# Channel Upgradation for Non-Binary Input Alphabets and MACs

Uzi Pereg and Ido Tal
Department of Electrical Engineering,
Technion, Haifa 32000, Israel.
Email: uzipereg@tx.technion.ac.il, idotal@ieee.org

***Abstract*—To be considered for an IEEE Jack Keil Wolf ISIT Student Paper Award.**

**Consider a single-user or multiple-access channel with a large output alphabet. A method to approximate the channel by an upgraded version having a smaller output alphabet is presented and analyzed. The gain in symmetric channel capacity is controlled through a fidelity parameter. The larger the fidelity parameter, the better the approximation on the one hand, but the larger the new output alphabet on the other.**

**The approximation method is instrumental when constructing polar codes. No assumption is made on the symmetry of the original channel, and the input alphabet need not be binary.**

***Index Terms*—Polar codes, multiple-access channel, sum-rate, channel degradation, channel upgradation.**

## I. Introduction

Polar codes have recently been invented by Arıkan [1]. In his seminal paper, Arıkan treated the following channel model, over which information is to be sent: a binary-input, memoryless, output-symmetric channel. The definition of polar codes was soon generalized to channels with prime input alphabet size [2]. A further generalization to a polar coding scheme for a multiple-access channel (MAC) with prime input alphabet size is presented in [3] and [4].

The communication schemes in [2]–[4] are explicit, have efficient encoding and decoding algorithms, and achieve symmetric capacity (sum-rate symmetric capacity in the MAC setting). However, [2]–[4] do not discuss how an efficient construction of the underlying polar code is to be carried out. The problem of constructing polar codes for these settings was discussed in [5], in which a degraded approximation of the bit-channels is derived. The current paper is the natural counterpart of [5], since we now derive an upgraded approximation.

In addition to single-user and multiple-access channels, polar codes have been used to tackle many classical information theoretic problems. Of these, we single-out the wiretap channel [6], since the results in this paper are especially relevant when using polar codes to code for the wiretap setting, as was done in [7]–[10]. Namely, in brief, if we are to transmit information over a bit-channel, it must be an almost pure-noise channel to Eve. In order to validate this property computationally, it suffices to show that an upgraded version of the bit-channel is almost pure-noise.

The same problem we consider in this paper — approximating a channel with an upgraded version having a prescribed output alphabet size — was recently considered by Ghayoori and Gulliver in [11]. Broadly speaking, the method presented in [11] builds upon the pair and triplet merging ideas presented in [12] and analyzed in [13]. In [11], it is stated that the resulting approximation is expected to be close to the original channel. As yet, we are not aware of an analysis making this claim precise. In this paper, we present an alternative upgrading approximation method which seems easier to analyze. Thus, with respect to our method, we are able to derive an upper bound on the gain in symmetric capacity. The bound, is given as Theorem 1 below, and is the main analytical result of this paper.

Let the underlying MAC have an input alphabet of size $p$ and $t$ users ($t = 1$ if we are in fact considering a single-user channel). We would like to mention up-front that the running time of our upgradation algorithm grows very fast in $q = p^t$. Thus, our algorithm can only be argued to be practical for small values of $q$.

The structure of this paper is as follows. In Section II we set up the basic concepts and notation that will be used later on. Section III describes the binning operation as it is used in our algorithm. The binning operation is a preliminary step used later on to define the upgraded channel. Section IV contains our approximation algorithm, as well as the statement of Theorem 1. Section V is devoted to proving Theorem 1. The full paper [14] is available online, and contains the proof of all the statements in this paper.

## II. Preliminaries

### A. Multiple Access Channel

Let $W : \mathcal{X}^t \to \mathcal{Y}$ designate a generic $t$-user MAC, where $\mathcal{X} = \{0, 1, \ldots, p-1\}$ is the input alphabet, $p$ is a positive integer[1], and $\mathcal{Y}$ is the finite[2] output alphabet. Denote a vector of user inputs by $\mathbf{u} \in \mathcal{X}^t$, where $\mathbf{u} = (u^{(l)})_{l=1}^t$. Our MAC is defined through the probability function $W$, where $W(y|\mathbf{u})$ is the probability of observing the output $y$ given that the user input was $\mathbf{u}$.

### B. Degradation and Upgradation

The notions of a (stochastically) degraded and upgraded MAC are defined in an analogous way to that of a degraded

[1] Following the observation in [15], we do not constrain ourselves to an input alphabet which is prime.

[2] The assumption that $\mathcal{Y}$ is finite is only meant to make the presentation simpler. Our method readily generalizes to continuous output alphabet cases.

and upgraded single-user channel, respectively. That is, we say that a $t$-user MAC $Q : \mathcal{X}^t \to \mathcal{Z}$ is *degraded* with respect to $W : \mathcal{X}^t \to \mathcal{Y}$, if there exists a channel $\mathcal{P} : \mathcal{Y} \to \mathcal{Z}$ such that for all $z \in \mathcal{Z}$ and $\mathbf{u} \in \mathcal{X}^t$,

$$Q(z|\mathbf{u}) = \sum_{y \in \mathcal{Y}} W(y|\mathbf{u}) \cdot \mathcal{P}(z|y) \; .$$

We write $Q \preceq W$ to denote that $Q$ is degraded with respect to $W$.

Conversely, we say that a $t$-user MAC $Q^{'} : \mathcal{X}^t \to \mathcal{Z}^{'}$ is *upgraded* with respect to $W : \mathcal{X}^t \to \mathcal{Y}$ if $W$ is degraded with respect to $Q^{'}$. We denote this as $Q^{'} \succeq W$. If $Q$ satisfies both $Q \preceq W$ and $Q \succeq W$, then $Q$ and $W$ are said to be *equivalent*. We express this by $W \equiv Q$.

### C. The Sum-Rate Criterion

Let a $t$-user MAC $W : \mathcal{X}^t \to \mathcal{Y}$ be given. Next, let $\mathbf{U} = (U^{(l)})_{l=1}^{t}$ be a random variable uniformly distributed over $\mathcal{X}^t$. Let $Y$ be the random variable one gets as the output of $W$ when the input is $\mathbf{U}$. The sum-rate of $W$ is defined as the mutual information

$$R(W) = I(\mathbf{U}; Y) \; .$$

Note that by the data-processing inequality [16, Theorem 2.8.1], we have that $W \preceq Q$ implies that $R(W) \leq R(Q)$. Whereas $W' \succeq Q$ implies that $R(W') \geq R(Q)$. Thus, equivalent MACs have the same sum-rate.

In Section IV we show how to obtain an upgraded approximation of $W$. The original MAC $W : \mathcal{X}^t \to \mathcal{Y}$ is approximated by another MAC $Q^{'} : \mathcal{X}^t \to \mathcal{Z}^{'}$ with a smaller output alphabet size. Then, we bound the difference (increment) in the sum-rate. Our use of the sum-rate as the figure of merit is justified by [5, Lemma 2].

## III. The Binning Operation

### A. Regions and Bins

In [5], a binning operation was used to approximate a given channel by a degraded version of it. Our algorithm uses a related yet different binning rule, as a preliminary step towards upgrading the channel $W : \mathcal{X}^t \to \mathcal{Y}$.

Let the random variables $\mathbf{U}$ and $Y$ be as earlier. Assume that the output alphabet $\mathcal{Y}$ has been purged of all letters $y$ with zero probability, since these outputs never occur. Thus, we can define the function $\varphi_W : \mathcal{X}^t \times \mathcal{Y} \to [0,1]$ as the a posteriori probability (APP):

$$\varphi_W(\mathbf{u}|y) = \mathbb{P}(\mathbf{U} = \mathbf{u} | Y = y) = \frac{W(y|\mathbf{u})}{\sum\limits_{\mathbf{v} \in \mathcal{X}^t} W(y|\mathbf{v})} \; , \tag{1}$$

for every input $\mathbf{u} \in \mathcal{X}^t$ and every letter in the (purged) output alphabet $y \in \mathcal{Y}$. Next, for $y \in \mathcal{Y}$ let us denote $p_W(y) = \mathbb{P}(Y = y)$ , and define $\eta : [0,1] \to \mathbb{R}$ by $\eta(x) = -x \cdot \ln x$ , where $\ln(\cdot)$ stands for *natural* logarithm. Thus, the sum-rate can be expressed as

$$\begin{aligned} R(W) &= \ln q - \sum_{y \in \mathcal{Y}} p_W(y) H(\mathbf{U}|Y = y) \\ &= \ln q - \sum_{y \in \mathcal{Y}} p_W(y) \sum_{\mathbf{u} \in \mathcal{X}^t} \eta\left(\varphi_W(\mathbf{u}|y)\right) \; , \end{aligned}$$

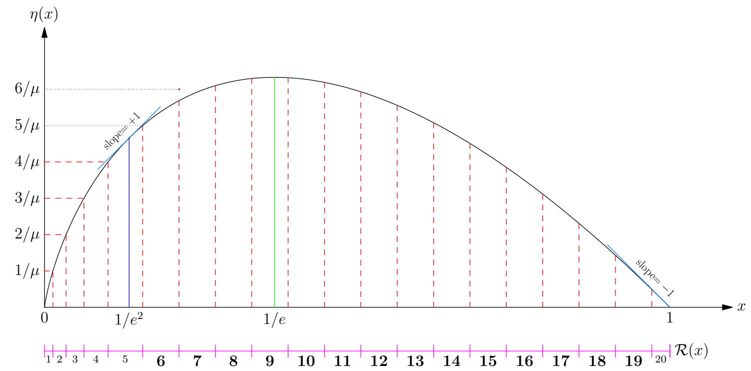

Fig. 1. Functions $\eta(x) = -x \cdot \ln x$ and $\mathcal{R}(x)$. The fidelity parameter $\mu$ is set to $\mu = 17.2$, which results in the number of regions being $M = 20$. The bold-faced regions have a horizontal increment (width) of exactly $1/\mu$, while their vertical increment is less than $1/\mu$, as the horizontal dotted lines in the figure demonstrate for region 6. As a leftover effect, the last region, 20, has horizontal and vertical increments which are both less than $1/\mu$.

where $H(\mathbf{U}|Y = y)$ is the conditional entropy of $\mathbf{U}$ given that $Y = y$ (measured in natural units).

As a first step towards the definition of our bins, we quantize the domain of $\eta(x)$ with resolution specified by a fidelity parameter $\mu$, where

$$\mu \geq \max(5, q(q-1)) \; . \tag{2}$$

The interval $[0,1)$ is partitioned into $M = M_\mu$ non-empty regions of the form $[b_i, b_{i+1})$ , $i \in \{1, 2, \ldots, M\}$ . Starting from $b_1 = 0$, the endpoint of the $i$th region is given by

$$b_{i+1} = \max \left\{ \; 0 < x \leq 1 \; : \; \begin{array}{c} x \leq b_i + \frac{1}{\mu} \; , \\ \\ |\eta(x) - \eta(b_i)| \leq \frac{1}{\mu} \end{array} \right\} \tag{3}$$

And so it is inferred that for all regions $1 \leq i < M$ (all regions but the last), there is *either* a horizontal *or* vertical increment of $1/\mu$. That is, $b_{i+1} - b_i = \frac{1}{\mu}$ or $|\eta(b_{i+1}) - \eta(b_i)| = \frac{1}{\mu}$ , but typically not both (Figure 1).

Denote the region to which $x$ belongs by $\mathcal{R}(x) = \mathcal{R}_\mu(x)$. Namely, $\mathcal{R}(x) = i$ iff $x \in [b_i, b_{i+1})$ , with the exception of $x = 1$ belonging to the last region, meaning $\mathcal{R}(1) = M$.

Based on the quantization regions defined above, we define our binning rule. Two output letters $y_1, y_2 \in \mathcal{Y}$ are said to be in the same bin if for all $\mathbf{u} \in \mathcal{X}^t$ we have that $\mathcal{R}(\varphi_W(\mathbf{u}|y_1)) = \mathcal{R}(\varphi_W(\mathbf{u}|y_2))$ .

### B. Merging of letters in the same bin

Recall that our ultimate aim is to approximate the original channel $W : \mathcal{X}^t \to \mathcal{Y}$ by an upgraded version having a smaller output alphabet. As we will see, the output alphabet of the approximating channel will be a union of two sets. In this subsection, we define one of these sets, denoted by $\mathcal{Z}$.

Figuratively, we think of $\mathcal{Z}$ as the result of merging together all the letters in the same bin. That is, the size of $\mathcal{Z}$ is the number of non-empty bins, as each non-empty bin corresponds to a distinct letter $z \in \mathcal{Z}$. Denote by $\mathcal{B}(z)$ the set of letters in $\mathcal{Y}$ which form the bin associated with $z$. Thus, all the symbols

$y \in \mathcal{B}(z)$ can be thought of as having been merged into one symbol $z$.

As we will see, the size of $\mathcal{Z}$ can be upper-bounded by an expression that is not a function of $|\mathcal{Y}|$.

### *C. The APP measure* $\psi$

In this subsection, we define an a posteriori probability measure on the input alphabet $\mathcal{X}^t$, given a letter from the merged output alphabet $\mathcal{Z}$. We denote this APP measure as $\psi(\mathbf{u}|z)$, defined for $\mathbf{u} \in \mathcal{X}^t$ and $z \in \mathcal{Z}$. The measure $\psi(\mathbf{u}|z)$ will be used in Section IV in order to define the approximating channel.

For each bin define the *leading input* as

$$\mathbf{u}^* = \mathbf{u}^*(z) \triangleq \arg\max_{\mathbf{u}\in\mathcal{X}^t} \left[ \max_{y\in\mathcal{B}(z)} \varphi_W(\mathbf{u}|y) \right] , \tag{4}$$

where ties are broken arbitrarily. For $z \in \mathcal{Z}$, let

$$\psi(\mathbf{u}|z) = \min_{y\in\mathcal{B}(z)} \varphi_W(\mathbf{u}|y) \quad \text{for all } \mathbf{u} \neq \mathbf{u}^*, \tag{5a}$$

$$\psi(\mathbf{u}^*|z) = 1 - \sum_{\mathbf{u}\neq\mathbf{u}^*} \psi(\mathbf{u}|z) . \tag{5b}$$

Informally, we note that if the bins are "sufficiently narrow" (if $\mu$ is sufficiently large), then $\psi(\mathbf{u}|z)$ is close to $\varphi_W(\mathbf{u}|y)$, for all $\mathbf{u} \in \mathcal{X}^t$, $z \in \mathcal{Z}$, and $y \in \mathcal{B}(z)$. The above will be made exact in Lemma 9 below.

## IV. The Upgraded Approximation

Now we are in position to define our $t$-user MAC approximation $Q^{'} : \mathcal{X}^t \to (\mathcal{Z} \cup K)$, where $K$ is a set of additional symbols to be specified in this section. We refer to these new symbols as "boost" symbols; that is, noiseless symbols.

Let $y \in \mathcal{Y}$ and $\mathbf{u} \in \mathcal{X}^t$ be given, and let $z$ correspond to the bin $\mathcal{B}(z)$ which contains $y$. Define the quantity $\alpha_{\mathbf{u}}(y)$ as

$$\alpha_{\mathbf{u}}(y) \triangleq \begin{cases} \frac{\psi(\mathbf{u}|z)}{\varphi_W(\mathbf{u}|y)} \cdot \frac{\varphi_W(\mathbf{u}^*|y)}{\psi(\mathbf{u}^*|z)} & \text{if } \varphi_W(\mathbf{u}|y) \neq 0 , \\ 1 & \text{if } \varphi_W(\mathbf{u}|y) = 0 . \end{cases} \tag{6}$$

As stated in Sub-section V-C, $\alpha_{\mathbf{u}}(y)$ is indeed well defined and is between 0 and 1. Next, for $\mathbf{u} \in \mathcal{X}^t$, let

$$\varepsilon_{\mathbf{u}} \triangleq \sum_{y\in\mathcal{Y}} (1 - \alpha_{\mathbf{u}}(y)) W(y|\mathbf{u}) . \tag{7}$$

We now define $K$, the set of output "boost" symbols. Namely, we define a boost symbol for each non-zero $\varepsilon_{\mathbf{u}}$,

$$K = \{ k_{\mathbf{u}} : \mathbf{u} \in \mathcal{X}^t , \varepsilon_{\mathbf{u}} > 0 \} .$$

Lastly, the probability function $Q^{'}$ of our upgraded MAC is defined as follows. With respect to non-boost symbols, define for all $z \in \mathcal{Z}$ and $\mathbf{u} \in \mathcal{X}^t$,

$$Q^{'}(z|\mathbf{u}) = \sum_{y\in\mathcal{B}(z)} \alpha_{\mathbf{u}}(y) W(y|\mathbf{u}) . \tag{8a}$$

With respect to boost symbols, define for all $\kappa_{\mathbf{v}} \in K$ and $\mathbf{u} \in \mathcal{X}^t$,

$$Q^{'}(\kappa_{\mathbf{v}}|\mathbf{u}) = \begin{cases} \varepsilon_{\mathbf{u}} & \text{if } \mathbf{u} = \mathbf{v} , \\ 0 & \text{otherwise} . \end{cases} \tag{8b}$$

Note that if a boost symbol $\kappa_{\mathbf{u}}$ is received at the output of $Q^{'} : \mathcal{X}^t \to (\mathcal{Z} \cup K)$, we know for certain that the input was $\mathbf{U} = \mathbf{u}$.

The following theorem presents the properties of our upgraded approximation of $W$. The proof concludes Section V.

*Theorem* 1. Let $W : \mathcal{X}^t \to \mathcal{Y}$ be a $t$-user MAC, and let $\mu$ be a given fidelity parameter that satisfies (2) . Let $Q^{'} : \mathcal{X}^t \to (\mathcal{Z} \cup K)$ be the MAC obtained from $W$ by the above definition (8). Then,

(i) The MAC $Q^{'}$ is well defined and is upgraded with respect to $W$.

(ii) The increment in sum-rate is bounded by

$$R(Q^{'}) - R(W) \leq \frac{q-1}{\mu} (2 + q \cdot \ln q) .$$

(iii) The output alphabet size of $Q^{'}$ is bounded by $(2\mu)^q + q$.

Note that the input alphabet size $q$ is usually considered to be a given parameter of the communications system. Therefore, we can think of $q$ as being a constant. In this view, Theorem 1 claims that our upgraded-approximation has a sum-rate deviation of $\mathcal{O}(\frac{1}{\mu})$, and an output-alphabet of size $\mathcal{O}(\mu^q)$.

## V. Analysis

We now examine the algorithm step by step, and state the relevant lemmas and properties for each step (proved in [14]). This eventually leads up to the proof of Theorem 1.

### *A. Quantization Properties*

In Section III-A, we have quantized the domain of the function $\eta(x) = -x \cdot \ln x$ for the purpose of binning. Now, we would like to discuss a few properties of this definition, which are exemplified in Figure 1.

*Lemma* 2. Let the extreme points $\{b_i : 1 \leq i \leq M+1\}$ partition the domain interval $0 \leq x \leq 1$ into quantization regions (intervals), as in Section III-A (see (3)). Thus,

(i) if $0 \leq b_i < b_{i+1} < \frac{1}{e^2}$ , then $\eta(b_{i+1}) - \eta(b_i) = 1/\mu$ .

(ii) Otherwise, if $\frac{1}{e^2} \leq b_i < b_{i+1} < 1$ , then $b_{i+1} - b_i = 1/\mu$ .

The following corollary will be used to bound the number of bins, namely $|\mathcal{Z}|$, later on.

*Corollary* 3. The number of quantization regions $M = M_\mu$ is bounded by $M \leq 2\mu$ .

The corollary, following the lemma below, will play a significant role in the proof of Theorem 1.

*Lemma* 4. Given $x \in [0, 1)$, let $i = \mathcal{R}(x)$. That is, $b_i \leq x < b_{i+1}$ . Also, let $0 < \delta \leq b_{i+1} - b_i$ , such that $x + \delta \leq 1$. Then, $|\eta(x + \delta) - \eta(x)| \leq \frac{1}{\mu}$ .

The corollary below is an immediate consequence of Lemma 4.

*Corollary* 5. All $x_1$ and $x_2$ that belong to the same quantization region (that is: $\mathcal{R}(x_1) = \mathcal{R}(x_2)$) satisfy

$$|\eta(x_1) - \eta(x_2)| \leq \frac{1}{\mu} .$$

The following lemma claims that each quantization interval, save the last, is at least as wide as the previous intervals.

*Lemma* 6. Let the width of the $i$th quantization interval be denoted by $\Delta_i = b_{i+1} - b_i$ , for $i = 1, 2, \ldots, M$ . Then the sequence $\{\Delta_i\}_{i=1}^{M-1}$ (the last interval excluded) is a non-decreasing sequence.

Following the quantization definition, the output letters in $\mathcal{Y}$ were divided into bins (Section III-B). Each bin is represented by a single letter in $\mathcal{Z}$. The following lemma upper bounds the size of $\mathcal{Z}$.

*Lemma* 7. The size of $\mathcal{Z}$ is bounded by $|\mathcal{Z}| \leq (2\mu)^q$ .

Consider a given bin (and a given $z \in \mathcal{Z}$). Depending on $\mathbf{u} \in \mathcal{X}^t$, all $y \in \mathcal{B}(z)$ share the same

$$i(\mathbf{u}) = i_z(\mathbf{u}) \triangleq \mathcal{R}\left(\varphi_W(\mathbf{u}|y)\right) . \tag{9}$$

Thus the bin can be characterized by the set of region-indices $\{\, i_z(\mathbf{u}) : \mathbf{u} \in \mathcal{X}^t \,\}$ . The following lemma claims that the leading input, defined in (4), is in the *leading region*.

*Lemma* 8. Consider a given $z \in \mathcal{Z}$. Let $i(\mathbf{u})$ be given by (9) for all $\mathbf{u} \in \mathcal{X}^t$, and let $\mathbf{u}^*$ be as in (4). Then

$$i(\mathbf{u}^*) = \max \left\{\, i(\mathbf{u}) : \mathbf{u} \in \mathcal{X}^t \,\right\} .$$

### *B. Properties of $\psi$*

Recall that the APP measure $\psi(\mathbf{u}|z)$ was defined in Subsection III-C. We start this subsection by claiming that $\psi$ is "close" to the APP of the original channel.

*Lemma* 9. For each $z \in \mathcal{Z}$, let $\mathbf{u}^* = \mathbf{u}^*(z)$ be the leading-input defined by (4), and let $\psi(\mathbf{u}|z)$ be the probability measure on $\mathbf{u} \in \mathcal{X}^t$ defined in (5). Then, for all $z \in \mathcal{Z}$ and $y \in \mathcal{B}(z)$,

$$|\eta\left(\varphi_W(\mathbf{u}|y)\right) - \eta\left(\psi(\mathbf{u}|z)\right)| \leq \begin{cases} \frac{1}{\mu} & \text{if } \mathbf{u} \neq \mathbf{u}^* , \\ \frac{q-1}{\mu} & \text{if } \mathbf{u} = \mathbf{u}^* . \end{cases}$$

Let $z \in \mathcal{Z}$ and $y \in \mathcal{B}(z)$ be given. Thus, we claim without proof that

$$\psi(\mathbf{u}^*|z) \geq \varphi_W(\mathbf{u}^*|y) \geq \frac{1}{q} - \frac{1}{\mu} > 0 . \tag{10}$$

The following lemma states that $\psi$ and $\varphi_W$ are close in a multiplicative sense as well, when we are considering $\mathbf{u}^*$.

*Lemma* 10. Consider a given $z \in \mathcal{Z}$. Then, for all $y \in \mathcal{B}(z)$,

$$0 \leq 1 - \frac{q(q-1)}{\mu} \leq \frac{\varphi_W(\mathbf{u}^*|y)}{\psi(\mathbf{u}^*|z)} \leq 1 .$$

### *C. The MAC $W'$*

We now define the channel $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$, an upgraded version of $W : \mathcal{X}^t \to \mathcal{Y}$. The definition makes heavy use of $\alpha_{\mathbf{u}}(y)$, defined in (6). We show in [14], using (10) and the above definitions, that $\alpha_{\mathbf{u}}(y)$ is well defined for all $y \in \mathcal{Y}$, and that

$$0 \leq \alpha_{\mathbf{u}}(y) \leq 1 . \tag{11}$$

We now define $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$, an upgraded version of $W$. For all $y \in \mathcal{Y}$ and for all $\mathbf{u} \in \mathcal{X}^t$, define

$$W'(y|\mathbf{u}) = \alpha_{\mathbf{u}}(y) \cdot W(y|\mathbf{u}) . \tag{12a}$$

Whereas, for all $\kappa_{\mathbf{v}} \in K$ and for all $\mathbf{u} \in \mathcal{X}^t$, define

$$W'(\kappa_{\mathbf{v}}|\mathbf{u}) = \begin{cases} \varepsilon_{\mathbf{u}} = \sum_{y \in \mathcal{Y}} (1 - \alpha_{\mathbf{u}}(y)) W(y|\mathbf{u}) & \text{if } \mathbf{u} = \mathbf{v} , \\ 0 & \text{otherwise.} \end{cases} \tag{12b}$$

We conclude this subsection with two lemmas that will be useful in the proof of Theorem 1.

*Lemma* 11. The MAC $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$ is well-defined and is upgraded with respect to $W : \mathcal{X}^t \to \mathcal{Y}$. That is, $W' \succeq W$ .

*Lemma* 12. Let $\varepsilon_{\mathbf{u}}$ be given by (7) for all $\mathbf{u} \in \mathcal{X}^t$. Then,

$$\frac{1}{q} \sum_{\mathbf{u} \in \mathcal{X}^t} \varepsilon_{\mathbf{u}} \leq \frac{q(q-1)}{\mu} .$$

### *D. Consolidation*

In the previous section, we defined $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$ which is an upgraded version of $W : \mathcal{X}^t \to \mathcal{Y}$. Note that the output alphabet of $W'$ is *larger* than that of $W$, and our original aim was to *reduce* the output alphabet size. We do this now by consolidating letters which essentially carry the same information.

Consider the output alphabet $\mathcal{Y} \cup K$ of our upgraded MAC $W'$, compared to the original output alphabet $\mathcal{Y}$. Note that, while the output letters $y \in \mathcal{Y}$ are the same output letters we started with, their APP values are *modified* and satisfy the following.

*Lemma* 13. Let $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$ be the MAC defined in Subsection V-C. Then, all the output letters $y \in \mathcal{B}(z)$ have the same modified APP values (for each $\mathbf{u} \in \mathcal{X}^t$ separately). Namely, for all $\mathbf{u} \in \mathcal{X}^t$, and for all $z \in \mathcal{Z}$ and $y \in \mathcal{B}(z)$, we have that $\varphi_{W'}(\mathbf{u}|y) = \psi(\mathbf{u}|z)$ .

Note that consolidating symbols with equal APP values results in an equivalent MAC (proof is omitted). Thus, we consolidate (map w.p.1) all the members of $\mathcal{B}(z)$ to $z$ :

$$Q'(z|\mathbf{u}) = \begin{cases} \sum_{y \in \mathcal{B}(z)} W'(z|\mathbf{u}) & \text{if } z \in \mathcal{Z} , \\ \\ W'(z|\mathbf{u}) & \text{if } z \in K . \end{cases} \tag{13}$$

for all $z \in \mathcal{Z} \cup K$ and for all $\mathbf{u} \in \mathcal{X}^t$. Based on (12), it can be easily shown that the alternative definition above agrees with the definition of $Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K)$ in (8).

The rest of this section is dedicated to proving Theorem 1.

*Proof of Theorem 1:*

We first prove part (i). Since $Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K)$ is a result of applying consolidation on $W' : \mathcal{X}^t \to (\mathcal{Y} \cup K)$, it follows that $Q'$ is well defined as well.

According to Lemma 11, $W' \succeq W$. Since $W'$ and $Q'$ are equivalent, and since upgradation transitivity immediately follows from the definition, it follows that $Q' \succeq W$.

We now move to part (ii) of the theorem, which concerns the sum-rate difference. Recall that the random variable $Y$ has been defined as the output of $W : \mathcal{X}^t \to \mathcal{Y}$ when the input is $\mathbf{U}$. Similarly, define $Z'$ as the output of $Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K)$ when the input is $\mathbf{U}$.

To estimate the APPs for $Q^{'} : \mathcal{X}^t \to (\mathcal{Z} \cup K)$, we may use (1) and (13). By Lemma 13, we have that $\varphi_{Q^{'}}(\mathbf{u}|z) = \psi(\mathbf{u}|z)$ , for all $\mathbf{u} \in \mathcal{X}^t$ and for all $z \in \mathcal{Z}$ (for all non-boost output symbols). Whereas for boost-symbols, $\varphi_{Q^{'}}(\mathbf{u}|\kappa) \in \{0,1\}$ for all $\mathbf{u} \in \mathcal{X}^t$ and for all $\kappa \in K$. Denote the entropy of the probability distribution defined in Section III-C by $H_\psi(\mathbf{U}|Z=z) = \sum_{\mathbf{u}\in\mathcal{X}^t} \eta\left[\psi(\mathbf{u}|z)\right]$ . Thus,

$$\begin{aligned} R(Q^{'}) = \ln q - \sum_{z\in\mathcal{Z}} p_{Q^{'}}(z) H_\psi(\mathbf{U}|Z=z) \\ - \sum_{\kappa\in K} p_{Q^{'}}(\kappa) H(\mathbf{U}|Z^{'}=\kappa) \ . \end{aligned}$$

However, the last term is zero due to the following observation. Given that the output of the MAC $Q^{'}$ is $\kappa_{\mathbf{v}}$ for some $\mathbf{v} \in \mathcal{X}^t$, the input $\mathbf{U}$ is known to be $\mathbf{v}$ (it is deterministic). Hence $H(\mathbf{U}|Z'=\kappa_{\mathbf{v}}) = 0$ for all $\kappa_{\mathbf{v}} \in K$. Hence

$$R(Q^{'}) = \ln q - \sum_{z\in\mathcal{Z}} p_{Q^{'}}(z) H_\psi(\mathbf{U}|Z=z) \ . \qquad (14)$$

Next we define a new auxiliary quantity to ease the proof. But first, define the random variable $Z$ as the letter in the merged output alphabet $\mathcal{Z}$ corresponding to $Y$. Namely, the realization $Z = z$ occurs whenever $Y$ is contained in $\mathcal{B}(z)$. The probability of that realization is

$$p_{\mathcal{B}}(z) \triangleq \mathbb{P}(Z=z) = \sum_{y\in\mathcal{B}(z)} p_W(y) \ . \qquad (15)$$

Note that the joint distribution $p_{\mathcal{B}}(z) \cdot \psi(\mathbf{u}|z)$ does *not* necessarily induce a true MAC (for a uniformly distributed input vector $\mathbf{U}$). Nevertheless, we plug this joint distribution into the sum-rate expression, with due caution. In other words, we define a new quantity $J(\mathbf{U};Z)$, which is a surrogate for mutual information. Namely, define

$$\begin{aligned} J(\mathbf{U};Z) &\triangleq \ln q - \sum_{z\in\mathcal{Z}} p_{\mathcal{B}}(z) \cdot H_\psi(\mathbf{U}|Z=z) \\ &= \ln q - \sum_{z\in\mathcal{Z}} p_{\mathcal{B}}(z) \sum_{\mathbf{u}\in\mathcal{X}^t} \eta\left[\psi(\mathbf{u}|z)\right] \ . \qquad (16) \end{aligned}$$

Now, we would like to bound the increment in sum-rate. To this end, we prove two bounds and then sum. First, note that

$$\begin{aligned} &J(\mathbf{U};Z) - R(W) = \sum_{y\in\mathcal{Y}} p_W(y) \sum_{\mathbf{u}\in\mathcal{X}^t} \eta(\varphi_W(\mathbf{u}|y)) \\ &- \sum_{z\in\mathcal{Z}} p_{\mathcal{B}}(z) \sum_{\mathbf{u}\in\mathcal{X}^t} \eta(\psi(\mathbf{u}|z)) = \\ &\sum_{z\in\mathcal{Z}} \sum_{y\in\mathcal{B}(z)} p_W(y) \sum_{\mathbf{u}\in\mathcal{X}^t} \left[\eta(\varphi_W(\mathbf{u}|y)) - \eta(\psi(\mathbf{u}|z))\right] \le \\ &\sum_{z\in\mathcal{Z}} \sum_{y\in\mathcal{B}(z)} p_W(y) \cdot |\eta(\varphi_W(\mathbf{u}^*|y)) - \eta(\psi(\mathbf{u}^*|z))| + \\ &\sum_{z\in\mathcal{Z}} \sum_{y\in\mathcal{B}(z)} p_W(y) \sum_{\mathbf{u}\neq\mathbf{u}^*} |\eta(\varphi_W(\mathbf{u}|y)) - \eta(\psi(\mathbf{u}|z))| \le \\ &2 \cdot \frac{q-1}{\mu} \ , \qquad (17) \end{aligned}$$

where the last inequality is due to Lemma 9.

For the second bound, we subtract (16) from (14) to get

$$R(Q^{'}) - J(\mathbf{U};Z) = \sum_{z\in\mathcal{Z}} (p_{\mathcal{B}}(z) - p_{Q^{'}}(z)) H_\psi(\mathbf{U}|Z=z)$$

By (8a), (11), and (15), the parenthesized difference on the RHS is non-negative. Thus,

$$\begin{aligned} &R(Q^{'}) - J(\mathbf{U};Z) \le \ln q \cdot \sum_{z\in\mathcal{Z}} (p_{\mathcal{B}}(z) - p_{Q^{'}}(z)) = \\ &\ln q \cdot \left[1 - \sum_{z\in\mathcal{Z}} p_{Q^{'}}(z)\right] = \ln q \cdot \frac{1}{q} \cdot \sum_{\mathbf{u}\in\mathcal{X}^t} \left[1 - \sum_{z\in\mathcal{Z}} Q^{'}(z|\mathbf{u})\right] = \\ &\ln q \cdot \frac{1}{q} \cdot \sum_{\mathbf{u}\in\mathcal{X}^t} \varepsilon_{\mathbf{u}} \ . \end{aligned}$$

Hence, by Lemma 12 we have a second bound:

$$R(Q^{'}) - J(\mathbf{U};Z) \le \ln q \cdot \frac{q(q-1)}{\mu} \ . \qquad (18)$$

The proof follows by adding the bounds (17) and (18).

Our last task is to prove part (iii) of the theorem, which bounds the output alphabet size. Recall that $|\mathcal{Z}|$ is bounded by Lemma 7. Recalling that the number of boost symbols is bounded by $|K| \le |\mathcal{X}^t| = q$, the proof easily follows. ■

## REFERENCES

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3051–3073, 2009.

[2] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," arXiv:0908.0302v1, 2009.

[3] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," arXiv:1006.4255v1, 2010.

[4] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inform. Theory*, vol. 58, pp. 5437–5448, 2012.

[5] I. Tal, A. Sharov, and A. Vardy, "Constructing polar codes for non-binary alphabets and MACs," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2012)*, Cambridge, Massachusetts, 2012, pp. 2132–2136.

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

[7] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inform. Theory*, vol. 57, pp. 6428–6443, 2011.

[8] E. Hof and S. Shamai, "Secrecy-achieving polar-coding for binary-input memoryless symmetric wire-tap channels," arXiv:1005.2759v2, 2010.

[9] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Commmun. Lett.*, vol. 14, pp. 752–754, 2010.

[10] O. O. Koyluoglu and H. El Gamal, "Polar coding for secure transmission and key agreement," in *Proc. IEEE Intern. Symp. Personal Indoor and Mobile Radio Comm.*, Istanbul, Turkey, 2010, pp. 2698–2703.

[11] A. Ghayoori and T. A. Gulliver, "Upgraded approximation of non-binary alphabets for polar code construction," arXiv:1304.1790v3, 2013.

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

[13] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2011)*, Saint Petersburg, Russia, 2011, pp. 11–15.

[14] U. Pereg and I. Tal, "Channel upgradation for non-binary input alphabets and MACs," arXiv:1308.5793v1, 2013.

[15] E. Şaşoğlu, "Polar codes for discrete alphabets," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2012)*, Cambridge, Massachusetts, 2012, pp. 2137–21 141.

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