Channel Upgrading for Semantically-Secure Encryption on Wiretap Channels

Ido Tal
Department of Electrical Engineering
Techion, Haifa 32000, Israel
idotal@ieee.org

Alexander Vardy
University of California San Diego
La Jolla, CA 92093, USA
avardy@ucsd.edu

***Abstract*—Bellare and Tessaro recently introduced a new coding scheme, based on cryptographic principles, that guarantees strong security for a wide range of symmetric wiretap channels. This scheme has numerous advantages over alternative constructions, including constructions based on polar codes. However, the BT coding scheme achieves secrecy capacity only under a certain restrictive condition. Specifically, let V be the main channel (from Alice to Bob) and let W be wiretap channel (from Alice to Eve). Suppose that W has a finite output alphabet $\mathcal{Y}$, and let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the input and output of W, respectively. Then the rate of the BT scheme is upper-bounded by $\mathsf{capacity}(V) - \Psi(W)$, where**

$$\Psi(W) \stackrel{\text{def}}{=} \log_2 |\mathcal{Y}| - H(\boldsymbol{Y}|\boldsymbol{X})$$

For symmetric channels, it clear that $\Psi(W)$ equals the capacity of W if and only if uniform input to W produces uniform output. Unfortunately, few symmetric DMCs satisfy this condition.

In this paper, we show how the Bellare-Tessaro coding scheme can be extended to achieve secrecy capacity in the case where W is an *arbitrary* symmetric DMC. To this end, we solve the following problem. Given W and $\varepsilon > 0$, we construct another channel Q such that W is degraded with respect to Q while the difference between $\Psi(Q)$ and the capacity of W is at most ε.

I. Introduction

The wiretap setting was introduced by Wyner [1]. Figuratively, it consists of three players, Alice, Bob, and Eve. Alice wishes to send Bob confidential information, and this can be done by using a channel C. However, everything that Alice inputs to the channel C is also fed into the channel $W : \mathcal{X} \to \mathcal{Y}$, from Alice to Eve. Eve plays the role of the eavesdropper. Thus, in brief, Alice must code her information so that Bob can decode it, but Eve cannot. In order to do so, Alice has access to a private source of random bits.

Recently, Bellare, Tessaro, and Vardy introduced [**?**] a coding scheme by which the above can be attained. The distinction of the BT scheme is that it guarantees *semantic security*, a stronger form of security than was originally proposed by Wyner in [1]. The notion of semantic security is described fully in [**?**]. In brief, we say that a method achieves σ bits of semantic security if the following holds. For all distributions on the message set of Alice, for all functions f of the message, and for all strategies Eve might employ, the probability of Eve guessing the value of f correctly increases by no more than $2^{-\sigma}$ between the case in which Eve does not have access to the output of W and the case that she does. That is, having access to W hardly helps Eve, for sufficiently large σ.

The method by which Alice achieves this security is by transmitting, apart from the message, r random bits. These bits serve to garble the communication on the channel W. On the other hand, since the main channel C must be utilized in order to transmit these random bits, which Bob doesn't care about, they can be thought of as overhead.

Semantic security is a strong form of security. In fact, it is the "holly grail" of cryptography. In his seminal paper, Wyner discussed a weaker form of security (which we will not define). Loosely put, it was shown in [1] that in order to achieve this security, asymptotically in n, the codeword length, a fraction $r/n = I(W)$ of random bits is both required and sufficient.

For the BT scheme, the number of random bits used is closely tied to a function that we now define. Let $W : \mathcal{X} \to \mathcal{Y}$ be a symmetric, discrete, memoryless channel (abbreviated SDMC and defined shortly). The function $\Psi(W)$ is defined as

$$\Psi(W) \stackrel{\text{def}}{=} \log_2 |\mathcal{Y}| \; + \sum_{y \in \mathcal{Y}} W(y|0) \log_2 W(y|0) \; .$$

Next, define $W(y)$ as the probability of receiving y, assuming that the input to W is uniform. That is,

$$W(y) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} W(y|x) \; .$$

Since W is an SDMC, its capacity [2, Theorem 4.5.2.] is

$$I(W) = \sum_{y \in \mathcal{Y}} -W(y) \log_2 W(y) - \sum_{y \in \mathcal{Y}} -W(y|0) \log_2 W(y|0) \; .$$

By a direct application of Jensen's inequality, we have that

$$\Psi(W) \geqslant I(W) \; . \tag{1}$$

The above inequality is tight iff a uniform distribution on the input to W does not result in the output being uniformly distributed.

The following proposition links the number of random bits r employed by the BT scheme to the number of security bits σ required and the codeword length n.

*Proposition 1 (Restatment of [**?**]):* Let $W : \mathcal{X} \to \mathcal{Y}$ be the SDMC from Alice to Eve. Then, the BT scheme achieves at least σ bits of semantic security with a codeword length of n and r random bits, provided that

$$r = 2(\sigma + 1) + \sqrt{n} \log_2(|\mathcal{Y}| + 3) \sqrt{2(\sigma + 3)} + n \cdot \Psi(W) \; . \tag{2}$$

Proposition 1 implies that, asymptotically in n, a fraction of $r/n = \Psi(W)$ random bits are needed in order to achieve security in the BT scheme. In light of (1), this does not compare favorably to Wyner's original scheme, in which only a fraction of $I(W)$ random bits are needed.

The structure of this paper is as follows. In Section II we define the concepts of an equivalent channel. We then show a generic method by which an SDMC W can be replaced by an equivalent SDMC Q for which $\Psi(Q)$ is arbitrarily close to $I(W)$. Since Q is equivalent to W, we may substitute W by Q in Proposition 1 and show that, asymptotically, a fraction of $I(W)$ random bits are needed for the BT scheme to achieve security, making it comparable to Wyner's original scheme. Recalling (2), we see that apart from the coefficient $\Psi(W)$ of n, there is also a coefficient of $\sqrt{n}$, which is an increasing function of the output alphabet size of W. Thus, when refining to a non-asymptotic setting, we note that when we replace W by Q we must also be mindful of the output alphabet size of Q. Section III addresses the question of finding the best equivalent Q, when we constrain the output alphabet size of Q. In fact, we need not constrict ourselves to replace W by an equivalent Q, it suffices for Q to be upgraded (a concept to be defined) with respect to W. We show in Section IV that if W has binary input, then this observation can greatly reduce the number of random bits we are required to transmit.

II. Symmetric Channels and Letter Splitting

In this section, we first set up the basic notation and concepts used throughout. We then prove the claim that a fraction of $I(W)$ random bits asymptotically suffices for security.

A. Channels

A discrete memoryless channel (DMC) W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ will be denoted as $W : \mathcal{X} \to \mathcal{Y}$. Recall that W is a DMC if both $\mathcal{X}$ and $\mathcal{Y}$ are discrete, and the probability of receiving a vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$ given that the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ was transmitted over W is $W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i)$.

We say that the DMC W is symmetric (SDMC) if [2, Page 94] the set of output symbols $\mathcal{Y}$ can be partitioned into subsets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ such that the following holds. For a each such subset $\mathcal{Y}_t$, let A_t be the matrix for which the rows are indexed by $\mathcal{X}$, the columns by $\mathcal{Y}_t$, and for which entry $(x, y) \in \mathcal{X} \times \mathcal{Y}_t$ is equal to $W(y|x)$. Then, each row of A_t is a permutation of each other row, and each column is a permutation of each other column.

B. Degraded, Upgraded, and Equivalent Channels

A DMC $W : \mathcal{X} \to \mathcal{Y}$ is (stochastically) degraded with respect to a DMC $Q : \mathcal{X} \to \mathcal{Z}$, denoted $W \preceq Q$, if there exists an intermediate channel $P : \mathcal{Z} \to \mathcal{Y}$ such that

$$W(y|x) = \sum_{z \in \mathcal{Z}} Q(z|x) \cdot P(y|z) \ .$$

Namely, W is the result of concatenating the channel P to Q. Alternatively, we say that Q is upgraded with respect to W, and denote this as $Q \succeq W$. If W is both upgraded and degraded with respect to Q, then we say that W and Q are equivalent, and denote this as $Q \equiv W$. From the data-processing inequality [3, Theorem 2.8.1] we have that

$$Q \succeq W \quad \text{implies} \quad I(Q) \geqslant I(W) \ . \tag{3}$$

Thus, $Q \equiv W$ implies $I(Q) = I(W)$.

The following is a corollary to Proposition 1. Its proof is essentially: If Eve has the channel Q, she can simulate having the channel W.

Corollay 2: Let $W : \mathcal{X} \to \mathcal{Y}$, σ, r, and n be as in Proposition 1. Suppose the SDMC $Q : \mathcal{X} \to \mathcal{Z}$ is upgraded with respect to W. Then, we can substitute W by Q in Proposition 1. That is, we can take

$$r = 2(\sigma+1) + \sqrt{n}\log_2(|\mathcal{Z}|+3)\sqrt{2(\sigma+3)} + n \cdot \Psi(Q) \ . \tag{4}$$

C. Letter Splitting

Let $W : \mathcal{X} \to \mathcal{Y}$ be a given SDMC with a corresponding partition $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$. We say that a function $s : \mathcal{Y} \to \mathbb{N}$ from $\mathcal{Y}$ to the positive integers is an *output letter split* of W if for all $1 \leqslant t \leqslant T$ and all $y, y' \in \mathcal{Y}_t$ we have that $s(y) = s(y')$. Thus, the split function s assigns the same value to every letter in a subset $\mathcal{Y}_t$, and we can thus abuse notation and define $s(\mathcal{Y}_t)$ as that value. The SDMC corresponding to this split, $Q : \mathcal{X} \to \mathcal{Z}$ is defined as follows. The output alphabet of Q is gotten by duplicating each letter $y \in \mathcal{Y}$ and making $s(y)$ distinct copies:

$$\mathcal{Z} = \bigcup_{y \in \mathcal{Y}} \{y_1, y_2, \ldots, y_s \mid s = s(y)\} \ .$$

The transition probabilities of Q are given by

$$Q(y_i|x) = W(y|x)/s(y) \ , \quad (x, y_i) \in \mathcal{X} \times \mathcal{Z} \ .$$

Namely, each letter y is duplicated $s(y)$ times, and the conditional probability of receiving each copy is simply $1/s(y)$ times the corresponding probability in the original channel W.

Note that since W is a SDMC, then so is Q. Also, note that W and Q are equivalent channels, $W \equiv Q$.

Let an SDMC $W : \mathcal{X} \to \mathcal{Y}$ be given. As previously discussed, a discrepancy between $I(W)$ and $\Psi(W)$ arises when a uniform distribution on the input to W does not result in a uniform distribution on the output. We will now use the above mentioned splitting operation to define a channel $Q : \mathcal{X} \to \mathcal{Z}$. The merit of Q will be that a uniform input distribution results in an output distribution that is close to uniform. The price we will pay for this quasi-uniformity is a larger output alphabet, compared to that of W (recall that the coefficient of $\sqrt{n}$ in (4) is an increasing function of the output alphabet size).

Lemma 3: Let $W : \mathcal{X} \to \mathcal{Y}$ be an SDMC, and let $M \geqslant 1$ be a given positive integer. For each $y \in \mathcal{Y}$, define

$$s(y) = \lceil M \cdot W(y) \rceil \ ,$$

and let $Q : \mathcal{X} \to \mathcal{Z}$ be the channel resulting from the applying the letter splitting function s to W. Then,

$$\Psi(Q) - I(W) = \Psi(Q) - I(Q) \leqslant \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right) , \quad (5)$$

and

$$|\mathcal{Z}| \leqslant M + |\mathcal{Y}| . \quad (6)$$

Proof: The bound (6) follows by

$$|\mathcal{Z}| = \sum_{y\in\mathcal{Y}} s(y) = \sum_{y\in\mathcal{Y}} \lceil M \cdot W(y)\rceil \leqslant \sum_{y\in\mathcal{Y}} 1 + M \cdot W(y) = |\mathcal{Y}| + M .$$

We now prove (5). We start by simplifying $\Psi(Q) - I(Q)$ to

$$\Psi(Q) - I(Q) = \log_2|\mathcal{Z}| + \sum_{z\in\mathcal{Z}} Q(z)\log_2 Q(z) = \log_2|\mathcal{Z}| + \sum_{y\in\mathcal{Y}} W(y)\log_2\left(\frac{W(y)}{s(y)}\right) .$$

We have already proved (6), and thus have an upper bound on the first term. Thus, we concentrate now on the second term.

$$\sum_{y\in\mathcal{Y}} W(y)\log_2\left(\frac{W(y)}{s(y)}\right) = \sum_{y\in\mathcal{Y}} W(y)\log_2\left(\frac{W(y)}{\lceil M\cdot W(y)\rceil}\right) \leqslant \sum_{y\in\mathcal{Y}} W(y)\log_2\left(\frac{W(y)}{M\cdot W(y)}\right) = -\log_2 M .$$

Combining the above two bounds, we get

$$\Psi(Q) - I(Q) \leqslant \log_2(|\mathcal{Y}| + M) - \log_2 M = \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right) .$$

■

Combining Corollary 2 with Lemma 3 gives us the following theorem.

Theorem 4: Let $W : \mathcal{X} \to \mathcal{Y}$, σ, r, and n be as in Proposition 1. Let $M \geqslant 1$ be a parameter we are allowed to choose. Then, the number of random bits needed to achieve semantic security is at most

$$r = 2(\sigma+1) + \sqrt{n}\log_2(M + |\mathcal{Y}| + 3)\sqrt{2(\sigma+3)} + n\cdot\left(I(W) + \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right)\right) . \quad (7)$$

Setting, say, $M = n$ and taking $n \to \infty$ gives us

$$\lim_{n\to\infty}\frac{r}{n} = I(W) .$$

III. Optimal Letter Splitting

Recall that Theorem 4 arises from choosing a specific letter-splitting function, $s(y) = \lceil n \cdot W(y)\rceil$. Although the theorem states that this choice is a good choice asymptotically, a natural direction to pursue now is the finite n case. That is, given W, σ, and n, we may ask what is the best letter-splitting function one can choose so that (4) is minimized. We do not know how to answer this question. However, let us pose a related one. Namely, suppose $W : \mathcal{X} \to \mathcal{Y}$ is such that the subsets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ are all of the same size, μ. For example, if $|\mathcal{X}| = 2$, then our assumption holds when $\mathcal{Y}$ does not contain an erasure symbol, in which case we can always find a partition for which $\mu = 2$. We now ask, suppose we are given a parameter M which is a multiple of μ, and wish to find a letter-splitting function s for which $\sum_{y\in\mathcal{Y}} s(y) = M$ and for which the resulting channel Q has a minimum $\Psi(Q)$ value. We now show that a greedy algorithm can find such a letter-splitting function efficiently.

Algorithm 1: Greedy algorithm to find optimal splitting function

input : Channel $W : \mathcal{X} \to \mathcal{Y}$, a partition $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ where each subset is of size μ, a positive integer M which is a multiple of μ

output: A letter-splitting function s such that $\sum_{y\in\mathcal{Y}} s(y) = M$ and $\Psi(Q)$ is minimal

```
// Initialization
```
$s(\mathcal{Y}_1) = s(\mathcal{Y}_2) = \cdots = s(\mathcal{Y}_T) = 1$

```
// Main loop
```
for $i = 1, 2, \ldots, \frac{M}{\mu} - T$ **do**

  $t = \arg\max_{1\leqslant t\leqslant T} \sum_{y\in\mathcal{Y}_t} W(y)\log_2\left(\frac{s(\mathcal{Y}_t)+1}{s(\mathcal{Y}_t)}\right)$

  $s(\mathcal{Y}_t) = s(\mathcal{Y}_t) + 1$

return s

Theorem 5: Given a valid input to Algorithm 1, the output is a valid letter-splitting function s, such that $\sum_{y\in\mathcal{Y}} s(y) = M$ and the resulting channel Q is such that $\Psi(Q)$ is minimized.

Proof: First, note that after the initialization step, we have $\sum_{y\in\mathcal{Y}} s(y) = \mu \cdot T$. Each iteration obviously increments the sum by μ, so at the end we indeed have a letter-splitting function s such that $\sum_{y\in\mathcal{Y}} s(y) = M$.

With respect to optimality, first note that a channel Q which results from a splitting function has $I(Q) = I(W)$ constant, since $Q \equiv W$. Thus, minimizing $\Psi(Q)$ is equivalent to maximizing

$$I(Q) - \Psi(Q) = \sum_{y\in Y} -W(y)\log_2\left(\frac{W(y)}{s(y)}\right) - \log_2 M .$$

Clearing away constant terms, our target function becomes

$$\sum_{y\in\mathcal{Y}} W(y)\log_2 s(y) .$$

Recall that we must have $s(y) \geqslant 1$ for all $y \in \mathcal{Y}$. We now rephrase our optimization problem in an equivalent manner. Define the set

$$A = \bigcup_{y\in\mathcal{Y}} \bigcup_{i=1}^{M/\mu - T} \left\{\delta(y,i) = W(y)\log_2\left(\frac{i+1}{i}\right)\right\} .$$

Then, finding the optimal $s(y)$ is equivalent to choosing $M/\mu - T$ numbers from the set A such that their sum is maximal, and they satisfy the following constraint: If $\delta(y,i)$

was picked and $i > 1$, then $\delta(y, i-1)$ must be picked as well. To see the equivalence, define $s(y)$ as the largest i such that $\delta(y,i)$ was picked, or as 1 if no such δ was picked. The important point to note now is that the last constraint is redundant. That is, note that since $\log_2$ is a concave function, we have a "diminishing returns" effect

$$\log_2(i+2) - \log_2(i+1) < \log_2(i+1) - \log_2(i) \; .$$

Thus, the optimal solution will satisfy the constraint. Therefore, we can forget about the constraint and simply pick the M largest elements of A. It should now be easy to see that that is exactly what Algorithm 1 does. ∎

We note that the complexity of Algorithm 1 is $O(\log(T) \cdot (M/\mu - T))$, provided that one uses a heap as implemented in [4, Chapter 6] to find the maximal element to add at each iteration.

IV. Optimization for Continuous Alphabets

Recall that Theorem 4 has given us a method by which, asymptotically, only a fraction $r/n = I(W)$ random bits need to be utilized in order to achieve semantic security. However, the underlying assumption was that that the channel W had a finite output alphabet size. Specifically, Theorem 4 does not apply if $W : \mathcal{X} \to \mathcal{Y}$ is the binary-input Gaussian channel (BAWGN). More so, $\Psi(W)$ is undefined in this case, so we cannot even fall back to relying on Proposition 1. However, one need not go to such extremes. Even if W does have a finite output alphabet, but that alphabet size is rather large, we stand to lose much as implied by the coefficient of $\sqrt{n}$ in (7). Luckily, if W is a channel with binary input, we can derive a bound on r which is not a function of the output alphabet size of W. Thus, in this section, we will assume that $|\mathcal{X}| = 2$.

Apart from our assumption on a binary input, symmetry (yet to be defined in a non-DMC setting), and memorylessness, we make the following assumption for simplicity of exposition. Let the input alphabet of W be $\mathcal{X} = \{-1, 1\}$ and let the output alphabet be $\mathcal{Y} = \mathbb{R}$ the real numbers. Let the p.d.f. of W be f, and let it be symmetric:

$$f(y|1) = f(-y|-1) \; , \quad y \in \mathbb{R} \; . \tag{8}$$

Next, we assume that

$$f(y|1) \geqslant f(y|-1) \; , \quad y \geqslant 0 \; , \tag{9}$$

and also that the likelihood ratios increase with y. That is, for $y_1 < y_2$ we have

$$\frac{f(y_1|1)}{f(y_1|-1)} \leqslant \frac{f(y_2|1)}{f(y_2|-1)} \; , \quad -\infty < y_1 < y_2 < \infty \; . \tag{10}$$

The above implicitly assumes that $f(y|x) > 0$ for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Note that the above conditions hold for the BAWGN channel.

Recall that previously, we replaced W by an equivalent channel Q. In this section, we will replace W by an *upgraded* channel Q. As in Section III our figure of merit will be $\Psi(Q)$. We should first note that we can build on prior art and prior sections to upgrade W to Q such that $\Psi(Q) - I(W)$ is arbitrarily small. Namely, [5, Section VI] contains a method by which W can we upgraded to a DMC W' with a finite output alphabet size such that $I(W) - I(W')$ is arbitrarily small. With W' at hand, we can use the method in Section II to manufacture a DMC Q such that $\Psi(Q) - I(W')$ is arbitrarily small. Thus, we can combine the two methods and deduce that we can manufacture a Q such that $\Psi(Q) - I(W)$ is arbitrarily small. However, as explained earlier, apart for the difference $\Psi(Q) - I(W)$, we also care about the output alphabet size of Q. It turns out that if we were to use the above solution, requiring that $\Psi(Q) - I(W) \leqslant 1/M$ would imply that the output alphabet size of Q would be $O(M^2)$. In contrast, the method we will now show requires an output alphabet size of only $2M$.

We start by recursively defining the following equi-probable regions. Denote $y_0 = 0$. Next, for $1 \leqslant i < M$ and an already calculated y_{i-1}, let $y_i > y_{i-1}$ be such that

$$\int_{-y_i}^{-y_{i-1}} f(y|1)\, dy + \int_{y_{i-1}}^{y_i} f(y|1)\, dy = \frac{1}{M} \; .$$

Lastly, by abuse of notation, we "define" $y_M = \infty$. Thus, for $1 \leqslant i \leqslant M$, the regions

$$A_i = \{y \; : \; -y_i < y \leqslant -y_{i-1}\} \cup \{y \; : \; y_{i-1} \leqslant y < y_i\}$$

form a partition of $\mathcal{Y} = \mathbb{R}$ and are equi-probable with respect to $f(y|1)$ as well as $f(y|-1)$, by the symmetry condition (8).

For future reference, let us define for $0 \leqslant i \leqslant M$ the sets

$$B_i = \{y \; : \; y_{i-1} \leqslant y < y_i\}$$

Next, for $1 \leqslant i < M$, define

$$\lambda_i = \frac{f(y_i|1)}{f(y_1|-1)} \; ,$$

and let $\lambda_M = \infty$. Then, by (10) and (9), we have for $1 \leqslant i \leqslant M$ that

$$1 \leqslant \lambda_{i-1} = \inf_{y \in B_i} \frac{f(y|1)}{f(y|-1)} \leqslant \sup_{y \in B_i} \frac{f(y|1)}{f(y|-1)} \leqslant \lambda_i \; . \tag{11}$$

We now define the upgraded channel $Q : \mathcal{X} \to \mathcal{Z}$ and $Q' : \mathcal{X} \to \mathcal{Z}$. The output alphabet of Q is

$$\mathcal{Z} = \{z_1, \bar{z}_1, z_2, \bar{z}_2, \ldots, z_M, \bar{z}_M\} \; . \tag{12}$$

The channel Q is defined as

$$Q(z|1) = \begin{cases} \frac{\lambda_i}{M(\lambda_i+1)} & \text{if } z = z_i \text{ and } \lambda_i \neq \infty \; , \\ \frac{1}{M(\lambda_i+1)} & \text{if } z = \bar{z}_i \text{ and } \lambda_i \neq \infty \; , \\ \frac{1}{M} & \text{if } z = z_i \text{ and } \lambda_i = \infty \; , \\ 0 & \text{if } z = \bar{z}_i \text{ and } \lambda_i = \infty \; , \end{cases} \tag{13}$$

and

$$Q(z_i|-1) = Q(\bar{z}_i|1) \; , \quad Q(\bar{z}_i|-1) = Q(z_i|1) \; . \tag{14}$$

For analysis purposes, we now define an additional SDMC $Q' : \mathcal{X} \to \mathcal{Z}$. Note that Q and Q' share the same output alphabet. The channel Q' is defined similarly, but with an "index shift". That is, define $\lambda_0 = 1$. Then,

$$Q'(z|1) = \begin{cases} \frac{\lambda_{i-1}}{M(\lambda_{i-1}+1)} & \text{if } z = z_i \ , \\ \frac{1}{M(\lambda_{i-1}+1)} & \text{if } z = \bar{z}_i \ , \end{cases} \tag{15}$$

and

$$Q'(z_i|-1) = Q'(\bar{z}_i|1) \ , \quad Q'(\bar{z}_i|-1) = Q'(z_i|1) \ . \tag{16}$$

Lemma 6: The above channels satisfy $Q' \preceq W \preceq Q$.

Proof: The proof of $W \preceq Q$ is given in [5, Lemma 16]. We now show that $Q' \preceq W$. Since degradation is a transitive relation, we start by degrading W to a channel $\Phi : \mathcal{X} \to \mathcal{Z}$. We define Φ as follows.

$$\Phi(z|1) = \begin{cases} \int_{y \in B_i} f(y|1)\, dy & y \in B_i \ , \\ \int_{-y \in B_i} f(y|1)\, dy & -y \in B_i \ , \quad y \neq 0 \ , \end{cases}$$

and

$$\Phi(z_i|-1) = \Phi(\bar{z}_i|1) \quad \Phi(\bar{z}_i|-1) = \Phi(z|1) \ .$$

To show that $\Phi \preceq W$, we supply the intermediate channel:

$$P_1(z|y) = \begin{cases} z_i & y \in B_i \ , \\ \bar{z}_i & -y \in B_i \ , \quad y \neq 0 \ . \end{cases}$$

We now note that by (12),

$$\Phi(z_i|1) = \int_{y \in A_i} f(y|1) = \int_{y \in A_i} \frac{f(y|1)}{f(-y|1)} \cdot f(-y|1) \geqslant \int_{y \in A_i} \lambda_{i-1} \cdot f(-y|1) = \lambda_{i-1} \Phi(z_i|-1) \ .$$

Thus,

$$\Lambda_i = \Phi(z_i|1)/\Phi(z_i|-1) \geqslant \lambda_{i-1} \geqslant 1 \ .$$

To finish the proof, we now show an intermediate channel $P_2 : \mathcal{Z} \to \mathcal{Z}$ which degrades Φ to Q'. Specifically,

$$P_2(z'|z) = \begin{cases} 1 - p_i & z = z' = z_i \text{ or } z = z' = \bar{z}_i \ , \\ p_i & z = z_i, z' = \bar{z}_i \text{ or } z = \bar{z}_i, z' = z_i \ , \end{cases}$$

where

$$p_i = \left(\frac{\Lambda_i}{\Lambda_i + 1} - \frac{\lambda_{i-1}}{\lambda_{i-1} + 1} \right) \Big/ \left(\frac{\Lambda_i - 1}{\Lambda_i + 1} \right)$$

■

We are now ready to state our theorem, relating $\Psi(Q)$ to $I(W)$.

Theorem 7: Let $W : \mathcal{X} \to \mathcal{Y}$ be a continuous channel as defined above. For a given integer M, let $Q : \mathcal{X} \to \mathcal{Z}$ be the upgraded channel described previously. Then, $|\mathcal{Z}| = 2M$ and

$$\Psi(Q) - I(W) \leqslant \frac{1}{M} \ .$$

Proof: The claim $|\mathcal{Z}| = 2M$ is simply a restatement of (12). Next, note that for all $z \in \mathcal{Z}$

$$Q(z) = \frac{Q(z|1) + Q(z|-1)}{2} = \frac{1}{2M} \ .$$

Namely, a uniform input to Q results in a uniform output, and thus,

$$\Psi(Q) = I(Q) \ . \tag{17}$$

Hence, we are to prove that

$$I(Q) - I(W) \leqslant \frac{1}{M} \ .$$

By Lemma 6 and (3) we have that

$$I(Q') \leqslant I(W) \leqslant I(Q) \ .$$

Thus, it suffices to prove that

$$I(Q) - I(Q') = \frac{1}{M} \ . \tag{18}$$

To this end, define the function C as follows. For $0 \leqslant \lambda < \infty$

$$C[\lambda] = 1 - \frac{\lambda}{\lambda+1} \log_2 \left(1 + \frac{1}{\lambda} \right) - \frac{1}{\lambda+1} \log_2 (\lambda + 1) \ ,$$

and (for continuity) we define $C[\infty] = 1$. Now, a short calculation shows that for any SDMC Q with output alphabet $\mathcal{Z} = \{z_1, \bar{z}_1, z_2, \bar{z}_2, \ldots, z_M, \bar{z}_M\}$ and input alphabet $\mathcal{X} = \{-1, 1\}$ we have that

$$I(Q) = \sum_{i=1}^{M} Q(Z) C[Q(z|1)/Q(z|-1)] \ . \tag{19}$$

Specializing (19) to our Q gives

$$I(Q) = \frac{1}{M} \sum_{i=1}^{M} C[\lambda_i] \ ,$$

while specialzing (19) to Q' gives

$$I(Q') = \frac{1}{M} \sum_{i=1}^{M} C[\lambda_{i-1}] \ ,$$

Thus,

$$I(Q) - I(Q') = \frac{1}{M} \left(C[\lambda_M] - C[\lambda_0] \right) = \frac{1}{M} \ .$$

■

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J*, vol. 54(8), pp. 1355–1387, 1975.

[2] R. G. Gallager, *Information Theory and Reliable Communications*. New York: John Wiley, 1968.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, Massachusetts: The MIT Press, 2001.

[5] I. Tal and A. Vardy, "How to construct polar codes," *submitted to IEEE Trans. Inform. Theory, available online as* `arXiv:1105.6164v2`, 2011.