erasure channels changes regarding to their erasure probability. In Fig. 8 the horizontal axis is marked with the code rate and the vertical axis is marked with the magnitude of the probability of block error between transmitter and receiver. The capacity of BEC is $C(\epsilon) = (1 - \epsilon)$ and we note that reliable communication (zero probability of bit error) is accomplished with capacity-achieving rates.

In Fig. 9 we plot the bit error rate (Pe) as a function of the indices of the channels $W_N^{(i)}$ normalized to 1, for different code rates. We observe that as the code rate R grows, the fraction of indices $i \in \{1 \dots, N\}$ for which the probability of error is 0.5, hence they are totally useless, grows too.

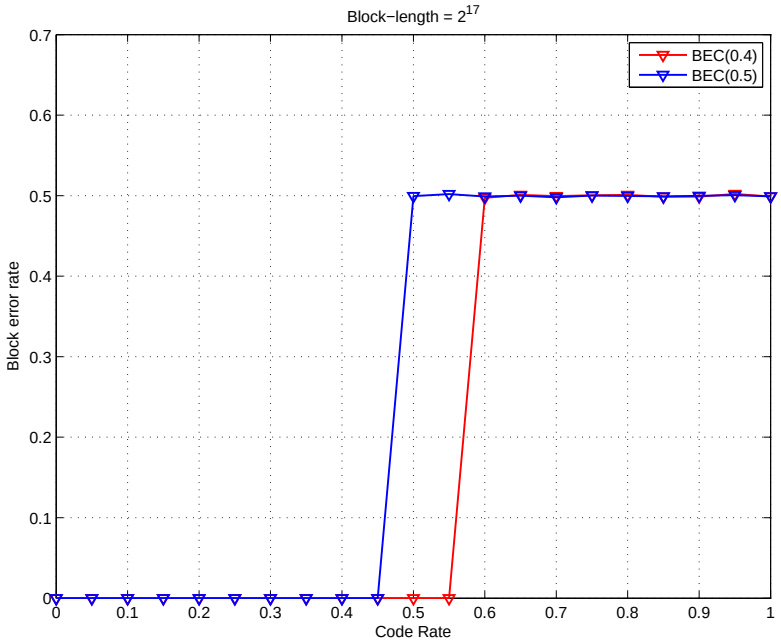

Figure 8: Rate vs BER for the BEC.

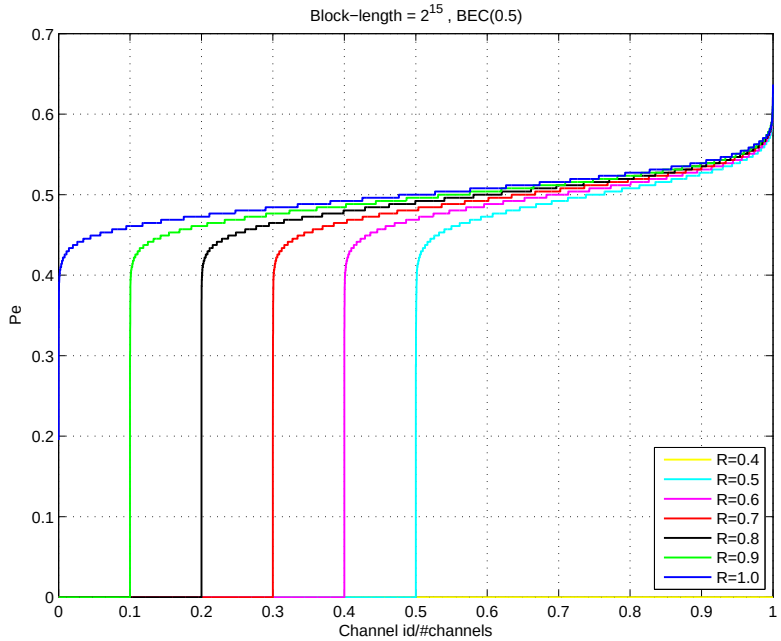

Figure 9: Normalized channel indices vs P_e for the BEC.

4 Polar Codes for Secrecy

The intuition behind the presented secrecy coding scheme is that we take advantage of the fact that the wiretap channel C_2 is degraded with respect to C_1, hence the capacity of C_1 is greater than the capacity of C_2. This is illustrated in Fig. 10.

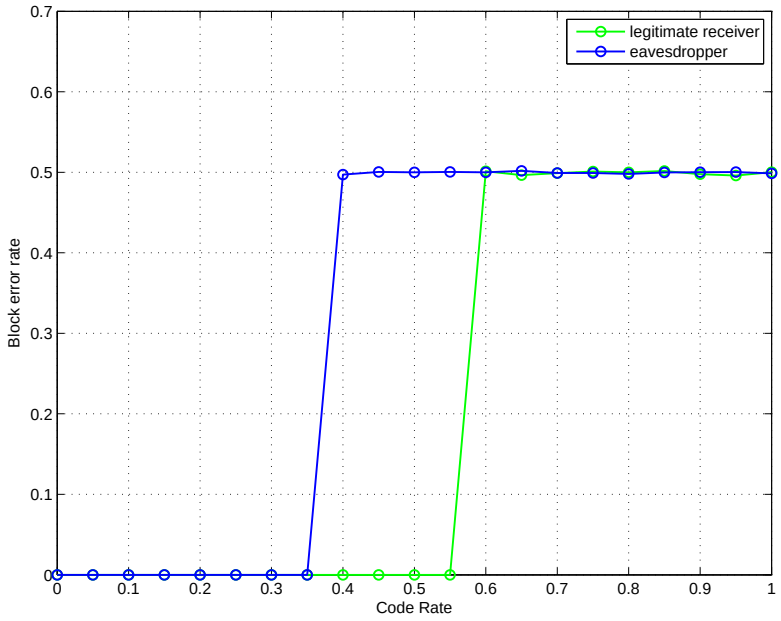

Figure 10: Rate vs reliability.

We consider a special case of the wiretap-channel system of Fig. 1, wherein the main channel $C_1 = W^*(y|x)$ and Eves wiretap channel $C_2 = W(z|x)$ are BECs, and C_2 is degraded with respect to C_1. Note that the scenario result is satisfied immaterial of whether the eavesdropper adheres successive decoding. In fact, we consider that the eavesdropper knows everything that the legitimate receiver is aware of. As shown in Fig. 11, this secrecy polar scheme is based on transmitting the secret message only over those bit-channels $W_N^{(i)}$ that are bad for Eve, while flooding the bit-channels that are good for Eve with random bits.

For a given block length $n = 2^i$, $i \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$ and let as define three subsets of $[n]$ as follows:

$$\mathcal{R} \doteq A(W) \tag{26}$$

$$\mathcal{S} \doteq A(W^*) \setminus A(W) \tag{27}$$

$$\mathcal{B} \doteq A^c(W^*). \tag{28}$$

A is the set of the indices to the good channels defined in section 3.4, $A(W)$ refers to the information set of channel W, and A^c is the compliment of A over $[n]$. Note that the sets $\mathcal{R}, \mathcal{S}, \mathcal{B}$ are disjoint and $\mathcal{R} \cup \mathcal{S} \cup \mathcal{B} = [n]$. Let $|\mathcal{R}| = r$ and $|\mathcal{S}| = k$.

4.1 The Encoding and Decoding Algorithms

The encoder is a function $\mathcal{E} : \{0,1\}^k \times \{0,1\}^r \to \{0,1\}^n$. It accepts as input a message $\mathbf{u} \in \{0,1\}^k$ and a vector $\mathbf{e} \in \{0,1\}^r$. We assume that $\mathbf{e}$ is selected by Alice uniformly at random from $\{0,1\}^r$. The encoder first constructs the vector $\boldsymbol{v} \in \{0,1\}^n$, by setting $\boldsymbol{v}_{\mathcal{R}} = \mathbf{e}$, $\boldsymbol{v}_{\mathcal{S}} = \boldsymbol{v}$, and $\mathbf{u}_{\mathcal{B}} = \mathbf{0}$ and then outputs $\mathcal{E}(\mathbf{u}, \mathbf{e}) := \boldsymbol{v}G_N$ as shown in section 3.3.

The decoder is a function $\mathcal{D} : \mathcal{Y}^n \to \{0,1\}^k$. It accepts as input a vector $\mathbf{y} \in \mathcal{Y}^n$ at the output of the main channel . It then invokes successive cancellation decoding for the polar code $\mathbb{C}_n(\mathcal{S} \cup \mathcal{R})$, used over W^*, to produce the vector $\hat{\boldsymbol{v}} \in \{0,1\}^n$. The decoder outputs $\mathcal{D}(\mathbf{y}) := \hat{\boldsymbol{v}}_{\mathcal{S}}$.

The sets $\mathcal{R}, \mathcal{S}, \mathcal{B}$, the polar generator matrix G_N, and the value of the frozen bits, are all known to both the legitimate user and the eavesdropper.

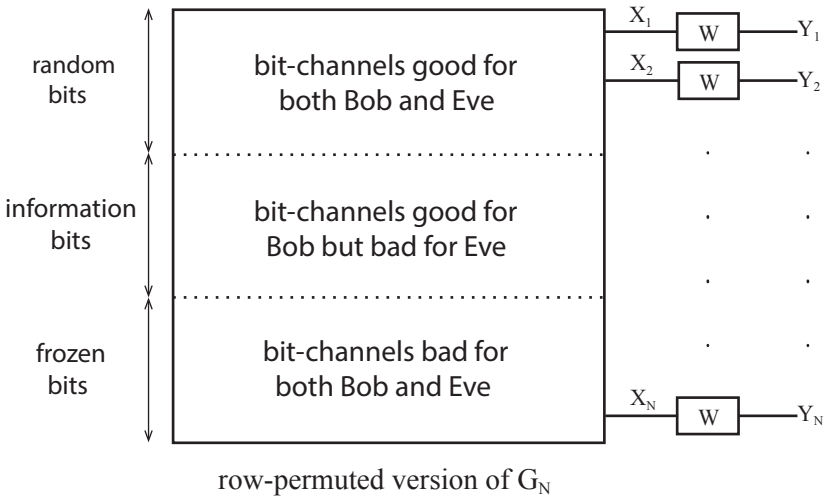

Figure 11: Secrecy structure.

4.2 Secrecy achieving properties

The information bits $\mathbf{U}^k$ reach Bob via good (almost noiseless) bit-channels. Thus, Bob should be able to reconstruct them with very high probability. On the other hand, these same bits pass through bad (almost useless) bit-channels on their way to Eve. Thus, Eve should not be able to deduce much information about $\mathbf{U}^k$ from her observations $\mathbf{U}^n$, and $H(\mathbf{U}^k \mid \mathbf{Z}^n)$ should be close to $H(\mathbf{U}^k)$. However, this simple intuition is misleading, because it does not show how the random bits in Fig. 11 help keep Eve ignorant. It may appear that this randomness is not really needed. For example, what would happen if the vector that serves as the second input to our encoder is not chosen at random from $\{0,1\}^r$ but rather set to an *a priori* fixed value? Since the channels are symmetric, any fixed value is as good as any other, so we may as well assume $\mathbf{e} = \mathbf{0}$. This does not seem to affect the argument in the foregoing paragraph and, according to this argument, $H(\mathbf{U}^k \mid \mathbf{Z}^n)$ would still be close to $H(\mathbf{U}^k)$.

In fact, this is not true. The reason is that channels seen by individual input bits as they undergo the encoding transformation depend on the distribution of other input bits. Specifically, if $\mathbf{e} = \mathbf{0}$ or $\mathbf{e}$ is fixed, the resulting encoder will not be secure. To prove that we need the following simple lemma [7].

Lemma 1: Let $\mathcal{F}$ be an arbitrary subset of $[n]$ of size k, and suppose that the polar code $\mathbb{C}_n(\mathcal{F})$ is used to communicate over a BSM channel W. Further, assume that the message $\mathbf{U}^k$ at the input to the encoder for $\mathbb{C}_n(\mathcal{F})$ is uniformly random over $\{0,1\}^k$, and let $\mathbf{Z}^n$ denote the random vector at the channel output. Then $I(\mathbf{U}^k; \mathbf{Z}^n) \geq kC(W)$.

Proof: Let $\mathbf{V}$ be the random vector obtained by setting $\mathbf{V}_{\mathcal{F}} = \mathbf{U}$ and $\mathbf{V}_{\mathcal{F}^c} = \mathbf{0}$. Then the codeword transmitted over the channel is

$$\mathbf{X} = \mathbf{V}G_N = \mathbf{U}M \tag{29}$$

where M is a $k \times n$ row submatrix of G_N. Since G_N is nonsingular, $rank(M) = k$ and there exists a subset $\mathcal{T}$ of $[n]$ of size k such that the corresponding k columns of M are linearly independent. This implies that there is a one-to-one correspondence between $\mathbf{U}$ and $\mathbf{X}_{\mathcal{T}}$. Hence, $I(\mathbf{U}^k; \mathbf{Z}^n) = I(\mathbf{X}_{\mathcal{T}}^k; \mathbf{Z}^n)$. Furthermore, since the random vector $\mathbf{U}$ is uniform over $\{0,1\}^k$, so is the vector $\mathbf{X}_{\mathcal{T}}$. Equivalently, its components $\{X_i : i \in \mathcal{T}\}$ are i.i.d. random variables Ber(1/2), and we can conclude that

$$I(\mathbf{X}_{\mathcal{T}}^k; \mathbf{Z}^n) \geq I(\mathbf{X}_{\mathcal{T}}^k; \mathbf{Z}_{\mathcal{T}}^n) = \sum_{i \in \mathcal{T}} I(X_i; Z_i) = kC(W) \tag{30}$$

where the last two equalities follow from the fact that the channel W is memoryless and symmetric. ■

Now suppose that the input to our encoder $\mathcal{E}(\cdot,\cdot)$ is a message chosen uniformly at random from $\{0,1\}^k$ along with $\mathbf{e} = \mathbf{0}$. This is a special case of the situation considered in Lemma 1, with the set $\mathcal{F}$ given by (27). Hence, $I(\mathbf{U}^k;\mathbf{Z}^n) \geq kC(W)$, and security condition (2) cannot be satisfied: a significant fraction of message bits, at least $C(W)$, is potentially exposed.

This polar coding scheme satisfies the reliability and security conditions (1) and (2) while its rate approaches the secrecy capacity. For the wiretap-channel system considered in this work, the secrecy capacity is $C_s = C_1 - C_2$ (see (10)). Let $R_s = k/n$ donate the rate of our coding scheme.

Proposition 1:

$$\lim_{n\to\infty} R_s = C(W^*) - C(W)$$

Proof: Observe that

$$R_s = \frac{|\mathcal{S}|}{n} = \frac{|A(W^*)|}{n} - \frac{|A(W)|}{n} \tag{31}$$

where we have used the definition of $\mathcal{S}$ in (27) and the fact that $A(W)$ is a subset of $A(W^*)$. ■

4.3 Performance Results

To illustrate the performance of this secrecy scheme, the following three figures are chosen as the most representative of the performance of polar codes for the wiretap communication model. We consider the wiretap-channel system of Fig. 1, wherein the main channel is a BEC with erasure probability $\epsilon_1 = 0.4$ and the wiretap channel is also a BEC. The horizontal axis is marked with the indices of the channels $W_N^{(i)}$ normalized to 1 and the vertical axis is marked with the probability of bit error (P_e).

The first one shows the BER at the legitimate receiver and the eavesdropper for various block lengths 2^n. The wiretap channel here is a BEC with erasure probability $\epsilon_2 = 0.5$. As we expect, the fraction of the bit-channels that are totally useless approaches better the secrecy capacity as the block length grows.

Next we have the performance of polar codes for three cases of the wiretap channel (i.e., for BEC(0.5),BEC(0.6) and BEC(0.8)), hence three values of the secrecy capacity C_s and we confirm that

our coding scheme corresponds correctly to the changes of the wiretap channel.

And finally, in Fig. 14 for the wiretap channel with erasure probability $\epsilon_2 = 0.5$, we consider two different code rates and we observe that a larger rate performs better according to the fraction of the channels $W_N^{(i)}$ that can be used for secrecy but inserts error at the legitimate receiver's transmission.

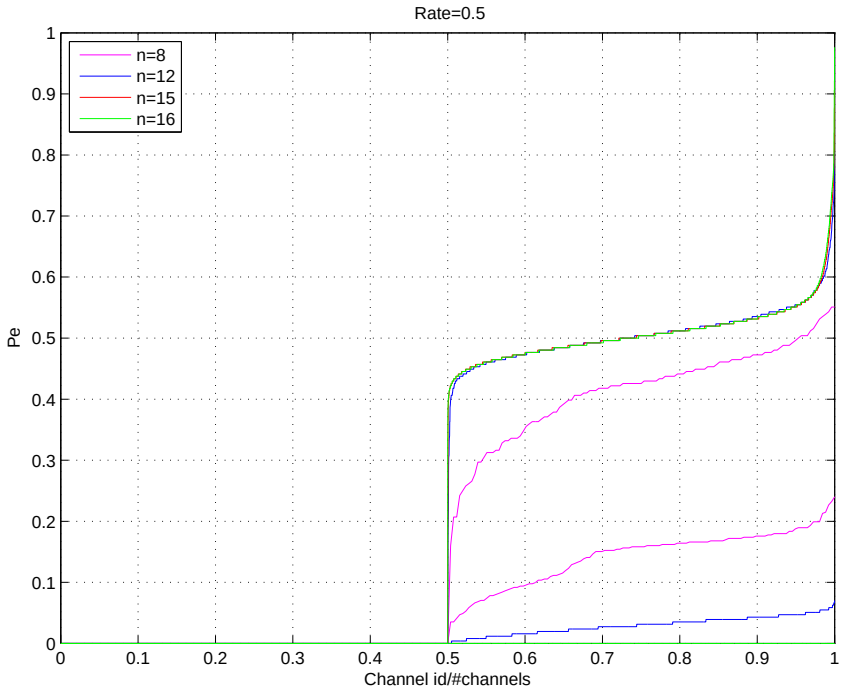

Figure 12: Normalized channel indices vs P_e for the wiretap channel (different block lengths).

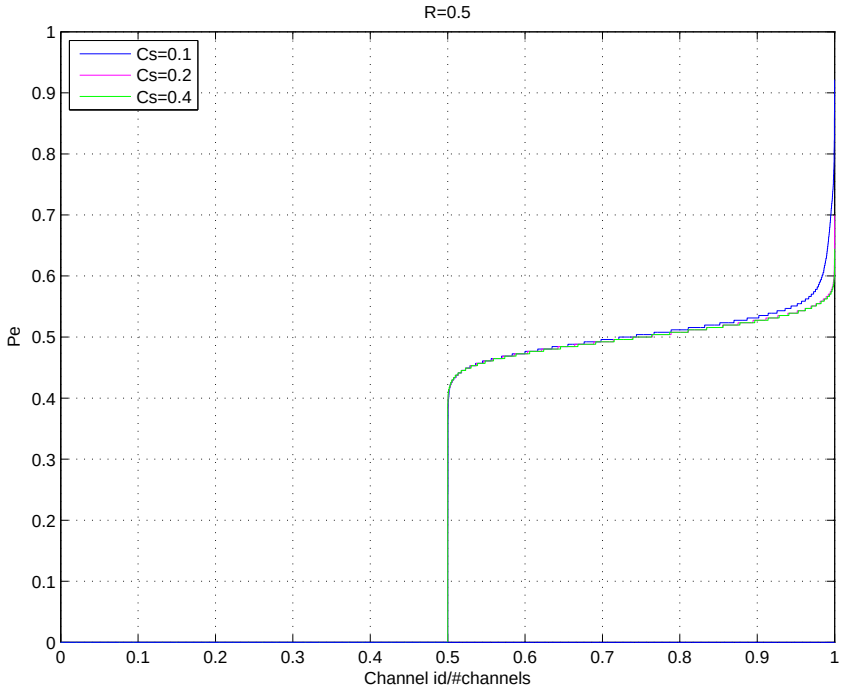

Figure 13: Normalized channel indices vs P_e for the wiretap channel (different secrecy capacities).

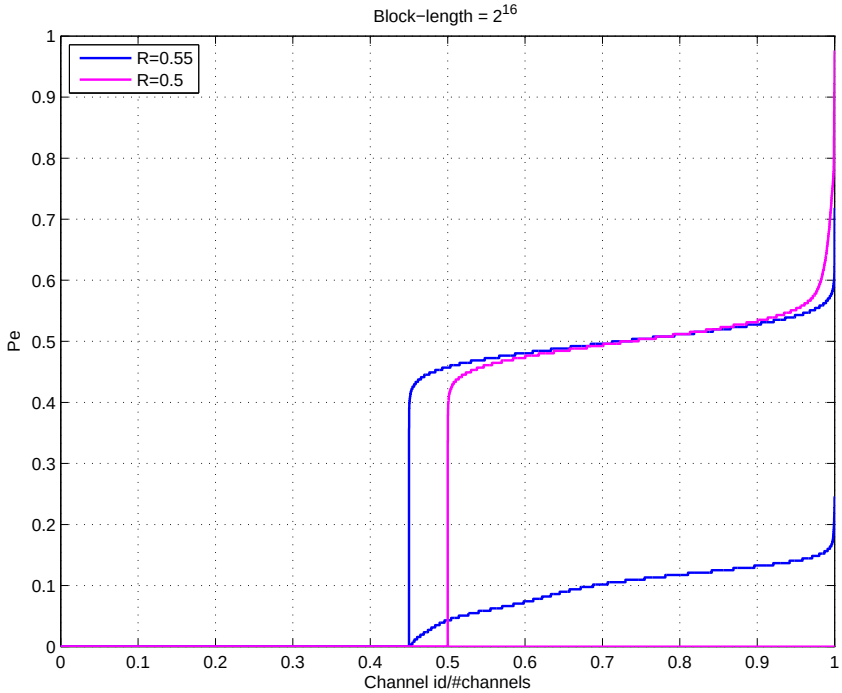

Figure 14: Normalized channel indices vs P_e for the wiretap channel (different code rates).

References

[1] C.E. Shannon, "Communication theory of secrecy systems," Bell Syst. Tech.J., vol. 28, pp. 656-715, 1949.

[2] A.D. Wyner, "The wire-tap channel," Bell Syst. Tech.J., vol. 54, no. 8, pp. 1355-1387, Oct 1975.

[3] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. McLaughlin, and J.-M. Merolla, "Applications of LDPC codes to the wiretap channel," IEEE Trans. Inf. Theory, vol. 53, no. 8, pp. 2933-2945, Aug. 2007.

[4] 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.

[5] R. Liu, Y. Liang, H. V. Poor, and P. Spasojevic, "Secure nested codes for type II wiretap channels," in Proc. IEEE Information Theory Workshop (ITW), Lake Tahoe, CA, September 2-6, 2007, pp. 337-342.

[6] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 30513073, Jul. 2009.

[7] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," IEEE Trans. Inform. Theory, vol. 57, no. 10, pp. 64286443, October 2011.

[8] E. Hof and S. Shamai, "Secrecy-achieving polar-coding, in Proc. IEEE Information Theory Workshop, Dublin, Ireland, Aug.-Sep. 2010, pp. 15.

[9] O. O. Koyluoglu and H. El-Gamal, "Polar Coding for Secure Transmission and Key Agreement," IEEE Trans. on Information Forensics and Security, vol. 7, no. 5, Oct. 2012

[10] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer and M. Skoglund, "Nested Polar Codes for Wiretap and Relay Channels, IEEE Communications Letters, vol. 14, no. 8, Aug. 2010.

[11] I. Csiszár and J. Körner, "Broadcast channels with confidential messages, IEEE Trans. Inf. Theory, vol. IT-24, no. 3, pp. 339348, May 1978.

[12] S. Leung-Yan-Cheong, "On a special class of wire-tap channels, IEEE Trans. Inf. Theory, vol. IT-23, no. 5, pp. 625627, Sep. 1977.