ACHIEVING SECRECY CAPACITY OF THE WIRETAP CHANNEL AND BROADCAST CHANNEL WITH A CONFIDENTIAL COMPONENT

TALHA CIHAD GULCU∗ AND ALEXANDER BARG∗∗

ABSTRACT. The wiretap channel model is one of the first communication models with both reliability and security constraints. Explicit capacity-achieving schemes for various models of the wiretap channel have received considerable attention in recent literature. In this paper we address the original version of this problem, showing that capacity of the general (not necessarily degraded or symmetric) wiretap channel under a “strong secrecy constraint” can be achieved using an explicit scheme based on polar codes. We also extend our construction to the case of broadcast channels with confidential messages defined by Csiszár and Körner (1978), achieving the entire capacity region of this communication model.

Index terms: Polar codes, chaining construction, strong secrecy, coordinate partition.

1. INTRODUCTION

The wiretap channel model was introduced by Wyner in 1975 [23]. In this model, there are two receivers Y, Z and a single transmitter X. The transmitter aims at sending messages to Receiver 1 through a communication channel W1. The information sent from X to Y is also received by Receiver 2 through another channel W2. The transmission problem in the system W(W1, W2) calls for designing a coding system that supports communication between X and Y in a way that is both reliable and secure. The reliability requirement is the usual one for communication systems, namely, that the error probability of decoding the information by Y be made arbitrarily low by increasing the block length of the encoding. At the same time, the transmission needs to be made secure in the sense that the information extracted by Receiver 2 about the message of X approaches zero as a function of the block length.

To describe the problem in formal terms, denote the input alphabet of the transmitter by X, and the output alphabets of the channels W1 and W2 by Y and Z, respectively. The messages that the transmitter can convey to Receiver 1 form a finite set denoted below by M. For transmission over the channel the message is encoded using a mapping f : M → XN, where XN is an N-fold repetition of the input alphabet. We say that f is a length-N block encoder of the transmitter. It will turn out that better transmission rates can be obtained by using a randomized version of the encoder, i.e., a mapping sends M to a probability distribution on XN. In other words, the message m ∈ M is encoded as a sequence xN ∈ XN with probability f(xN|m), and the encoder is defined as a matrix of conditional probabilities (f(xN|m))xN∈XN for xN ∈ XN.

The decoder of Receiver 1 is a mapping ϕ : YN → M. We also denote by PY|X and PZ|X the conditional distributions induced by the channels W1 and W2, respectively, and define the induced distributions PN Y|X, PN Z|X, where, for instance, PN Y|X(yN|xN) = PN Y|X(yi|xi), where yi and xi refer to the i-th symbol of the vectors yN and xN, respectively.

∗ Department of ECE and Institute for Systems Research, University of Maryland, College Park, MD 20742, Email: gulcu@umd.edu. Research supported in part by NSF grant CCF1217245.
∗∗ Department of ECE and Institute for Systems Research, University of Maryland, College Park, MD 20742, and IITP, Russian Academy of Sciences, Moscow, Russia. Email: abarg@umd.edu. Research supported in part by NSF grants CCF1217245, CCF1217894, and CCF1422955.
Definition 1.1. We say that the encoder-decoder pair \((f, \phi)\) gives rise to \((N, \epsilon)\)-transmission over the wiretap channel \(W\) if

\[
\sum_{x^N \in X^N} f(x^N|m) P_{Y|X}^N(\phi(y^N) = m|x^N) \geq 1 - \epsilon \quad \forall m \in M
\]  

(1)

\[
I(M; Z^N) \leq \epsilon.
\]  

(2)

where \(M\) is the message random variable (RV) and \(Z^N\) is the RV that corresponds to the observations of Receiver 2.

In Definition 1.1, Eq. (1) represents the reliability of communication condition while (2) answers the security of transmission requirement. We note that in many works on transmission with a secrecy constraint the security condition was formulated in a more relaxed way, namely as the inequality

\[
(1/N) I(M; Z^N) < \epsilon.
\]  

(3)

This is particularly true about pre-1990s works in information theory, but also applies to some very recent works on the wiretap channel, e.g., \([20, 9, 11]\). However, as shown by Maurer in \([13, 14]\), this constraint does not fulfill the intuitive security requirements in the system. More specifically, it is possible to construct examples in which inequality (3) is satisfied and at the same time Receiver 2 is capable of learning \(N^{1-\epsilon}\) out of \(N\) bits of the encoding \(x^N\). In view of this, Maurer suggested (2) as a better alternative to condition (3). As a result, currently (3) is called the “weak security constraint” as opposed to the stronger constraint (2). In this paper we design coding schemes that provide strong secrecy, so below we work only with condition (2).

The secrecy capacity of the wiretap channel is defined as follows.

Definition 1.2. The value \(R > 0\) is called an achievable rate for the wiretap channel \(W\) if there exists a sequence of message sets \(M_N\) and encoder-decoder pairs \((f_N, \phi_N)\) giving rise to \((N, \epsilon_N)\) transmission with \(\epsilon_N \to 0\) and \(\frac{1}{N} \log |M_N| \to R\) as \(N \to \infty\). The secrecy capacity \(C_s\) is the supremum of achievable rates for the wiretap channel.

The following theorem provides an expression for \(C_s\).

Theorem 1. ([7]; see also [8, p.411]) The secrecy capacity of the wiretap channel \(W\) equals

\[
C_s = \max I(V; Y) - I(V; Z)
\]  

(4)

where the maximum is computed over all RVs \(V, X, Y, Z\) such that the Markov condition \(V \to X \to Y, Z\) holds true, and such that \(P_{Y|X} = W_1, P_{Z|X} = W_2\).

While most general constructive coding schemes for the wiretap channels rely on polar codes, there were some constructive solutions even before the publication of Arıkan’s seminal work [2]. At the same time, these schemes applied only to some special cases of the channels \(W_1, W_2\). For instance the case when \(W_1\) is noiseless and \(W_2\) is a binary erasure channel was addressed in [18, 20] which show that in this case the capacity \(C_s\) can be achieved using low-density parity-check codes. The results in [20] are based on the weak security assumption while strong security is considered in [18]. Moreover, [20] extends the construction to the cases when both \(W_1\) and \(W_2\) are erasure channels, and when \(W_1\) is noiseless and \(W_2\) is a binary symmetric channel. As usual with the application of low-density codes, the constructions are based on code ensembles and strictly speaking are not explicit.

Another special case of the wiretap channel relates to the combinatorial version of the erasure channel (the so-called wiretap channel of type II) in which Receiver 2 can choose to observe any \(t\) symbols out of \(N\) transmitted symbols. Constructive capacity-achieving solutions for this case are based on MDS codes [21] or extractors [6].

In [5], it is shown that \(C_s\) is achievable with strong security using invertible extractors, if both \(W_1\) and \(W_2\) are binary symmetric channels. Both encoding and decoding algorithms in [5] have polynomial complexity. Moreover, [5] also claims that its proof method can be easily extended to other wiretap channels as long as both \(W_1\) and \(W_2\) are symmetric.
After the introduction of polar codes by Arıkan, achieving $C_s$ via polar coding has been considered by different works, mostly under the degradedness assumption. Recall that a channel $W_2 : \mathcal{X} \rightarrow \mathcal{Z}$ is called degraded with respect to a channel $W_1 : \mathcal{X} \rightarrow \mathcal{Y}$ if there exists a stochastic $|\mathcal{Y}| \times |\mathcal{Z}|$ matrix $P_{Z|Y}(z|y)$ such that for all $x \in \mathcal{X}, z \in \mathcal{Z}$

$$P_{Z|X}(z|x) = \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x)P_{Z|Y}(z|y).$$

(5)

The wiretap channel $W$ is called degraded if the channel to the eavesdropper is degraded with respect to the main channel. In this case Theorem 1 affords a simpler formulation because there is no need in the auxiliary RV $V$. Namely in the degraded case the secrecy capacity equals [8, Probl. 17.8]

$$C_s = C(W_1) - C(W_2)$$

(6)

(this specialization is true under slightly more general assumptions, but we will not need them below).

Communication over degraded wiretap channels using polar codes was considered in a number of papers, notably, [12, 8, 1, 11]. The main result of these works is that secrecy capacity (5) can be attained under the weak security constraint. We note that the degraded case is easier to handle with polar codes because of the specific nature of the polar codes construction (more on this below in Sect. 5). Another step was made by [17] which suggested a polar coding scheme that attains the rate $C_s$ of a symmetric degraded wiretap channel $W$ under the strong security requirement (2).

The problem of attaining secrecy capacity of the general wiretap channel (4) under the strong secrecy condition and without the degradedness assumption was further studied in [19]. A polar coding scheme suggested in this work attains a transmission rate of $\max_{p_X(x)} [H(X|Z) - H(X|Y)]$. Clearly $H(X|Z) - H(X|Y) \leq C_s$ for all $X$ since one can take $V = X$ in the Markov chain $V \rightarrow X \rightarrow Y, Z$, which appears in Theorem 1. It is not immediately clear for which channels the result of [19] actually attains the secrecy capacity of $W$. At the same time, the coding scheme employed in [19] relies on two nested layers of the polarizing transform. The decoder for the second (outer) layer works with the probability distribution generated by the first decoder, which is not easily computable. Thus, the low complexity decoding claim of the construction made in [19] is not supported by the known decoding procedures for polar codes. For these reasons the construction in [19] does not resolve the question of constructing an explicit capacity-achieving scheme for the non-degraded case of wiretap channels.

In related works [22, 16] the problem of constructing capacity achieving schemes for wiretap channels was addressed for the case of quantum channels. The constructions suggested in these works attain symmetric secrecy capacity of quantum wiretap channels. These constructions require a shared secret key between the transmitter and Receiver 1. This requirement seems to be intrinsic to polar code constructions for this problem including our work. However the constructions in [22, 16] require a positive-rate shared key, which results in a communication scheme that transmits at rates separated from capacity.

To summarize, to the best of our knowledge the question of constructing explicit capacity-achieving transmission schemes for the non-degraded wiretap channel with strong secrecy is an open problem. It is this problem that we aim to solve in this paper by removing the degradedness assumption (5). We also do not assume that either of the channels $W_1, W_2$ is symmetric. The main idea of our work is to exploit the Markov chain conditions intrinsic to secure communication problems using polar codes. In Section 4 we propose a polar coding scheme that attains the secrecy capacity (4) under the strong security assumption. Both the encoding and decoding complexity estimates of our construction are $O(N \log N)$, where $N$ is the length of the encoding. In Section 5.2 we generalize our construction to cover the case when a part of transmitter’s message is public, i.e., is designed to be conveyed both to Receivers 1 and 2. This model, called a broadcast channel with confidential messages, is in fact the principal model in the founding work of Csiszár and Körner [7] on this topic.

Apart from the basic polar coding results [2], our solution of the described problems relies on the previous work on the wiretap channel [17], the polar coding scheme for the broadcast channel of [15], and the construction of polar codes for general memoryless channels in [10]. A new idea introduced in our solution is related to a stochastic encoding scheme that emulates the random coding proof of the capacity theorem in [7], whereby polarization is used for the values of the auxiliary random variable $V$ in Theorem 1 followed by a stochastic
encoding into a channel codeword. Another insight, which is particularly useful for the broadcast channel result in Sect. [5] is related to a partition of the coordinates of the transmitted block that enables simultaneous decoding of different parts of the transmitted message by both receivers, whereby the decoder of polar codes is used by the receivers according to their high- and low-entropy bits. It becomes possible to show that the receivers recover the bits designed to communicate with each of them with high probability, and that the secret part of the message is not accessible to the unintended recipient.

2. PRELIMINARIES ON POLAR CODING

We begin with recalling basic notation for polar codes and then continue with the scheme for capacity-achieving communication on discrete binary-input channels.

Let $W$ be a binary-input channel with the output alphabet $Y$, input alphabet $X = \{0,1\}$, and the conditional probability distribution $W_{Y|X}(\cdot|\cdot)$. Throughout the paper we denote the capacity and the symmetric capacity of $W$ by $C(W)$ and $I(W)$, respectively. We say $W$ is symmetric if $W_{Y|X}(y|1), y \in Y$ can be obtained from $W_{Y|X}(y|0), y \in Y$ through a permutation $\pi : Y \to Y$ such that $\pi^2 = \text{id}$. Note that if $W$ is symmetric then $I(W) = C(W)$.

Given a binary RV $X$ and a discrete RV $Y$ supported on $Y$, define the Bhattacharyya parameter $Z(X|Y)$ as

$$Z(X|Y) = 2 \sum_{y \in Y} P_Y(y) \sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)}.$$  

The value $Z(X|Y), 0 \leq Z(X|Y) \leq 1$ measures the amount of randomness in $X$ given $Y$ in the sense that if it is close to zero, then $X$ is almost constant given $Y$, while if it is close to one, then $X$ is almost uniform on $\{0,1\}$ given $Y$. The Bhattacharyya parameter $Z(W)$ of a binary-input channel $W$ is defined as

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

Clearly if $P_X(0) = P_X(1) = 1/2$, then $Z(X|Y)$ coincides with the value $Z(W)$ for the communication channel $W : X \to Y$.

For $N = 2^n$ and $n \in \mathbb{N}$, define the polarizing matrix (or the Arıkan transform matrix) as $G_N = B_N F^\otimes n$, where $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is the Kronecker product of matrices, and $B_N$ is a “bit reversal” permutation matrix [2]. In [2], Arıkan showed that given a symmetric and binary input channel $W$, an appropriate subset of the rows of $G_N$ can be used as a generator matrix of a linear code that achieves the capacity of $W$ as $N \to \infty$.

2.1. Symmetric Channel Coding. The material in this section and Section [22] is well understood, but it merits some space in the present paper because it helps us to define the terminology that is useful for the main results below. Given a symmetric binary-input channel $W$, define the channel $W^N$ with input alphabet $\{0,1\}^N$ and output alphabet $Y^N$ by the conditional distribution

$$W^N(y^N|x^N) = \prod_{i=1}^N W(y_i|x_i)$$

where $W(\cdot|\cdot)$ is the conditional distribution that defines $W$. Define a combined channel $\widetilde{W}$ by the conditional distribution

$$\widetilde{W}(y^N|u^N) = W^N(y^N|u^N G_N).$$

In terms of $\widetilde{W}$, the channel seen by the $i$-th bit $U_i, i = 1, \ldots, N$ (also known as the bit-channel of the $i$-th bit) can be written as

$$W_i(y_i^N, u_i^{i-1}|u_i) = \frac{1}{2^{n-1}} \sum_{u \in \{0,1\}^{n-i}} \widetilde{W}(y_i^N|(u_i^{i-1}, u_i, u)).$$
where \( u^{i-1} = (u_1, u_2, \ldots, u_{i-1}) \). We see that \( W_i \) is the conditional distribution of \((Y^N, U^{i-1})\) given \( U_i \) provided that the channel inputs \( X_i \) are uniformly distributed for all \( i = 1, \ldots, N \). The bit-channels thus defined are partitioned into good channels \( \mathcal{S}_N(W, \beta) \) and bad channels \( \mathcal{B}_N(W, \beta) \) based on the value of their Bhattacharyya parameters. Bearing in mind our notation choices later in the paper, we denote them by

\[
\mathcal{L}_{X|Y} = \mathcal{L}_{X|Y}(N) = \{ i \in [N] : Z(W_i) \leq \delta_N \},
\]

\[
\mathcal{H}_{X|Y} = \mathcal{H}_{X|Y}(N) = \{ i \in [N] : Z(W_i) > 1 - \delta_N \},
\]

where \( [N] = \{1, 2, \ldots, N\} \) and \( \delta_N \triangleq 2^{-N^\beta}, \beta \in (0, 1/2) \). As shown in \([4]\), for any symmetric binary-input channel \( W \) and any constant \( \beta < 1/2 \),

\[
\lim_{N \to \infty} \frac{|\mathcal{L}_{X|Y}(N)|}{N} = C(W),
\]

\[
\lim_{N \to \infty} \frac{|\mathcal{H}_{X|Y}(N)|}{N} = 1 - C(W).
\]

Based on this equality, information can be transmitted over the good-bit channels while the remaining bits are fixed to some values known in advance to the receiver (in polar coding literature they are called frozen bits). The transmission scheme can be described as follows: A message of \( k = |\mathcal{L}_{X|Y}| \) bits is written in the bits \( u_i, i \in \mathcal{L}_{X|Y} \). The remaining \( N - k \) bits are set to 0. This determines the sequence \( u^N \) which is transformed into \( x^N = u^N G_N \), and the vector \( x^N \) is sent over the channel. Denote by \( y^N \) the sequence received on the output. The decoder finds an estimate of \( u^N \) by computing the values \( \hat{u}^i, i = 1, \ldots, N \) as follows:

\[
\hat{u}_i = \begin{cases} \arg\max_{u \in \{0,1\}} W_i(y^N, \hat{u}^{i-1}|u), & \text{if } i \in \mathcal{L}_{X|Y}, \\ 0, & \text{if } i \in \mathcal{H}_{X|Y}. \end{cases}
\]

The results of \([2, 4]\) imply the following upper bound on the error probability \( P_e = \Pr(\hat{u}^N \neq u^N) \):

\[
P_e \leq \sum_{i \in \mathcal{L}_{X|Y}} Z(W_i) \leq N2^{-N^\beta} \leq 2^{-N^{\beta'}}
\]

where \( \beta \) is any number in the interval \((0, 0.5)\) and \( \beta' = \beta'(N) < \beta \).

This describes the basic construction of polar codes \([2]\) which attains symmetric capacity \( I(W) \) of the channel \( W \) with a low error rate.

**Remark 2.1.** There is a subtle point about the limit relations in \((8)\). Even though asymptotically the bit-channels are either good or bad, it is not true that \( \mathcal{L}_{X|Y} = \mathcal{H}_{X|Y} \) because there is a nonempty subset of indices \( \mathcal{L}_{X|Y} \setminus \mathcal{H}_{X|Y} \) of cardinality \( o(N) \) that is neither good nor bad. This distinction has no import for the simple situation of transmitting over \( W \), but leads to complications in the multi-user systems considered below; see, e.g., \([16]\) in Section 4 We use the terms “not good,” “not bad” to describe the corresponding subsets of bit channels.

### 2.2. General Channel Coding

Let \( W \) be a binary-input discrete memoryless channel \( W : \mathcal{X} \to \mathcal{Y} \) and let \( P_X \) be the capacity achieving distribution of \( W \). If \( P_X \) is not uniform, then the basic scheme attains a transmission rate \( I(W) \) which is less than \( C(W) \). This scheme was extended by Honda and Yamamoto \([10]\) to cover the case of arbitrary distributions \( P_X \). To explain the idea in \([10]\), for a given block length \( N \) define the sets

\[
\mathcal{H}_X = \{ i \in [N] : Z(U_i|U^{i-1}) \geq 1 - \delta_N \}
\]

\[
\mathcal{L}_X = \{ i \in [N] : Z(U_i|U^{i-1}) \leq \delta_N \}
\]

\[
\mathcal{H}_{X|Y} = \{ i \in [N] : Z(U_i|U^{i-1}, Y^N) \geq 1 - \delta_N \}
\]

\[
\mathcal{L}_{X|Y} = \{ i \in [N] : Z(U_i|U^{i-1}, Y^N) \leq \delta_N \}
\]

where \( U^N, X^N, Y^N \) have the same meaning as above. It can be shown \([3, 4]\) that

\[
\lim_{N \to \infty} \frac{1}{N} |\mathcal{H}_X| = I(X)
\]
\[
\lim_{N \to \infty} \frac{1}{N} |\mathcal{H}_X| = H(X|Y).
\]

We note that the set \( \mathcal{L}_{X|Y} \) is the set of good bit channels defined in (7). Unlike the case of uniform \( P_X \), it is not possible to use all of these channels to transmit information over \( W \). This is because if \( i \notin \mathcal{H}_X \), then \( U_i \) cannot be used to carry information conditioned on previous bits \( U^{i-1} \). Hence [10] argued that the set of information indices should be chosen as \( \mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y} \) rather than \( \mathcal{L}_{X|Y} \).

Since \( \mathcal{H}_{X|Y} \subseteq \mathcal{H}_X \) and the number of indices that are neither in \( \mathcal{H}_X \) nor in \( \mathcal{L}_{X|Y} \) is \( o(N) \), we have
\[
\lim_{N \to \infty} \frac{1}{N} |\mathcal{I}| = \lim_{N \to \infty} \frac{1}{N} |\mathcal{H}_X| - |\mathcal{H}_{X|Y}| = C(W),
\]
i.e., transmitting the information using the bits \( U_i, i \notin \mathcal{I} \) attains the capacity of the channel \( W \).

The code construction in [10] makes use of the following partition of the coordinate set \([N]\):
\[
\mathcal{F}_r = \mathcal{H}_X \cap \mathcal{L}_{X|Y},
\]
\[
\mathcal{F}_d = \mathcal{H}_X^c,
\]
\[
\mathcal{I} = \mathcal{H}_X \cap \mathcal{L}_{X|Y}
\]
where the superscript \( ^c \) refers to the complement of the subset in \([N]\). In terms of this partition, the encoding is done as follows. The information bits are stored in \( \{u_i, i \in \mathcal{I}\} \) while the bits \( \{u_i, i \in \mathcal{F}_r\} \) are taken to be independent samples of a Bernoulli-(1/2) RV and are shared with the receiver. Lastly, the bits \( \{u_i, i \in \mathcal{F}_d\} \) are determined according to the rule
\[
u_i = \arg\max_{u \in \{0, 1\}} P_{U_i|U^{i-1}}(u|u^{i-1}).
\]

Similarly to the symmetric channel coding case, once \( u^N \) is determined, the transmitter finds \( x^N = x^N G_N \) and sends it over the channel. The receiver uses the following successive decoding function: for \( i = 1, 2, \ldots, N \) let
\[
\hat{u}^i = \begin{cases} 
\arg\max_{u \in \{0, 1\}} P_{U_i|U^{i-1}, Y^N}(u|\hat{u}^{i-1}, y^N), & i \in \mathcal{I} \\
\arg\max_{u \in \{0, 1\}} P_{U_i|U^{i-1}}(u|\hat{u}^{i-1}), & i \in \mathcal{F}_d
\end{cases}
\]

(13)
The probability of decoding error can be bounded above similarly to [10]:
\[
P_e \leq \sum_{i \in \mathcal{I}} Z(U_i|U^{i-1}, Y^N) \leq N2^{-N\beta} \leq 2^{-N\beta'}
\]
where the parameters have the same meaning as before.

This completes the description of the capacity-achieving transmission scheme of [10]. We will rely in part on these ideas in our construction of a coding scheme for the wiretap channel.

3. A Closer Look at Prior Works on Polar Coding for the Wiretap Channel

To explain our proposal we will first discuss some of the schemes available in the literature. We begin with the transmission scheme of [12] (see also [11]). As already remarked, these works are concerned with the special case when the channel \( W_2 \) is degraded with respect to \( W_1 \) and aim to attain the rate value (6) with weak secrecy. Let \( X^N \) be a random uniform vector over \( \{0, 1\}^N \). Similarly to \( \mathcal{H}_{X|Y} \) and \( \mathcal{L}_{X|Y} \) given by (11), define the following subsets of indices:
\[
\mathcal{H}_{X|Z} = \{ i \in [N] : Z(U_i|U^{i-1}, Z^N) \geq 1 - \delta_N \}
\]
\[
\mathcal{L}_{X|Z} = \{ i \in [N] : Z(U_i|U^{i-1}, Z^N) \leq \delta_N \}
\]
where $U^N = X^N G_N$, and $Z^N$ is the output that Receiver 2 observes when the transmitter sends $X^N$. Partition the set $[N]$ as follows:

$$
\mathcal{R} = \mathcal{L}_{X|Z} \\
\mathcal{J} = \mathcal{L}_{X|Y} \setminus \mathcal{L}_{X|Z} \\
\mathcal{B} = \mathcal{L}^c_{X|Y}.
$$

(14)

Note that the degradedness assumption (5) implies the inclusion $\mathcal{L}_{X|Z} \subseteq \mathcal{L}_{X|Y}$. The coding scheme for the wiretap channel relies on the partition (14) and is summarized in Figure 1. The information is stored in the bits $u_i$, $i \in \mathcal{I}$. The bits in the coordinates in $\mathcal{R}$ are chosen randomly while the bits in $\mathcal{B}$ form a subset of the frozen bits.

Attainability of the rate (6) using this coding scheme is proved in the cited papers. An essential remark here is that the bits $u_i$, $i \in \mathcal{R}$ are randomly selected because fixing their values contradicts even the weak security constraint, let alone the stronger one.

Apart from transmitting the information, the coding scheme aims to convey the bits in $\tilde{\mathcal{R}}_2$ to Receiver 1 using the good indices of Receiver 1, at the same time preserving the security requirement. This is accomplished using the “chaining” construction proposed in [17] and shown in Fig. 2. As the figure suggests, the bits in $\tilde{\mathcal{R}}_2(j)$ contained in block $j$ are transmitted over the channel as a part of the message of block $j - 1$, for all $j = 2, \ldots, m$. This enables Receiver 1 to recover these bits reliably as a part of the successive decoding procedure for block $j$, which is performed similarly to (13). At the same time, because of the inclusion $\mathcal{J} \subseteq \mathcal{H}_{X|Z}$, Receiver 2 does not have the resources for their reliable decoding, which provides the desired security.

Figure 1. Block diagram of the coding scheme in [12]. Good bit channels and bad bit channels are as defined by (7).

We note that generally $\mathcal{H}^c_{X|Z} \not\subseteq \mathcal{L}_{X|Y}$, and even though the number of coordinates in $\mathcal{L}^c_{X|Y} \cap \mathcal{H}^c_{X|Z}$ behaves as $o(N)$, this constitutes an obstacle to achieving strong security. To bypass it, [17] uses a different partition of the coordinates, namely

$$
\tilde{\mathcal{J}} = \mathcal{L}_{X|Y} \cap \mathcal{H}_{X|Z} \\
\tilde{\mathcal{B}} = \mathcal{L}^c_{X|Y} \cap \mathcal{H}_{X|Z} \\
\tilde{\mathcal{R}}_1 = \mathcal{L}_{X|Y} \cap \mathcal{H}^c_{X|Z} \\
\tilde{\mathcal{R}}_2 = \mathcal{L}^c_{X|Y} \cap \mathcal{H}^c_{X|Z}.
$$

(15)

The term “chaining” was introduced later in [15].
The analysis of the transmission is performed based on $m$ blocks of $N$ bits as opposed to a single block. The seed for the transmission is provided by choosing $|\tilde{R}_2| = o(N)$ random bits which are shared with Receiver 1 (more on this below). In each of the blocks 1 to $m$, the bits indexed by the set $\tilde{E}$ are used to send the message. The bits in $\tilde{R}_1$ are selected randomly, and the bits $u_i, i \in \tilde{B}$ are frozen, i.e., assigned arbitrarily and shared in advance with Receiver 1 (they may be also known to Receiver 2 without compromising secrecy).

The assignment of bits in the set $\tilde{R}_2$ in block $j$ depends on the block index. In block 1 these bits are set equal to the message bits of the seed block. In block $j = 2, \ldots, m$ the bits indexed by the set $\tilde{R}_2$ are set to be equal to the bits in the set $\tilde{E}$ in block $j - 1$, representing the chaining procedure.

Having formed the sequence $u^N(j)$ in block $j = 1, \ldots, m$, the encoder passes it through the polarizing transform and transmits the sequence $x^N = u^N G_N$ over the channel. The only remaining problem is to convey to Receiver 1 the bits of $\tilde{R}_2$ of the first block. This is done by performing the seed transmission of a block which encodes the $|\tilde{R}_2|$ bits using some error correcting code of length $N$. This code is chosen to fulfill the reliability and security requirements, which is possible because the rate of the code for large $N$ can be made arbitrarily close to zero. The fact that the seed code needs to encode only a small number $o(N)$ of message bits follows from the degradedness assumption, which is therefore essential in this construction.

As shown in [17], this scheme satisfies both constraints (1) and (2) under the assumption that the channel to the eavesdropper is degraded with respect to the channel $W_1$. The rate of communication between the transmitter and Receiver 1 can be made arbitrarily close to the value $I(W_1) - I(W_2)$ since the assumption that $W_2$ is degraded with respect to $W_1$ ensures $|\tilde{R}_2| = o(N)$, i.e., there is no asymptotic loss in rate by removing the bits $\{u_i, i \in \tilde{E}\}$ from the message in order to support the strong security condition.

4. POLAR CODING FOR THE WIRETAP CHANNEL

In this section, we show that secrecy capacity for the wiretap channel given by Theorem 1 is achievable using polar codes. For this purpose, we consider the RVs $V, X, Y, Z$ as described by Theorem 1, i.e., we assume some fixed distributions $P_V, P_{X|V}$ and the conditional distributions $P_{Y|X} = W_1, P_{Z|X} = W_2$ that satisfy the Markov condition $V \rightarrow X \rightarrow Y, Z$ and maximize the expression in (4). Define the RV $T^N = V^N G_N$, where $V^N$ denotes $N$ independent realizations of $V$. The transformation $V^N \rightarrow T^N$ induces conditional distributions $P_{T_i|T^{i-1}}$ derived from the corresponding distributions of the RVs $V_i$. Define the sets $\mathcal{H}_V, \mathcal{L}_V, \mathcal{H}_{V|Y}, \mathcal{L}_{V|Y}$ as follows:

\[
\mathcal{H}_V = \{i \in [N] : Z(T_i|T^{i-1}) \geq 1 - \delta_N\}
\]

\[
\mathcal{L}_V = \{i \in [N] : Z(T_i|T^{i-1}) \leq \delta_N\}
\]

\[
\mathcal{H}_{V|Y} = \{i \in [N] : Z(T_i|T^{i-1}, Y^N) \geq 1 - \delta_N\}
\]

\[
\mathcal{L}_{V|Y} = \{i \in [N] : Z(T_i|T^{i-1}, Y^N) \leq \delta_N\}
\]
\[
L_{V|Y} = \{i \in [N] : Z(T_i|T_i-1, Y^N) \leq \delta_N \}
\]
and define the sets \(\mathcal{H}_{V|Z}, \mathcal{L}_{V|Z}\) analogously. The cardinalities of these sets satisfy \(\frac{1}{N}|\mathcal{H}_V| \to H(V), \frac{1}{N}|\mathcal{L}_V| \to 1 - H(V|Y), \frac{1}{N}|\mathcal{H}_{V|Z}| \to H(V|Z)\) as \(N \to \infty\).

Define a partition of \([N]\) into the following sets which will be used to describe the coding scheme:

\[
\begin{align*}
\mathcal{J} &= \mathcal{H}_V \cap \mathcal{L}_{V|Y} \cap \mathcal{H}_{V|Z} \\
\mathcal{B} &= \mathcal{H}_V \cap \mathcal{L}_{V|Y} \cap \mathcal{H}_{V|Z} \\
\mathcal{R}_1 &= \mathcal{H}_V \cap \mathcal{L}_{V|Y} \cap \mathcal{H}_{c|Z} \\
\mathcal{R}_2 &= \mathcal{H}_V \cap \mathcal{L}_{c|Y} \cap \mathcal{H}_{c|Z} \\
\mathcal{D} &= \mathcal{H}_c 
\end{align*}
\]

(16)

The partition of \([N]\) that thus arises is illustrated in Figure 3. It will be seen that the subsets \(\mathcal{J}, \mathcal{R}_1, \mathcal{R}_2, \mathcal{B}\) in our coding scheme play the role similar to that of the analogously denoted subsets in (15). Importantly, the cardinality of \(\mathcal{R}_2\) is not \(o(N)\) any more, which requires adjustments in the transmission scheme. Moreover, there is an extra randomness needed to determine the sequence to be transmitted, as will be seen in the encoding algorithm below.

**Figure 3.** Partition of \(N\) coordinates of the block for transmission over the wiretap channel \(W\); see (16). The highlighted part of the top block represents high-entropy coordinates for the distribution \(P_Y\). Similarly, in the middle block we highlight the high-entropy coordinates of the distribution \(P_{V|Z}\) and in the bottom block the low-entropy coordinates for the distribution \(P_{V|Y}\).

**Encoding:** We build on the chaining idea of [17], connecting multiple blocks in a cluster whose performance in transmission will attain the desired goals. The cluster consists of a seed block formed of \(|\mathcal{R}_2|\) random bits and a number, \(m\), of other blocks. The seed block consists of \(|\mathcal{R}_2|\) random bits. Even if the cardinality of the set \(\mathcal{R}_2\) constitutes a nonvanishing proportion of \([N]\), the rate of the seed \(|\mathcal{R}_2|/mN\) can be made arbitrarily small by choosing \(m\) sufficiently large. (For example, one can set \(m = N^\alpha\) for some \(\alpha > 0\), and let \(N \to \infty\).)

Let us describe the encoding and decoding procedures. The transmission is accomplished using \(m\) blocks of length \(N\) each and the seed block. The other blocks are constructed as follows. Every block \(t^n = t^N(j), j = 1, \ldots, m\) contains a group of almost deterministic bits, denoted by \(\mathcal{D}\) in Figure 4. The values of these bits are assigned according to the rule

\[
t_i = \arg\max_{i \in \{0,1\}} P_{T_i|T_{i-1}}(a_i|i-1), \quad i \in \mathcal{D}
\]

(17)
for all \(j = 1, \ldots, m\). The same rule (17) is used to assign values to frozen bits \(t_i, i \in B\) in each of the \(m\) blocks. The remaining sets are filled as follows. For block 1, the bits in the set \(R_2\) are assigned the value of the bits of the seed block, while for block \(j = 2, \ldots, m\) these bits are set to be equal to the bits in \(E(j - 1)\) of block \(j - 1\). The messages are stored in the bits indexed by \(N \setminus E\). The randomly chosen bits in \(E\) are written in the coordinates that are good for Receiver 1 and contained in the bad (high-entropy) set of Receiver 2. These bits are transmitted to Receiver 1 in block \(j\) and used for the decoding of the message contained in block \(j + 1\), for all \(j = 1, 2, \ldots, m - 1\). Finally, the bits in \(R_1\) are assigned randomly and uniformly for each of the \(m\) blocks. The diagram of the chaining construction for encoding is given in Figure 4.

Once the blocks \(t^N(j), j = 1, 2, \ldots, m\) are formed, we find \(m\) sequences \(v^N(j) = t^N(j)G_N\) by using the polarizing transform. Finally, given \(v^N\), the codeword to be sent over the wiretap channel will be chosen as \(x^N\) with probability \(P_{X^N|V^N}(x^N|v^N) = \prod_{i=1}^{N} P_{X|V}(x^i|v^i)\), where \(P_{X|V}\) is the conditional distribution induced by the joint distribution of the RVs \(V\) and \(X\). This logic is suggested by the proof of the capacity theorem, Theorem 1, which first considers “transmitting” the RV \(V^N\) to the receivers, and then choosing \(X^N\) so as to satisfy the Markov chain condition in the statement.

**Decoding:** Denote by \(E(0)\) the message sequence encoded in the seed block, and by \(E(j), j = 1, \ldots, m\) the corresponding sequences in the other blocks (see Fig. 4). Let \(y^N(1), \ldots, y^N(m)\) be the sequences that Receiver 1 observes on the output of the channel \(W_1\). The decoding rule is as follows:

\[
\hat{t}_i = \begin{cases} 
\arg\max_{t \in \{0,1\}} P_{T_i|T_i-1}(t|\hat{t}_{i-1}), & \text{if } i \in D \cup B \\
\arg\max_{t \in \{0,1\}} P_{T_i|T_i-1,Y^N}(t|\hat{t}_{i-1}, y^N), & \text{if } i \in J \cup R_1 \\
E_i(j-1), & \text{if } i \in R_2
\end{cases}
\]

(18)

where \(P_{T_i|T_i-1}\) and \(P_{T_i|T_i-1,Y^N}\) are the conditional distributions induced by the joint distribution of the RVs \(V^N\) and \(Y^N\) (this rule is applied to each of the blocks \(j = 1, \ldots, m\), and \(j\) is mostly omitted from the notation).

Let us show that the described scheme attains the secrecy capacity of \(W\). Namely, the following is true.

**Proposition 2.** For any \(\gamma > 0, \epsilon > 0\) and \(N \to \infty\) it is possible to choose \(m\) so that the transmission scheme described above attains the transmission rate \(R\) that is within \(\gamma\) of the secrecy capacity of \(W\) \([1]\) and the information leaked to Receiver 2 satisfies the strong secrecy condition \([2]\).

**Proof.** Assume that \(X, V, Y, Z\) are as given by Theorem 1. The rate of the proposed coding scheme is

\[
m(|J| - |E|) / mN = \frac{|\mathcal{H}_V \cap \mathcal{L}_{Y^Z} | - |\mathcal{H}_V \cap \mathcal{H}_{Y}^Z|}{N}
\]
which approaches $I(V; Y) - I(V; Z)$ as $N \to \infty$. According to Theorem 1, this is the target rate that we want to achieve for given $V$ and $X$ satisfying $V \to X \to Y, Z$ and $P_{Y|X} = W_1, P_{Z|X} = W_2$.

Introduce the following RVs: Let $M^m = (M_1, M_2, \ldots, M_m)$ form a sequence of message bits $\{t_i, t_i \in \mathcal{E}\}$ transmitted in blocks 1, \ldots, $m$, and let $Z^m = (Z^N(1), \ldots, Z^N(m))$ be a sequence of observations of Receiver 2 as a result of the transmission of the $m$ blocks.

Reliability: The claim of low error probability for Receiver 1 follows from the results of [10]. Namely, since the message bits are entirely contained in the set of good bits for channel $W_1$, the successive decoding procedure [18] will recover their values with error at most $mN2^{-N\beta}, \beta \in (0, 1/2)$. We conclude that the probability that Receiver 1 decodes the information bits correctly approaches 1 as $N$ tends to infinity.

Security: We will show that condition (2) is fulfilled for the sequence of $m$ blocks of transmission. The proof follows the argument in [17]. Writing $E_j$ instead of $\mathcal{E}(j)$, we have

$$I(M^m; Z^m) \leq I(M^m E_m; Z^m)$$

$$= I(M^m E_m; Z_m) + I(M^m E_m; Z^m - Z_m)$$

$$= I(M_m E_m; Z_m) + I(M^m E_m; Z^m - Z_m)$$

$$\leq I(M_m E_m; Z_m) + I(M^m E_m; Z^m - Z_m) + I(Z_m; Z^m - Z_m) + I(E_m - Z^m - Z_m | M^m E_m Z_m)$$

$$= I(M_m E_m; Z_m) + I(M^m (E_m - E_m) Z_m; Z^m - Z_m)$$

$$= I(M_m E_m; Z_m) + I(M^m E_m - E_m; Z^m - Z_m)$$

$$\leq \sum_{j=1}^{m} I(M_j E_j; Z_j)$$

where the second equality relies on the Markov condition $M^m - M_m E_m - Z_m$ and the next-to-last line on the condition $M_m E_m Z_m - M^m E_m - Z_m - Z_m$. Since the message bits are entirely within the set of indices $\mathcal{H}_{V|Z}$, the corresponding Bhattacharyya parameters are close to 1. Invoking the relation $Z(X|Y)^2 \leq H(X|Y)$ [3], we find for all $j$ that $I(M_j E_j; Z_j) \leq 2N2^{-N\beta}$. Hence, $I(M^m; Z^m) \leq 2mN2^{-N\beta}$, which implies that $\lim_{N \to \infty} I(M^m; Z^m) = 0$ as required.

We conclude that a secrecy rate of $I(V; Y) - I(V; Z)$ is achievable for any $V$ such that $V \to X \to Y, Z$ holds. Therefore, the secrecy capacity $C_s$ given by [4] is also achievable.

5. POLAR CODING FOR BROADCAST CHANNEL WITH CONFIDENTIAL MESSAGES

In this section we observe that ideas of the previous section together with some earlier works enable us to extend our code construction to a more general communication model introduced in [7].

5.1. The Model. Consider a pair of discrete memoryless channels with one transmitter $X$ and two receivers $Y, Z$. As before, let $W_1: X \to Y$ and $W_2: X \to Z$ denote the channels and let $\mathcal{X}$ and $\mathcal{Y}, \mathcal{Z}$ denote the input alphabet and the output alphabets. We assume that the system transmits three types of messages:

(i), a message $s_1 \in S_1$ from $X$ to $Y$ for which there are no secrecy requirements;
(ii), a message $s_2 \in S_2$ from $X$ to $Y$ which is secret from $Z$;
(iii), a message $t \in T$ from $X$ to $Y$ and $Z$, called the "common message".

Following [7], we call this communication scheme a broadcast channel with confidential messages (BCC).

As before, a block encoder for the BCC is a mapping $f: S_1 \times S_2 \times T \to \mathcal{X}^N$. A stochastic version of the encoder is a probability matrix $f(x^N|s_1, s_2, t)$ with columns indexed by $x^N \in \mathcal{X}^N$ and rows by the triples $(s_1, s_2, t)$. Given such a triple, the stochastic encoder samples from the conditional probability distribution on $\mathcal{X}^N$. In accordance with the problem statement, there are two decoders: The decoder of Receiver 1 is defined by a mapping $\phi: Y^N \to S_1 \times S_2 \times T$ and the decoder of Receiver 2 is a mapping $\psi: Z^N \to T$. 
Denote the rate of the common message \( t \) by \( R_o \), and denote the rates of the secret and non-secret messages to \( Y \) by \( R_s \) and \( R_1 \), respectively. The analogs of the standard definitions, Defns. [1,1] 1 [2] in this case look as follows.

**Definition 5.1.** The encoder-decoder mappings \((f, \phi, \psi)\) give rise to \((N, \epsilon)\) transmission over the BCC if for every \( s_1 \in \mathcal{S}_1 \), \( s_2 \in \mathcal{S}_2 \), \( t \in \mathcal{T} \), decoder \( \phi \) outputs the transmitted triple \((s_1, s_2, t)\) and decoder \( \psi \) outputs the message \( t \) with probability greater than \( 1 - \epsilon \), i.e.,
\[
\sum_{x^N \in \mathcal{X}^N} f(x^N | s_1, s_2, t) P_{Y|X}^N (\phi(\gamma^N)) = (s_1, s_2, t) | x^N) \geq 1 - \epsilon
\]
\[
\sum_{x^N \in \mathcal{X}^N} f(x^N | s_1, s_2, t) P_{Z|X}^N (\psi(\zeta^N)) = t | x^N) \geq 1 - \epsilon.
\]

**Definition 5.2.** \((R_1, R_s, R_0)\) is an achievable rate triple for the BCC if there exists a sequence of message sets \( \mathcal{S}_1 N, \mathcal{S}_2 N, \mathcal{T}_N \) and encoder-decoder triples \((f_N, \phi_N, \psi_N)\) giving rise to \((N, \epsilon_N)\) transmission with \( \epsilon_N \to 0 \), such that
\[
\lim_{N \to \infty} \frac{1}{N} \log |\mathcal{S}_1 N| = R_1
\]
\[
\lim_{N \to \infty} \frac{1}{N} \log |\mathcal{S}_2 N| = R_s
\]
\[
\lim_{N \to \infty} \frac{1}{N} \log |\mathcal{T}_N| = R_0
\]
\[
\lim_{N \to \infty} I(S_{2N}; Z^N) = 0.
\]
where \( S_{2N} \) is the random variable that corresponds to the “secret message.”

Note that our definition takes into account the formulation in [8] and is slightly different from the one in [7] (where we write \((R_1, R_s, R_0), \mathcal{T}\) has \((R_1 + R_s, R_s, R_0)\).)

The following theorem gives the achievable rate region for the triple \((R_1, R_s, R_0)\).

**Theorem 3.** [7, p. 414] The capacity region of the BCC \( \mathcal{R} \) consists of those triples of nonnegative numbers \((R_1, R_s, R_0)\) that satisfy, for some RVs \( U \to V \to X \to Y, Z \) with \( P_{Y|X} = W_1 \) and \( P_{Z|X} = W_2 \), the inequalities
\[
R_0 \leq \min[I(U; Y), I(U; Z)]
\]
\[
R_s \leq I(V; Y | U) - I(V; Z | U)
\]
\[
R_0 + R_1 + R_s \leq I(V; Y | U) + \min[I(U; Y), I(U; Z)].
\]
Moreover, it may be assumed that \( V = (U, V') \) and the range sizes of \( U \) and \( V' \) are at most \(|X| + 3\) and \(|X| + 1\).

One can define the secrecy capacity \( C_s \) in terms of the capacity region of the BCC, and then recover Theorem [1] as a particular case of Theorem [3].

5.2. **Polar Coding for the Csiszár-Körner Region.** In this section, we aim to show that the capacity region of the BCC can be achieved using polar codes. In the first two steps we design a scheme that achieves the rate pairs \((R_0, R_s)\) in (21)-(22), and in the last step we show that for any such pair \((R_0, R_s)\) any rate value
\[
R_1 \leq I(V; Y | U) + \min[I(U; Y), I(U; Z)] - R_0 - R_s
\]
is also achievable. Finally, the security condition in (20) will be shown in Proposition [4] below.

The overall encoding scheme is stochastic and assumes some fixed joint distribution of the RVs \( U, V, X, Y, Z \) such that the constraints of Theorem [3] are satisfied. Since the results below are valid for any such distribution, this will enable us to claim achievability of the rate region in this theorem. The encoder is formed of two stages performed in succession. At the outcome of the first stage, which deals with the common message \( s_2 \), the encoder computes a sequence of \( m \) blocks of \( N \) bits denoted below by \( q^N(j), j = 1, \ldots, m \). These blocks are used in the
second stage to construct the data encoding that is going to be sent to both receivers. Namely, it will be seen that the transformed blocks $w^N = q^N G_N$ can at the same time encode the common message to both receivers and also encode side information for Receiver 1 to ensure reliable transmission of the confidential message. The actual sequences to be transmitted are computed in the second stage based on the sequences $w^N(j)$. This is done by first constructing sequences $t^N(j)$ using the ideas developed in Sect. 4 and by using a stochastic mapping of these sequences on the codewords $x^N(j)$. Upon transmitting, these codewords are received by Receiver 1 as $y^N(j)$ and by Receiver 2 as $z^N(j)$. We will argue that the receivers can independently perform decoding procedures that recover the three desired types of messages reliably (and when appropriate, also securely).

5.2.1. The common-message encoding. The proof of the fact that any $R_0$ satisfying (21) is achievable follows from the polar coding scheme for the superposition region given in [15]. Given the RVs $U \rightarrow V \rightarrow X \rightarrow Y, Z$ with $P_{Y|X} = W_1$ and $P_{Z|X} = W_2$, let $U^N, V^N, X^N, Y^N, Z^N$ be $N$ repetitions of the RVs $U, V, X, Y, Z$. Set

$$Q^N = U^N G_N,$$

where $G_N$ is Arıkan’s transform. As before, lowercase letters denote realizations of these RVs.

Define the sets $\mathcal{H}_U, \mathcal{L}_{U|Y}, \mathcal{L}_{U|Z}$ as follows:

$$\mathcal{H}_U = \{ i \in [N] : Z(Q_i | Q^{i-1}) \geq 1 - \delta_N \},$$

$$\mathcal{L}_{U|Y} = \{ i \in [N] : Z(Q_i | Q^{i-1}, Y^N) \leq \delta_N \},$$

$$\mathcal{L}_{U|Z} = \{ i \in [N] : Z(Q_i | Q^{i-1}, Z^N) \leq \delta_N \}.$$

The cardinalities of these sets, normalized by $N$, approach respectively $H(U), 1 - H(U|Y), 1 - H(U|Z)$ as $N \rightarrow \infty$.

Now, we observe that for Receiver 1 to recover $q^N$ correctly, the indices of the information bits should be a subset of $\mathcal{J}_u^{(1)} = \mathcal{H}_U \cap \mathcal{L}_{U|Y}$. Similarly, for Receiver 2 to recover the sequence $q^N$ correctly, the information bits should be placed in only those positions of $q^N$ that are indexed by the set $\mathcal{J}_u^{(2)} = \mathcal{H}_U \cap \mathcal{L}_{U|Z}$. Therefore, choosing the indices of information bits as $\mathcal{J}_u = \mathcal{J}_u^{(1)} \cap \mathcal{J}_u^{(2)}$ ensures that the message embedded into $q^N$ will be decoded correctly by both receivers. In this case, the rate of the common message is $R_0 = |\mathcal{J}_u^{(1)} \cap \mathcal{J}_u^{(2)}| / N$. Given that

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{J}_u^{(1)}| = I(U; Y)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{J}_u^{(2)}| = I(U; Z)$$

we conclude the common message rate $R_0$ attains the value $\min[I(U; Y); I(U; Z)]$ only if either $\mathcal{J}_u^{(2)} \subseteq \mathcal{J}_u^{(1)}$ or $\mathcal{J}_u^{(1)} \subseteq \mathcal{J}_u^{(2)}$ holds starting from some $N$. However, this does not have to be the case. To overcome this problem, [15] proposed the following coding scheme. Define the sets

$$D^{(1)} = \mathcal{J}_u^{(1)} \setminus \mathcal{J}_u^{(2)}$$

$$D^{(2)} = \mathcal{J}_u^{(2)} \setminus \mathcal{J}_u^{(1)}.$$

Without loss of generality, assume that $I(U; Y) \leq I(U; Z)$, which implies that $|D^{(2)}| \geq |D^{(1)}|$ starting with some $N$. To describe the encoding procedure, consider $m$ blocks of $N$ coordinates each. In block 1, we use the positions indexed by $D^{(1)}$ to store message bits and assign the bits indexed by $D^{(2)}$ to some fixed values that are available to Receiver 1. In block $j$, $j = 2, \ldots, m - 1$, we again use the positions indexed by $D^{(1)}$ to store message bits and copy the part of Block $j - 1$ indexed by the coordinates in $D^{(1)}$ into the positions indexed by a subset of coordinates $E^{(2)} \subset D^{(2)}$ in block $j$ (thus $|E^{(2)}| = |D^{(1)}|$). Fill the remaining $|D^{(2)} \setminus E^{(2)}|$ bits in each block $j \in \{2, \ldots, m - 1\}$ with random and independent bits and communicate them to Receiver 1. These bits can be the same for each block as long as they are independent and uniform within the same block, so this part of the scheme has negligible impact on the overall rate. In the final block $m$, we assign the bits indexed by $D^{(1)}$ to
some fixed values that are available to Receiver 2 and copy the bits in $D(1)(m - 1)$ to the positions in $E(2)(m)$. The remaining $|D(2)\setminus E(2)|$ coordinates in block $m$ are filled with random bit values. The block diagram of the described coding scheme is shown by Figure 5.

![Diagram of coding scheme](image)

**Figure 5.** Encoding for the common-message case: The structure of the blocks $q^N(j)$, $j = 1, \ldots, m$

5.2.2. The secret-message encoding. In this section we describe the construction of sequences $x^N$ that are sent by transmitter $X$. The construction relies on the sequences $u^N(j)$ constructed by $X$ in the first stage. These sequences can be thought of as side information that enables Receiver 1 to reconstruct the secret message.

The transmission scheme we propose to achieve the rate $R_s$ that satisfies (22), is very similar to the scheme described for the wiretap channel problem in Section 4. Our solution consists of choosing the indices of information bits and random bits appropriately and using a chaining scheme quite similar to the one shown in Figure 4.

Let $U^N, V^N, X^N, Y^N, Z^N$ be as defined in Section 5.2.1, and let $T^N = V^N G_N$. Viewing $U$ as side information about $V$, we define the sets

$$
\mathcal{H}_{V|U} = \{ i \in [N] : Z(T_i|T^{i-1}, U^N) \geq 1 - \delta_N \}
$$

$$
\mathcal{L}_{V|U,Y} = \{ i \in [N] : Z(T_i|T^{i-1}, U^N, Y^N) \leq \delta_N \}
$$

$$
\mathcal{K}_{V|U,Z} = \{ i \in [N] : Z(T_i|T^{i-1}, U^N, Z^N) \geq 1 - \delta_N \}
$$

whose cardinalities, normalized by $N$, approach respectively the values $H(V|U), 1 - H(V|U, Y), H(V|U, Z)$ as $N \to \infty$.

The intuition behind the construction presented below can be described as follows. First, note that the coordinates of $T^N$ indexed by

$$
\mathcal{J}^{(1)} = \mathcal{H}_{V|U} \cap \mathcal{L}_{V|U,Y}
$$

(26)
can be decoded by Receiver 1, and so they can be used to send more data in addition to the common message. Of these bits, the part indexed by $(\mathcal{H}_{V|U} \cap \mathcal{L}_{V|U,Y}) \setminus (\mathcal{H}_{V|U} \cap \mathcal{K}_{V|U,Z})$ can be used to transmit the confidential message. Then, given that $\frac{1}{N}|(\mathcal{H}_{V|U} \cap \mathcal{L}_{V|U,Y}) \cap \mathcal{J}^{(1)}| \to I(V; Y|U)$ and $\frac{1}{N}|(\mathcal{J}^{(1)} \cap \mathcal{K}_{V|U,Z})| \to I(V; Z|U)$, we obtain

$$
\lim_{N \to \infty} \frac{1}{N}|(\mathcal{H}_{V|U} \cap \mathcal{L}_{V|U,Y}) \setminus (\mathcal{H}_{V|U} \cap \mathcal{K}_{V|U,Z})| \geq I(V; Y|U) - I(V; Z|U)
$$
This implies that the proposed scheme transmits the secret message at rates arbitrarily close to the rate given by (22) (provided that it also satisfies the strong security condition).

Building on this observation, we proceed to describe the coding scheme, adding some details that make the secrecy part work. Define the sets \( J, \mathcal{R}_1, \mathcal{R}_2, \mathcal{B}, \mathcal{D} \) partition \([N]\). This partition is basically the same as in (16) (see also Figures 3 and 4) except for the fact that the high- and low-entropy subsets rely on entropy quantities that are additionally conditioned on \( U \).

The transmission scheme that we propose is formed of multiple blocks joined in clusters of \( m \) blocks. Similarly to the wiretap coding scheme, there is a seed block shared between the transmitter and Receiver 1. The seed block consists of \([\mathcal{R}_2]\) random bits. Even if the set \([\mathcal{R}_2] = \frac{1}{mN} \) can be made arbitrarily small by choosing \( m \) sufficiently large. (For example, one can set \( m = N^\alpha \) for some \( \alpha > 0 \), and let \( N \) tend to infinity.)

The encoding procedure is as follows. Our aim is to construct \( m \) blocks \( t_N(j) \) which will be used to form the transmitted sequences \( x_N(j), j = 1, \ldots, m \). Apart from the seed block, all the other blocks \( t_N \) contain a group of almost deterministic bits, denoted by \( \mathcal{D} \) in (27). For Block \( j, j = 1, \ldots, m \), the values of these bits are assigned according to the rule

\[
t_i = \text{argmax}_{t \in \{0,1\}} P_{T_i|T_{i-1},U,N}(t|t^{i-1}, u^N(j)).
\]

The same rule is used for the set of frozen bits \( \mathcal{B} \) in each of the \( m \) blocks.

The remaining subsets of coordinates in are filled as follows. For block 1, the bits in the set \( \mathcal{R}_2 \) are assigned the values of the bits of the seed block, while for block \( j, j = 2, \ldots, m \) these bits are set to be equal to the bits in \( \mathcal{E} \) of block \( j-1 \). The messages are stored in the bits indexed by \( J \setminus \mathcal{E} \) which are good for Receiver 1 and contained in the bad (high-entropy) set of Receiver 2. The indices in the subset \( \mathcal{E} \) are still good for Receiver 1 bad for Receiver 2. Nevertheless, they are filled with random bits that are used for decoding by Receiver 1 in the same way as done in Sect. 4. Finally, the bits in \( \mathcal{R}_1 \) are assigned randomly and uniformly for each of the \( m \) blocks. We once again refer to Fig. 4 which illustrates the described processing. Once the blocks \( t_N(j), j = 1, \ldots, m \) are formed, we compute \( m \) sequences \( v^N(j) = t_N(j)G_N \) by using the polarizing transform.

Finally, the codewords to be sent by the transmitter are computed as follows. The codeword \( x_N(j), j = 1, \ldots, m \) is sampled from \( \mathcal{X}^N \) according to the distribution \( P_{X^N|Y^N}(x^N|v^N) = \prod_{i=1}^N P_{X|Y}(x^i|v^i) \), where \( P_{X|Y} \) is the conditional distribution induced by the joint distribution of the RVs \( V \) and \( X \).

5.2.3. Decoding of the common message and the secret message. Assume that the transmitted sequence \( x_N^* \) is received as \( y^N \) by Receiver 1 and as \( z^N \) by Receiver 2. Importantly, by our construction these sequences follow the conditional distributions \( P_{Y_i|X} \) and \( P_{Z_i|X} \) given by the channels \( W_1 \) and \( W_2 \). We describe the decoding procedures by Receivers 1 and 2. Initially they perform similar operations aimed at recovering the common message. Once this is accomplished, Receiver 1 performs additional decoding to recover the secret message.

We begin with the common-message part. In accordance with (25), Receivers 1 and 2 decode the blocks \( q_N^*(j), j = 1, \ldots, m \) relying on an iterative procedure. As the construction suggests, Receiver 1 decodes in the forward direction, starting with block 1 and ending with block \( m \), and Receiver 2 decodes backwards, starting

\[\text{We again use the same notation as in (16); since the earlier notation is used only in Section 4, this should not cause confusion.}\]
with block \( m \) and ending with block 1. Let
\[
D(j) = \{ q_i, i \in \mathcal{D}(1) \}, j = 1, \ldots, m - 1
\]
\[
C(j) = \{ q_i, i \in \mathcal{E}(2) \}, j = 2, \ldots, m
\]
denote the subblocks \( \mathcal{D}, \mathcal{E} \) of the corresponding blocks (see Fig. 5).

The processing by Receiver 1 is as follows. For block 1, it computes
\[
\hat{q}_i = \begin{cases} 
q_i, & \text{if } i \in \mathcal{D}(2) \\
\arg\max_{q \in \{0,1\}} P_{Q_i|Q^{i-1},Y^N}(q|q^{i-1},y^N), & \text{if } i \in \mathcal{J}_u \cup \mathcal{D}(1)
\end{cases}
\]
(28)

For the remaining blocks \( j = 2, \ldots, m \) Receiver 1 computes the vector \( \hat{q}_i(j), i = 1, \ldots, N \) as follows:
\[
\hat{q}_i(j) = \begin{cases} 
D_i(j - 1), & \text{if } i \in \mathcal{E}(2) \\
q_i(j), & \text{if } i \in \mathcal{D}(2) \setminus \mathcal{E}(2) \\
\arg\max_{q \in \{0,1\}} P_{Q_i|Q^{i-1},Y^N}(q|q^{i-1}(j),y^N(j)), & \text{if } i \in \mathcal{J}_u \cup \mathcal{D}(1)
\end{cases}
\]
(29)

The processing by Receiver 2 is quite similar except that it starts with block \( m \) and advances “backwards” for \( j = m - 1, m - 2, \ldots, 1 \). For block \( m \) the rule is as follows:
\[
\hat{q}_i = \begin{cases} 
q_i, & \text{if } i \in \mathcal{D}(1) \\
\arg\max_{q \in \{0,1\}} P_{Q_i|Q^{i-1},Z^N}(q|q^{i-1},z^N), & \text{if } i \in \mathcal{J}_u \cup \mathcal{D}(2)
\end{cases}
\]
(30)

For blocks \( j = m - 1, m - 2, \ldots, 1 \), Receiver 2 computes its estimates of the vector \( q^N(j) \) as follows:
\[
\hat{q}_i(j) = \begin{cases} 
C_i(j + 1), & \text{if } i \in \mathcal{D}(1) \\
\arg\max_{q \in \{0,1\}} P_{Q_i|Q^{i-1},Z^N}(q|q^{i-1}(j),z^N(j)), & \text{if } i \in \mathcal{J}_u \cup \mathcal{D}(2)
\end{cases}
\]
(31)

(in (28)-(31) the notation is somewhat abbreviated to keep the formulas compact, e.g., no reference is made to the index of the receiver, and the block index \( j \) is often omitted).

The processing described above in (28)-(29) yields the sequences \( \hat{u}^N(j) = q^N(j)G_N, j = 1, \ldots, m \) which are used by Receiver 1 to recover the secret messages. Denote by \( \mathcal{E}(0) \) the message sequence encoded in the seed block, and let \( \mathcal{E}(j), j = 1, \ldots, m \) be the subblock of block \( j \) indexed by the set \( \mathcal{E} \). For \( j = 1, \ldots, m \), the decoding rule is as follows:
\[
\hat{t}_i(j) = \begin{cases} 
\arg\max_{t \in \{0,1\}} P_{T_i|T_i^{j-1},U^N}(t|\hat{u}^{i-1}(j),\hat{u}^N(j)), & \text{if } i \in \mathcal{D} \cup \mathcal{B} \\
\arg\max_{t \in \{0,1\}} P_{T_i|T_i^{j-1},U^N,Y^N}(t|\hat{u}^{i-1}(j),\hat{u}^N(j),y^N(j)), & \text{if } i \in \mathcal{J} \cup \mathcal{R}_1 \\
\mathcal{E}_i(j - 1), & \text{if } i \in \mathcal{R}_2
\end{cases}
\]
(32)

where \( P_{T_i|T_i^{j-1},U^N} \) and \( P_{T_i|T_i^{j-1},U^N,Y^N} \) are the conditional distributions induced by the joint distribution of the RVs \( U^N, V^N \) and \( Y^N \) (the notation is again abbreviated similarly to (18)).

5.2.4. Achievability of the rate region (21)-(23). The rate of the common message achieved by the construction in Sect. 5.2.2 is equal to
\[
R_0 = \frac{1}{mN} \left[ m|\mathcal{J}_u| + (m - 1)|\mathcal{D}(1)| \right].
\]
As \( N \) increases, we obtain
\[
\frac{m - 1}{m} I(U;Y) \leq R_0 \leq I(U;Y).
\]
For sufficiently large $m$ this quantity is arbitrarily close to the common-message rate value given in (21). Note that we have assumed that $I(U; Z) \geq I(U; Y)$; to handle the opposite case is suffices to interchange the roles of the pieces $D^{(1)}$ and $D^{(2)}$ in the common-message encoding and decoding procedures.

As shown in [15], both receivers can decode the common message correctly with probability of error at most $mN2^{-\beta} \beta \in (0, 1/2)$. This follows because in this stage, Receivers 1 and 2 aim only at decoding the bits corresponding to the index sets $D^{(1)} \cup J_u = H_U \cap L_{U|Y}$ and $D^{(2)} \cup J_u = H_U \cap L_{U|Z}$, respectively. That these bits can be recovered in a successive decoding procedure follows from the basic results on polar codes, [2, 4].

For the secret-message part of the communication the properties of the scheme are characterized by the following proposition.

**Proposition 4.** For any $\gamma > 0$, $\epsilon > 0$ and $N \to \infty$ it is possible to choose $m$ so that the transmission scheme described above attains a secrecy rate $R_s$ such that $R_s \geq I(V; Y|U) - I(V; Z|U) - \gamma$ and the information leaked to Receiver 2 satisfies the strong secrecy condition in Definition 5.2

**Proof.** Assume that $X, V, Y, Z$ are as given by Theorem 3. The rate of proposed coding scheme is

$$m(|J| - |E|) = \frac{m|H_V|U \cap L_{V|U,Y} - |H_V|U \cap H_{V|U,Z}|}{N}$$

which converges to $I(V; Y|U) - I(V; Z|U)$ as $N$ goes to infinity. Recalling Theorem 3, we note that this is the target value of the rate $R_s$ for given $U, V$ and $X$ satisfying $U \to V \to X \to Y, Z$ and $P_{Y|X} = W_1, P_{Z|X} = W_2$.

Introduce the following RVs: Let $M^m = (M_1, M_2, \ldots, M_m)$ be a sequence of message bits $\{t_i, i \in J \setminus E\}$ sent in blocks $1, \ldots, m$, and let $Z^m = (Z^N(1), \ldots, Z^N(m))$ be the RVs that represent the random observations of Receiver 2 upon transmitting $m$ blocks $X^N(j)$.

**Reliability:** The claim of low error probability for Receiver 1 follows from the results of [10]. Namely, since the message bits are entirely contained in the set of $L_{V|U,Y}$, the successive decoding procedure (32) will recover their values with error at most $mN2^{-\beta}, \beta \in (0, 1/2)$. We conclude that the probability that Receiver 1 decodes the information bits correctly approaches 1 as $N$ goes to infinity.

**Security:** We will show that condition (2) is fulfilled for the sequence of $m$ blocks of transmission. Note that Receiver 2 observes not only a realization of $Z^m$, but also estimates RVs $U^m = (U^N(1), \ldots, U^N(m))$ through procedure (30)-(31). For this reason the strong security condition to be proved takes the form

$$\lim_{N \to \infty} I(M^m; U^m Z^m) = 0. \quad (33)$$

The proof of (33) relies on the following lemma.

**Lemma 5.** Let $T[E] = \{t_i, i \in E\}$ and $T[J \setminus E] = \{t_i, i \in J \setminus E\}$, then

$$I(T[J \setminus E]; T[E] U^N Z^N) \leq 2N2^{-\beta} \quad (34)$$

**Proof.** From (27) we have the inclusion $J \subseteq H_{V|U,Z}$. Let us label the indices in $J$ as $a_1, a_2, \ldots, a_{|J|}$. Using the inequality $Z(X|Y)^2 \leq H(X|Y) \quad (3)$, we obtain

$$H(T_{a_i}|T^{a_i-1} U^N Z^N) \geq 1 - 2\delta_N \quad (35)$$

for all $i = 1, 2, \ldots, |J|$. Observe that (35) implies

$$H(T_{a_i}|T_{a_1}T_{a_2} \ldots T_{a_{i-1}} U^N Z^N) \geq 1 - 2\delta_N \quad (36)$$

for all $i$. Inequality (36) is valid for any subset of indices in $J$ including in particular the subset $E$, say $|E| = \{a_1, a_2, \ldots, a_{|E|}\}$, which implies

$$H(T[J \setminus E]T[E] U^N Z^N) = \sum_{i=|E|+1}^{|J|} H(T_{a_i}|T_{a_1}T_{a_2} \ldots T_{a_{i-1}} U^N Z^N) \geq |J \setminus E|(1 - 2\delta_N).$$
This completes the proof of (34).

The remaining part of the proof is similar to the calculation in (19). Again writing \( E_j \) instead of \( E(j) \), we have

\[
I(M^m; U^m Z^m) \leq I(M^m E_m; U^m Z^m)
\]

\[
= I(M^m E_m; U_m Z_m) + I(M^m E_m; U^m Z^m \vert U_m Z_m)
\]

\[
= I(M^m E_m; U_m Z_m) + I(M^m E_m; U^m Z^m \vert U_m Z_m)
\]

\[
\leq I(M^m E_m; U_m Z_m) + I(M^m E_m; U^m Z^m \vert U_m Z_m)
\]

\[
= I(M^m E_m; U_m Z_m) + I(M^m; U^m Z^m \vert U_m Z_m)
\]

\[
\leq \sum_{j=1}^m I(M_j E_j; U_j Z_j)
\]

where (37) relies on the Markov condition \( M^{m-1} - M^m E_m - U_m Z_m \), inequality (38) is due to the condition \( M^{m-1} - U^m Z^m \) and the fact that \( M^m E_m \) is independent of \( U^m Z_m \), and (39) follows from the independence of \( M^m E_m \) and the pair \( M^{m-1}, U^m Z^m \).

Since the message bits are entirely within the set of indices \( \mathcal{H}_{U,Z} \), the corresponding Bhattacharyya parameters are close to 1. Invoking the relation \( Z(X \mid Y)^2 \leq H(X \mid Y) \), we find for all \( j \) that \( I(M_j E_j; U_j Z_j) \leq 2N2^{-N^2} \). Hence, from (40) we obtain \( I(M^m; U^m Z^m) \leq 2N2^{-N^2} \), which in turn implies (33). This completes the proof.

Finally, let us show that the “additional-message” rate \( R_1 \) as given by (23) is also achievable. We have seen that any rate pair \( (R_0, R_s) \) satisfying (21)-(22) is achievable in the system. Moreover, observe that Receiver 1 decodes correctly messages at the rate of \( \min[I(U; Y), I(U; Z)] \) according to (28), (29), and additionally decodes messages at the rate of \( I(V; Y \mid U) \) owing to the part of the encoding \( \{t_i, i \in \mathcal{H}(U)\} \) given by (26). Since these two groups of information bits can be decoded simultaneously by Receiver 1, we conclude that it is possible to communicate to Receiver 1 an additional message at rate \( R_1 \).

6. Conclusion

In this paper, we have considered the wiretap channel problem (23) and the broadcast channel with confidential messages problem (7). For the wiretap channel problem, we have shown that the secrecy capacity \( C_s \) is achievable by polar codes. The solution we propose for the wiretap channel problem is different from the previous works in the sense that it does not have any assumptions on the wiretap channel and achieves the capacity of the channel. We also show that it is possible to build on this solution by adding a second layer of encoding which enables one to attain the capacity region of the broadcast channel with confidential messages model of Csiszár and Körner.

REFERENCES

[1] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, M. Skoglund, Nested polar codes for wiretap and relay channels, IEEE Communication Letters, no. 14, pp. 752–754, 2010.
[2] E. Arikan, Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, IEEE Trans. Inform. Theory 55 (2009), no. 7, 3051–3073.
[3] E. Arikan, Source polarization, Proc. IEEE Int. Symposium on Information Theory, Austin, TX, June 2010, pp. 899–903.
[4] E. Arikan and E. Telatar, On the rate of channel polarization, Proc. IEEE Int. Sympos. Inform. Theory, Seoul, Korea, 2009, pp. 1493–1495.
[5] M. Bellare, S. Tessaro, A. Vardy, Semantic security for the wiretap channel in Advances in Cryptology CRYPTO 2012, Lecture Notes in Computer Science, Vol. 7417, pp. 294–311.
[6] M. Cheraghchi, F. Didier, and A. Shokrollahi, Invertible extractors and wiretap protocols, IEEE Trans. Inform. Theory, 58 (2012), no. 2, 1254–1274.
[7] I. Csiszár and J. Körner, Broadcast channels with confidential messages, IEEE Trans. Inform. Theory, 24 (1978), no. 3, 339–348.
[8] I. Csiszár and J. Körner, Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.
[9] E. Hof and S. Shamai, *Secrecy-achieving polar-coding*, Proc. IEEE Information Theory Workshop, Dublin, Ireland, Sept. 2010, pp. 1–5; also arXiv:1005.2759.
[10] J. Honda and H. Yamamoto, *Polar coding without alphabet extension for asymmetric models*, IEEE Trans. Inform. Theory 59 (2013), no. 12, 7829–7838.
[11] O. O. Koyluoglu and H. El Gamal, *Polar coding for secure transmission and key agreement*, IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 2698–2703, 2010.
[12] H. Mahdavifar and A. Vardy, *Achieving the secrecy capacity of wiretap channels using polar codes*, IEEE Trans. Inform. Theory, 57 (2011), no. 10, 6428–6443.
[13] U. M. Maurer, *The strong secret key rate of discrete random triples*, in Communication and Cryptography–Two Sides of One Tapestry, R. Blahut. Ed. et al., Boston, MA, 1994, pp. 271–285.
[14] 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.
[15] M. Mondelli, H. Hassani, I. Sason, and R. Urbanke, *Achieving the superposition and binning regions for broadcast channels using polar codes*, arXiv:1401.6060.
[16] J. M. Renes and M. M. Wilde, *Polar codes for private and quantum communication over arbitrary channels*, IEEE Trans. Inform. Theory, 60 (2014), no. 6, 3090–3103.
[17] E. Şaşoğlu and A. Vardy, *A new polar coding scheme for strong security on wiretap channels*, Proc. IEEE Int. Symposium on Information Theory, Istanbul, Turkey, July 2013, pp.1117–1121.
[18] A. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S.W. McLaughlin. *Strong security for erasure wiretap channels*, in Proc. IEEE Information Theory Workshop, Dublin, Ireland, Sep. 2010.
[19] D. Sutter, J. M. Renes, R. Renner, *Efficient one-way secret-key agreement and private channel coding via polarization*, arXiv:1304.3658v1.
[20] A. Thangaraj, S. Dihidar, A.R. Calderbank, S.W. McLaughlin, and J. Merolla, *Applications of LDPC codes to the wiretap channel*, IEEE Trans. Inform. Theory, 53 (2007), no. 8, 2933–2945.
[21] V. K. Wei, *Generalized Hamming weights of linear codes*, IEEE Trans. Inform. Theory, 37 (1991), no. 5, 1412–1418.
[22] M. M. Wilde and J. M. Renes, *Polar codes for private classical communication*, Proc. 2012 IEEE Int. Sympos. Inform. Theory Appl. (ISITA), pp. 745–749.
[23] A. D. Wyner, *The wire-tap channel*, Bell Syst. Tech. J., vol. 54, no. 8, pp. 1355–1387, Oct. 1975.