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—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 2 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 2. Section V is devoted to proving Theorem 2.

II. PRELIMINARIES

A. Multiple Access Channel

Let \( W : \mathcal{X}^t \rightarrow \mathcal{Y} \) designate a generic \( t \)-user MAC, where \( \mathcal{X} = \{0, 1, \ldots, p - 1\} \) is the input alphabet, \( p \) is a positive integer, \( \mathcal{Y} \) is the finite output alphabet. Denote a vector of user inputs by \( \mathbf{u} \in \mathcal{X}^t \), where \( \mathbf{u} = (u^{(i)})_{i=1}^q \). 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 and upgraded single-user channel, respectively. That is, we say that a \( t \)-user MAC \( Q : \mathcal{X}^t \rightarrow \mathcal{Z} \) is degraded with respect to

\footnote{Following the observation in [14], we do not constrain ourselves to an input alphabet which is prime.}

\footnote{The assumption that \( \mathcal{Y} \) is finite is only meant to make the presentation simpler. That is, our method readily generalizes to the continuous output alphabet case.}
Let the random variables $U$ and $Y$ be as in Lemma 1. Assume that the output alphabet $Y$ has been purged of all letters $y$ with zero probability, since these outputs never occur. That is, assume $W(y|u) > 0$ for at least one $u \in X^t$. Thus we can indeed define the function $\varphi_W: X^t \times Y \rightarrow [0,1]$ as the a posteriori probability (APP):

$$\varphi_W(u|y) = \mathbb{P}(U = u|Y = y) = \frac{W(y|u)}{\sum_{v \in Y} W(v|y)},$$

for every input $u \in X^t$ and every letter in the (purged) output alphabet $y \in Y$. Next, for $y \in Y$ let us denote

$$p_W(y) = \mathbb{P}(Y = y),$$

and define $\eta: [0,1] \rightarrow \mathbb{R}$ by

$$\eta(x) = -x \cdot \ln x,$$

where $\ln(x)$ stands for natural logarithm. Using the above notation, the entropy of the input $U$ given the observation $Y = y$ is

$$H(U|Y = y) = \sum_{u \in X^t} \eta(\varphi_W(u|y)),$$

measured in natural units (nats). Thus, the sum-rate can be expressed as

$$R(W) = \ln q - \sum_{y \in Y} p_W(y) H(U|Y = y) = \ln q - \sum_{y \in Y} p_W(y) \sum_{u \in X^t} \eta(\varphi_W(u|y)).$$

As a first step towards the definition of our bins, we quantize the domain of $\eta(x)$ with resolution specified by a fidelity parameter $\mu$. That is, we partition $[0,1]$ into quantization-regions which depend on the value of $\mu$. Informally, we enlarge the width of each region until an increment of $1/\mu$ is reached, either on the horizontal or the vertical axis. To be exact, the interval $[0,1]$ is partitioned into $M = M_\mu$ non-empty regions of the form

$$[b_i, b_{i+1}) \quad , \quad i = 1, 2, \ldots, M.$$

Starting from $b_1 = 0$, the endpoint of the $i$th region is given by

$$b_{i+1} = \max \left\{ x \leq b_i + \frac{1}{\mu} \mid \eta(x) - \eta(b_i) \leq \frac{1}{\mu} \right\} \quad (2)$$

And so it is easily inferred that for all regions $1 \leq i < M$ (all regions but the last), there is either a horizontal or vertical increment of $1/\mu$:

$$|b_{i+1} - b_i| = \frac{1}{\mu} \quad , \quad \text{or} \quad |\eta(b_{i+1}) - \eta(b_i)| = \frac{1}{\mu},$$

but typically not both (Figure 1). For technical reasons, we will henceforth assume that

$$\mu \geq \max(5, q(q - 1)). \quad (3)$$

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 \( R(1) = M \).

Based on the quantization regions defined above, we define our binning rule. Two output letters \( y_1, y_2 \in Y \) are said to be in the same bin if for all \( u \in X^t \) we have that \( R(\varphi_W(u|y_1)) = R(\varphi_W(u|y_2)) \). That is, \( y_1 \) and \( y_2 \) share the same vector of region-indices,

\[
(i(u))_{u \in X^t} ,
\]

where \( i(u) \triangleq R(\varphi_W(u|y_1)) = R(\varphi_W(u|y_2)) \).

B. Merging of letters in the same bin

Recall that our ultimate aim is to approximate the channel \( W : X^t \to Y \) 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 \( Z \).

Figuratively, we think of \( Z \) as the result of merging together all the letters in the same bin. That is, the size of \( Z \) is the number of non-empty bins, as each non-empty bin corresponds to a distinct letter \( z \in Z \). Denote by \( B(z) \) the set of letters in \( Y \) which form the bin associated with \( z \). Thus, all the symbols \( y \in B(z) \) can be thought of as having been merged into one symbol \( z \).

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

C. The APP measure \( \psi \)

In this subsection, we define an a posteriori probability measure on the input alphabet \( X^t \), given a letter from the merged output alphabet \( Z \). We denote this APP measure as \( \psi(u|z) \), defined for \( u \in X^t \) and \( z \in Z \).

The measure \( \psi(u|z) \) will be used in Section IV in order to define the approximating channel. As we have previously mentioned, the output alphabet of the approximating channel will contain \( Z \). As we will see, \( \psi(u|z) \) will equal the APP of the approximating channel, for output letters \( z \in Z \).

For each bin define the leading input as

\[
u^* = \arg\max_{y \in \mathcal{B}(z)} \phi_W(u|y),
\]

where ties are broken arbitrarily. For \( z \in Z \), let

\[
\psi(u|z) = \min_{y \in \mathcal{B}(z)} \phi_W(u|y), \quad \text{for all } u \neq \nu^* ,
\]

and

\[
\psi(u^*|z) = 1 - \sum_{u \neq \nu^*} \psi(u|z).
\]

Informally, we note that if the bins are “sufficiently narrow” (if \( \mu \) is sufficiently large), then \( \psi(u|z) \) is close to \( \phi_W(u|y) \), for all \( u \in X^t \), \( z \in Z \), and \( y \in \mathcal{B}(z) \). The above will be made exact in Lemma 10 below.

IV. THE UPGRADED APPROXIMATION

Now we are in position to define our \( t \)-user MAC approximation \( Q : X^t \to (Z \cup K) \), where \( K \) is a set of additional symbols to be specified in this section. We refer to these new symbols as “boost” symbols.

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

\[
\alpha_u(y) \triangleq \frac{\psi(u|z)}{\phi_W(u|y)} \frac{\phi_W(u^*|y)}{\psi(u^*|z)}, \quad \text{if } \phi_W(u|y) \neq 0 .
\]

Otherwise, define

\[
\alpha_u(y) \triangleq 1, \quad \text{if } \phi_W(u|y) = 0 .
\]
By Lemma 13 in the next section, \( \alpha_u(y) \) is indeed well defined and is between 0 and 1. Next, for \( u \in X^t \), let
\[
\varepsilon_u \triangleq \sum_{y \in Y} (1 - \alpha_u(y)) W(y/u) .
\] (7)

We now define \( K \), the set of output “boost” symbols. Namely, we define a boost symbol for each non-zero \( \varepsilon_u \).
\[ K = \{ \kappa_u : u \in X^t, \varepsilon_u > 0 \} \] (8)

Lastly, the probability function \( Q' \) of our upgraded MAC is defined as follows. With respect to non-boost symbols, define for all \( z \in Z \) and \( u \in X^t \),
\[
Q'(z|u) = \sum_{y \in B(z)} \alpha_u(y) W(y/u) .
\] (9a)

With respect to boost symbols, define for all \( \kappa_v \in K \) and \( u \in X^t \),
\[
Q'(\kappa_v|u) = \begin{cases} 
\varepsilon_u & \text{if } u = v, \\
0 & \text{otherwise} .
\end{cases}
\] (9b)

Note that if a boost symbol \( \kappa_u \) is received at the output of \( Q' : X^t \rightarrow (Z \cup K) \), we know for certain that the input was \( U = u \).

The following theorem presents the properties of our upgraded approximation of \( W \). The proof concludes Section IV.

**Theorem 2.** Let \( W : X^t \rightarrow Y \) be a \( t \)-user MAC, and let \( \mu \) be a given fidelity parameter that satisfies (3). Let \( Q' : X^t \rightarrow (Z \cup K) \) be the MAC obtained from \( W \) by the above definition (9).

(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 2 claims that our upgraded-approximation has a sum-rate deviation of \( O(1/\mu) \), and an output-alphabet of size \( O(\mu^q) \).

V. ANALYSIS

Conceptually, for the purpose of analysis, the algorithm can be thought of as performing four steps. In the first step, an output alphabet \( Z \) is defined (Subsection III-B) by means of a quantization (Subsection III-A). In the second step, a corresponding APP measure \( \psi \) is defined (Subsection III-C). In the third step, the original output alphabet \( Y \) is augmented with “boost” symbols \( K \), and a new channel \( W' : X^t \rightarrow (Y \cup K) \) is defined. The APP measure \( \psi \) has a key role in defining \( W' \), which is upgraded with respect to \( W \). In the fourth step, we consolidate equivalent symbols in \( W' : X^t \rightarrow (Y \cup K) \) into a single symbol. The resulting channel is \( Q' : X^t \rightarrow (Z \cup K) \). On the one hand, \( Q' \) is equivalent to \( W' \), and thus upgraded with respect to the original channel \( W \). On the other hand, the output alphabet of \( Q' \) turns out to be \( Z \cup K \), a set typically much smaller than the original output alphabet \( Y \). The channels used throughout the analysis are depicted in Figure 2 along with the corresponding properties and the relations between them.

We now examine the algorithm step by step, and state the relevant lemmas and properties for each step. This eventually leads up to the proof of Theorem 2.

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.

Observing Figure 1, the reader may have noticed that regions entirely to the left of \( x = \frac{1}{e} \) have a vertical increment of \( \frac{1}{\mu} \). On the other hand, regions entirely to the right of \( x = \frac{1}{e} \), last region excluded, have a horizontal width of \( \frac{1}{\mu} \). The following lemma shows that this is always the case.

**Lemma 3.** 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 and (2). Thus,

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

(ii) Otherwise, if \( \frac{1}{e} \leq b_i < b_{i+1} < 1 \), then
\[
b_{i+1} - b_i = \frac{1}{\mu} .
\]

**Proof:** The derivative \( \eta'(x) = -(1 + \ln x) \) is strictly decreasing from \( +\infty \) at \( x = 0 \), to \( +1 \) at \( x = \frac{1}{e} \). Thus, for all \( 0 \leq x < \frac{1}{e^2} \),
\[
\eta'(x) > 1 .
\]

If \( 0 \leq b_i < b_{i+1} < \frac{1}{e^2} \), then we have by the fundamental theorem of calculus that
\[
\eta(b_{i+1}) - \eta(b_i) = \int_{b_i}^{b_{i+1}} \eta'(x) \, dx > \frac{1}{\mu} .
\]

Hence \( b_{i+1} < b_i + \frac{1}{\mu} \), which implies the first part of the lemma.

Moving forward on the \( x \)-axis, \( \eta'(x) \) keeps decreasing from \( +1 \) at \( x = \frac{1}{e} \), to \( -1 \) at \( x = 1 \). Thus for all \( \frac{1}{e} \leq x \leq 1 \),
\[
|\eta'(x)| \leq 1 .
\]
Hence, if \( \frac{1}{e} \leq b_i < b_{i+1} < 1 \), the second part follows by the triangle inequality:
\[
|\eta(b_{i+1}) - \eta(b_i)| \leq \int_{b_i}^{b_{i+1}} |\eta'(x)| \, dx \leq \frac{1}{\mu} .
\]

We are now ready to upper-bound \( M = M_\mu \), the number of quantization regions. The following corollary will be used to bound the number of bins, namely \( |Z| \), later on.

**Corollary 4.** The number of quantization regions, \( M = M_\mu \), satisfies
\[
M \leq 2\mu .
\]
such that

\[ \mu \]

where the last inequality follows from our assumption in (3).

Lemma 5. \( (\)second (braced) term is due to regions entirely within \([0, \frac{1}{2}]\), the second (braced) term is due to regions entirely within \([\frac{1}{2}, 1]\), where the 1 inside the braces is due to the last (rightmost) region. The 1 outside the brace is due to the possibility of a region that crosses \( x = \frac{1}{2} \). Hence, since \( \eta(1/e^2) = 2/e^2 \),

\[ M \leq \left( 1 + \frac{1}{e^2} \right) + 2 \leq 2\mu, \]

where the last inequality follows from our assumption in (5) that \( \mu \geq 5 \).

The corollary, following the lemma below, will play a significant role in the proof of Theorem 2. The lemma is proved in the appendix.

Lemma 5. 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 5.

Corollary 6. 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. This lemma is proved in the appendix as well.

Lemma 7. Let the width of the \( i \)th quantization interval be denoted by

\[ \Delta_i = b_{i+1} - b_i, \quad i = 1, 2, \ldots, M. \]

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 8. Let \( \mathcal{Z} \) be defined as in Section III-B. Then,

\[ |\mathcal{Z}| \leq (2\mu)^q. \]

Proof: The following proof is attained by essentially the same technique as in [5, Lemma 6]. The size of the merged output alphabet \( |\mathcal{Z}| \) is in fact the number of non-empty bins. Recall that two letters \( y_1, y_2 \in \mathcal{Y} \) are in the same bin if and only if \( \mathcal{R}(\phi_{W}(u|y_1)) = \mathcal{R}(\phi_{W}(u|y_2)) \) for all \( u \in \mathcal{X} \). Suppose there are \( M \) quantization regions. Since the number of values \( u \) can take is \( q \),

\[ |\mathcal{Z}| \leq M^q. \]

The proof follows from Corollary 4.

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

\[ i(u) = i_y(u) \triangleq \mathcal{R}(\phi_{W}(u|y)) \quad (10) \]

Thus the bin can be characterized by the set of region-indices

\[ \mathcal{L}(z) = \{ i_z(u) : u \in \mathcal{X}^t \} \quad (11) \]

According to the following lemma, the largest index in \( \mathcal{L}(z) \) belongs to the leading input \( u^* \), defined in (4). In other words the leading input is in the leading region.

Lemma 9. Consider a given \( z \in \mathcal{Z} \). Let \( i(u) \) be given by (10) for all \( u \in \mathcal{X}^t \), and let \( u^* \) be as in (4). Then

\[ i(u^*) = \max \{ i(u) : u \in \mathcal{X}^t \} \quad (12) \]

Proof: Define the leading output \( y^* \in \mathcal{B}(z) \) by

\[ y^* = y^*(z) \triangleq \arg \max_{y \in \mathcal{B}(z)} \phi_{W}(u^*|y) \quad (12) \]

By (4) and (12), we have that

\[ \phi_{W}(u^*|y^*) = \max_{y \in \mathcal{B}(z)} \phi_{W}(u^*|y) \quad (13) \]

Recalling the definition of our bins in Subsection III-A, we deduce that

\[ \mathcal{R}(\phi_{W}(u^*|y)) = \mathcal{R}(\phi_{W}(u^*|y^*)) \geq \mathcal{R}(\phi_{W}(u|y)) \quad (13) \]

for all \( y \in \mathcal{B}(z) \) and for all \( u \in \mathcal{X}^t \).
B. Properties of $\psi$

Recall that the APP measure $\psi(u|z)$ was defined in Subsection III.C. We start this subsection by showing that $\psi$ is “close” to the APP of the original channel.

Lemma 10. Let $W : \mathcal{X}^t \to \mathcal{Y}$ be a generic $t$-user MAC, and let $Z$ be the merged output alphabet conceived through applying the binning procedure to $Y$. For each $z \in Z$, let $u^* = u^*(z)$ be the leading-input defined by (5), and let $\psi(u|z)$ be the probability measure on $u \in \mathcal{X}^t$ defined in (5).

Then for all $z \in Z$ and $y \in B(z)$,

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

Proof: Consider a particular letter $y \in B(z)$. For all $u \neq u^*$, we have by (5b) that $\psi(u|z)$ belongs to the same quantization interval as $\varphi_W(u|y)$. Therefore, the first case arises due to Corollary 6.

As for the second case, let $\{\Delta_i\}_{i=1}^M$ be as in Lemma 7. Also, for the leading region $i^* = i(u^*)$, define the leading width by

$$\Delta^* = \Delta_{i^*}.$$  

As Lemma 9 declares the leading region to be the rightmost region in $Z$, it follows from Lemma 7 that either

$$i^* = M \quad \text{or} \quad \Delta^* = \max \{\Delta_i : i \in L(z)\}.$$  

In words, the leading region is either the last region or the widest.

Suppose first that $i^* < M$. Thus, the leading width is the largest. And so we claim that for all $u \neq u^*$,

$$0 \leq \varphi_W(u|y) - \psi(u|z) \leq \Delta_{i} \leq \Delta^*,$$

where $i = i(u) = R(\varphi_W(u|y))$. The leftmost inequality follows from (5a), while the middle follows from $\psi(u|z)$ and $\varphi_W(u|y)$ belonging to the same quantization interval. The rightmost inequality follows from our observation that $\Delta^* = \max \{\Delta_i : i \in L(z)\}$. Based on (5b), the above implies that

$$0 \leq \psi(u^*|z) - \varphi_W(u^*|y) \leq (q - 1)\Delta^*. \quad \text{(14)}$$

That is, $u^*$ may have been “pushed” several regions higher: $R(\psi(u^*|z)) \geq R(\varphi_W(u^*|y))$. However, Lemma 7 assures that $\Delta^*$ is no bigger than the width of subsequent regions. Thus

$$R(\psi(u^*|z)) - R(\varphi_W(u^*|y)) \leq q - 1,$$

from which the second part of the lemma follows by induction, applying Lemma 5.

If, on the other hand, $i^* = M$, then $\psi(u^*|z)$ must also belong to the last (and leading) region. The second part of the lemma follows then from Corollary 6. \hfill \blacksquare

The quantity $\psi(u^*|z)$ frequently appears as a denominator. The main use of the following lemma is to show that such an expression is well defined.

Lemma 11. For $z \in Z$, let $u^* = u^*(z)$ be the leading-input defined by (5), and let $\psi(u|z)$ be the probability measure on $u \in \mathcal{X}^t$ defined in (5). Then,

$$\psi(u^*|z) \geq \frac{1}{q}, \quad \text{(15)}$$

for all $z \in Z$.

Proof: Consider a given $z \in Z$. Let the leading-output $y^* \in B(z)$ be as in (12). On the one hand, since the sum of $\varphi_W(u|y^*)$ over $u \in \mathcal{X}^t$ is 1, there exists a $u \in \mathcal{X}^t$ such that

$$\varphi_W(u|y^*) \geq \frac{1}{q}. \quad \text{(16)}$$

On the other hand, by (13), we have that

$$\psi(u^*|y^*) \geq \varphi_W(u^*|y^*).$$

Thus,

$$\psi(u^*|z) \geq \varphi_W(u^*|y^*) \geq \frac{1}{q}, \quad \text{(17)}$$

where the leftmost inequality follows by (4). Let $z \in Z$ and $y \in B(z)$ be given. We will shortly make use of the quantity

$$\gamma(y) \triangleq \frac{\psi(u^*|y)}{\psi(u^*|z)}. \quad \text{(18)}$$

Note that by (15), $\gamma(y)$ is indeed well defined. Next, we claim that

$$\psi(u^*|z) \geq \varphi_W(u^*|y) \geq \frac{1}{q} - \frac{1}{\mu} > 0. \quad \text{(19)}$$

To justify this claim, note that the leftmost inequality follows from (16) and (17). The middle inequality follows from (2) and (17) (recall that $R(\varphi_W(u^*|y)) = R(\varphi_W(u^*|y^*))$ for all $y \in B(z)$). Finally, the rightmost inequality follows from (3). Therefore,

$$0 \leq \gamma(y) \leq 1. \quad \text{(20)}$$

Recall that by Lemma 10, we have that $\psi$ is close to the APP of the original channel, $\varphi_W$, in an additive sense (for large enough $\mu$). The following lemma states that $\psi$ and $\varphi_W$ are close in a multiplicative sense as well, when we are considering $u^*$. The proof is given in the appendix.

Lemma 12. Let $W : \mathcal{X}^t \to \mathcal{Y}$ be a $t$-user MAC, and let $\gamma(y)$ be given by (18). Then for all $y \in \mathcal{Y}$,

$$0 \leq 1 - \frac{\varphi_W(u^*|y)}{\mu} \leq \gamma(y) \leq 1. \quad \text{(21)}$$

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 $\alpha_u(y)$, defined in (6). Thus, as a first step, we prove the following Lemma.

Lemma 13. Let $\alpha_u(y)$, be as in (6). Then, $\alpha_u(y)$ is well defined and satisfies

$$0 \leq \alpha_u(y) \leq 1. \quad \text{(22)}$$
Proof: The claim obviously holds if $\varphi_W(u|y) = 0$ due to (6b). So, we henceforth assume that $\varphi_W(u|y) > 0$, and thus have that

$$\alpha_u(y) = \frac{\psi(u|z)}{\varphi_W(u|y)} \cdot \frac{\varphi_W(u|y)}{\psi(u|z)} .$$

By assumption, the first denominator is positive. Also, by (15), the second denominator is positive, and thus $\alpha_u(y)$ is indeed well defined.

We now consider two cases. If $u = u^*$, then $\alpha_u(y) = 1$, and the claim is obviously true. Thus, assume that $u \neq u^*$. Since we are dealing with probabilities, we must have that $\alpha_u(y) \geq 0$. Consider the two fractions on the RHS of (23).

By (5a), the first fraction is at most 1, and by (19) the second fraction is at most 1. Thus, $\alpha_u(y)$ is at most 1.

We now define $W': \mathcal{X} \to (\mathcal{Y} \cup \mathcal{K})$, an upgraded version of $W$. For all $y \in \mathcal{Y}$ and for all $u \in \mathcal{X}'$, define

$$W'(y|u) = \alpha_u(y) \cdot W(y|u) .$$

(24a)

Whereas, for all $\kappa \in \mathcal{K}$ and for all $u \in \mathcal{X}'$, define

$$W' (\kappa|u) =
\begin{cases}
\varepsilon_u = \sum_{y \in \mathcal{Y}} (1 - \alpha_u(y))W(y|u) & \text{if } u = \nu, \\
0 & \text{otherwise},
\end{cases}

(24b)

The following lemma states that $W'$ is indeed an upgraded version of $W$.

Lemma 14. Let $W : \mathcal{X} \to \mathcal{Y}$ be a $t$-user MAC, and let $W' : \mathcal{X} \to (\mathcal{Y} \cup \mathcal{K})$ be the MAC obtained by the procedure above. Then, $W'$ is well-defined and is upgraded with respect to $W$. That is,

$$W' \succeq W .$$

Proof: Based on Lemma 13, it can be easily verified that $W'$ is indeed well-defined. We define the following intermediate channel $\mathcal{P} : (\mathcal{Y} \cup \mathcal{K}) \to \mathcal{Y}$, and prove the lemma by showing that $W$ is obtained by the concatenation of $W'$ followed by $\mathcal{P}$. Define for all $y \in \mathcal{Y}$ and for all $y' \in (\mathcal{Y} \cup \mathcal{K})$,

$$\mathcal{P}(y|y') =
\begin{cases}
1 - \alpha_u(y)\cdot W(y|u) & \text{if } y' = y \in \mathcal{Y}, \\
0 & \text{if } y' = \kappa_u \in \mathcal{K},
\end{cases}
$$

Let $y \in \mathcal{Y}$ and $u \in \mathcal{X}$ be given. Now consider the sum

$$\sum_{y' \in \mathcal{Y} \cup \mathcal{K}} W'(y'|u) \cdot \mathcal{P}(y|y') =
W'(y|u) \cdot 1 + \sum_{\kappa \in \mathcal{K}} W'(\kappa|u) \cdot \mathcal{P}(y|\kappa) .$$

Consider first the case in which $\varepsilon_u = 0$. In this case, the sum term, in the RHS, is zero (see (8)). Moreover, (6b) and (7) imply that $\alpha_u(y) = 1$. And so we have, by (24a), that

$$\sum_{y' \in \mathcal{Y} \cup \mathcal{K}} W'(y'|u) \cdot \mathcal{P}(y|y') = W(y|u) .$$

Next, consider the case where $\varepsilon_u > 0$. We have that

$$\sum_{y' \in \mathcal{Y} \cup \mathcal{K}} W'(y'|u) \cdot \mathcal{P}(y|y') = W'(y|u) + \varepsilon_u \cdot \mathcal{P}(y|\kappa_u) = \alpha_u(y)W(y|u) + [1 - \alpha_u(y)] \cdot W(y|u) = W(y|u) .$$

A boost symbol carries perfect information about what was transmitted through the channel. We now bound from above the average probability of receiving a boost symbol. This result will be useful in the proof of Theorem 2, where we bound the sum-rate increment of our upgraded approximation.

Lemma 15. Let $\varepsilon_u$ be given by (7) for all $u \in \mathcal{X}'$. Then,

$$\frac{1}{q} \sum_{u \in \mathcal{X}'} \varepsilon_u \leq \frac{q(q - 1)}{\mu} .$$

Proof: By definition (7), we have that

$$\frac{1}{q} \sum_{u \in \mathcal{X}'} \varepsilon_u = \frac{1}{q} \sum_{u \in \mathcal{X}'} \sum_{y \in \mathcal{Y}} (1 - \alpha_u(y))W(y|u) = 1 - \frac{1}{q} \sum_{u \in \mathcal{X}'} \sum_{y \in \mathcal{Y}} \alpha_u(y)W(y|u) .$$

(25)

We now bound the second term. We have that

$$\frac{1}{q} \sum_{u \in \mathcal{X}'} \sum_{y \in \mathcal{Y}} \alpha_u(y)W(y|u) =
\frac{1}{q} \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{X}' : W(y|u) > 0} \alpha_u(y)W(y|u) =
\frac{1}{q} \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{X}' : W(y|u) > 0} \psi(u|z) \cdot \varphi_W(u|y) \cdot \gamma(y) \cdot W(y|u) \geq
\frac{1}{q} \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{X}' : \varphi_W(u|y) > 0} \psi(u|z) \cdot \varphi_W(u|y) \cdot \gamma(y) \cdot W(y|u) =
\frac{1}{q} \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{X}' : \varphi_W(u|y) > 0} \psi(u|z) \cdot \varphi_W(u|y) \cdot \gamma(y) \cdot W(y|u) =
\frac{1}{q} \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{X}' : \varphi_W(u|y) > 0} \psi(u|z) \cdot \varphi_W(u|y) \cdot \gamma(y) \cdot W(y|u) =
1 - \frac{q(q - 1)}{\mu} .$$

(26)

where the inequality is due to Lemma 12 and the equality that follows it is due to the observation below. If $\varphi_W(u|y) = 0$, then based on (19), we have that $u \neq u^*$. Therefore, by (5a), $\varphi_W(u|y) = 0$ implies that $\psi(u|z) = 0$ as well. That in turn leads to our observation that

$$\sum_{u \in \mathcal{X}' : \varphi_W(u|y) > 0} \psi(u|z) = \sum_{u \in \mathcal{X}'} \psi(u|z) = 1 .$$

(27)

As the second term of (25) is bounded by (26), the proof follows.
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 16.** Let \( W' : \mathcal{X}^t \to (\mathcal{Y} \cup K) \) be the MAC defined in Subsection \( \ref{section:appendix} \). Then, all the output letters \( y \in B(z) \) have the same modified APP values (for each \( u \in \mathcal{X}^t \) separately). Namely, \[ \varphi_{W'}(u|y) = \psi(u|z) , \]

for all \( u \in \mathcal{X}^t \), and for all \( z \in \mathcal{Z} \) and \( y \in B(z) \).

**Proof:** First consider the case where \( \varphi_{W'}(u|y) = 0 \). On the one hand, \( \varphi_{W'}(u|y) = 0 \) by \( \ref{eq:phi_W} \) and \( \ref{eq:psi_z} \). On the other hand, \( \ref{eq:psi_z} \) implies that \( u \neq u^* \), and thus \( \psi(u|z) = 0 \) as well, by \( \ref{eq:psi_u} \).

Now assume \( \varphi_{W'}(u|y) > 0 \). In that case,

\[
\varphi_{W'}(u|y) = \sum_{v \in \mathcal{X}^t} \frac{W'(y|u) \alpha_{y}(y|v)}{\sum_{v \in \mathcal{X}^t} W'(y|v)} = \frac{\sum_{v \in \mathcal{X}^t} W'(y|u) \alpha_{y}(y|v) \gamma(y) \cdot \psi(u|z)}{\sum_{v \in \mathcal{X}^t} W'(y|v) \gamma(y) \cdot \psi(v|z)} = \frac{\psi(u|z)}{\sum_{v \in \mathcal{X}^t} \psi(v|z)} = \psi(u|z) ,
\]

where the fourth equality follows from \( \ref{eq:psi_z} \).

We have seen in Lemma \( \ref{lemma:modified_App} \) that with respect to \( W' \), all the members of \( B(z) \) have the same APP values. As will be pointed in Lemma \( \ref{lemma:modified_App} \) in the sequel, consolidating symbols with equal APP values results in an equivalent channel. Thus consolidating all the members of every bin \( B(z) \) to one symbol \( z \) results in an equivalent channel \( Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K) \) defined by \( \ref{eq:Q_prime} \). Note that consolidation simply means mapping all the members of \( B(z) \) to \( z \) with probability 1. Formally, we have for all \( z \in \mathcal{Z} \cup K \) and for all \( u \in \mathcal{X}^t \),

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

Based on \( \ref{eq:psi_z} \), 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 \( \ref{eq:Q} \).

The rest of this section is dedicated to proving Theorem \( \ref{theorem:MAC_equivalence} \). But before that, we address the equivalence of \( W' \) and \( Q \) in Lemma \( \ref{lemma:equivalence} \) which is proved in the appendix. In essence, we claim afterward that due to this equivalence, showing that \( W' \succeq W \) implies that \( Q' \succeq W \).

**Lemma 17.** Let \( W : \mathcal{X}^t \to \mathcal{Y} \) be a t-user MAC, and let \( y_1, \ldots, y_r \in \mathcal{Y} \) be \( r \) letters of equal APP values, for some positive integer \( r \). That is, for all \( u \in \mathcal{X}^t \),

\[
\varphi(u|y_i) = \varphi(u|y_j) , \quad \text{for all } 1 \leq i \leq j \leq r .
\]

Now let \( Q : \mathcal{X}^t \to \mathcal{Z} \) be the t-user MAC obtained by consolidating \( y_1, \ldots, y_r \), to one symbol \( z \). This would make the output alphabet

\[
\mathcal{Z} = \mathcal{Y} \setminus \{ y_1, \ldots, y_r \} \cup \{ z \} .
\]

Then, \( W \equiv Q \) (the MACs \( W \) and \( Q \) are equivalent).

We have mentioned that equivalence of MACs is a transitive relation. Therefore, consolidating bin after bin we finally have by induction that \( W' \equiv Q' \).

**Proof of Theorem 2**

We first prove part \( \ref{part:MAC_equivalence} \) of the theorem, which claims that the approximation is well defined and upgraded with respect to \( W \). 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 \( \ref{lemma:equivalence} \), \( W' \succeq W \). Since \( W' \) and \( Q' \) are equivalent, and since upgradeation transitivity immediately follows from the definition, it follows that \( Q' \succeq W \).

We now move to part \( \ref{part:MAC_equivalence} \) 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 \( U \). Similarly, define \( Z' \) as the output of \( Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K) \) when the input is \( U \).

To estimate the APPs for \( Q' : \mathcal{X}^t \to (\mathcal{Z} \cup K) \), we may use \( \ref{eq:phi_W} \) and \( \ref{eq:psi_z} \). By Lemma \( \ref{lemma:modified_App} \) we have that

\[
\varphi_{Q'}(u|z) = \psi(u|z) ,
\]

for all \( u \in \mathcal{X}^t \) and for all \( z \in \mathcal{Z} \) (for all non-boost output symbols). Whereas for boost-symbols, \( \varphi_{Q'}(u|\kappa) \in [0,1] \) for all \( u \in \mathcal{X}^t \) and for all \( \kappa \in \mathcal{K} \). Denote the entropy of the probability distribution defined in Section \( \ref{section:entropy} \) by

\[
H_{\psi}(U|Z = z) = \sum_{u \in \mathcal{X}^t} \eta(\psi(u|z)) .
\]

Thus

\[
R(Q') = \ln q - \sum_{z \in \mathcal{Z}} p_{Q'}(z) H_\psi(U|Z = z) - \sum_{\kappa \in \mathcal{K}} p_{Q'}(\kappa) H(U|Z' = \kappa) .
\]

However, the last term is zero due to the following observation. Given that the output of the MAC \( Q' \) is \( \kappa \psi \) for some \( \psi \in \mathcal{X}^t \),
the input \( U \) is known to be \( v \) (it is deterministic). Hence

\[
H(U | Z') = \kappa_v = 0 \text{ for all } \kappa_v \in K. \text{ Hence}
\]

\[
R(Q') = \ln q - \sum_{z \in Z} p_{Q'(z)} H_{\psi}(U | Z = z). \tag{31}
\]

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 \( Z \) corresponding to \( Y \). Namely, the realization \( Z = z \) occurs whenever \( Y \) is contained in \( B(z) \). The probability of that realization is

\[
p_B(z) \triangleq p(Z = z) = \sum_{y \in B(z)} p_W(y). \tag{32}
\]

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

\[
J(U; Z) \triangleq \ln q - \sum_{z \in Z} p_B(z) \cdot H_{\psi}(U | Z = z)
= \ln q - \sum_{z \in Z} p_B(z) \sum_{u \in \mathcal{X}^t} \eta[\psi(u | z)], \tag{33}
\]

where \( H_{\psi}(U | Z = z) \) is given by \( (30) \).

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

\[
J(U; Z) - R(W) = \sum_{y \in \mathcal{Y}} p_W(y) \sum_{u \in \mathcal{X}^t} \eta(\psi_W(u | y))
- \sum_{z \in Z} p_B(z) \sum_{u \in \mathcal{X}^t} \eta(\psi(u | z)) = \tag{34}
\]

\[
\sum_{z \in Z} \sum_{y \in \mathcal{B}(z)} p_W(y) \sum_{u \in \mathcal{X}^t} \eta(\psi_W(u | y)) - \eta(\psi(u | z)) \leq
\]

\[
\sum_{z \in Z} \sum_{y \in \mathcal{B}(z)} p_W(y) \cdot \eta(\psi_W(u^* | y)) - \eta(\psi(u^* | z)) +
\]

\[
2 \cdot \frac{q - 1}{\mu},
\]

where the last inequality is due to Lemma \[10\].

For the second bound, we subtract \( (33) \) from \( (31) \) to get

\[
R(Q') - J(U; Z) = \sum_{z \in Z} (p_B(z) - p_{Q'(z)}) H_{\psi}(U | Z = z).
\]

By \( (9a), (22) \), and \( (32) \), the parenthesized difference on the

RHS is non-negative. Thus,

\[
R(Q') - J(U; Z) \leq \ln q \cdot \sum_{z \in Z} (p_B(z) - p_{Q'(z)}) = \ln q \cdot \left[ 1 - \sum_{z \in Z} p_{Q'(z)} \right] = \ln q \cdot \frac{1}{q} \sum_{u \in \mathcal{X}^t} \left[ 1 - \sum_{z \in Z} Q'(z | u) \right] = \ln q \cdot \frac{1}{q} \sum_{u \in \mathcal{X}^t} \varepsilon_u.
\]

Hence, by Lemma \[15\] we have a second bound:

\[
R(Q') - J(U; Z) \leq \ln q \cdot \frac{q(q - 1)}{\mu}. \tag{35}
\]

The proof follows by adding the bounds \( (34) \) and \( (35) \).

Our last task is to prove part (ii) of the theorem, which bounds the output alphabet size. Recall that \( |Z| \) is bounded by Lemma \[8\]. Recalling that the number of boost symbols is bounded by \( |K| \leq |\mathcal{X}^t| = q, \) the proof easily follows.

APPENDIX

**Proof of Lemma \[3\]** Let \( x, i \) and \( \delta \) be as in Lemma \[3\]. If \( x \) is in the last region, then the lemma simply follows from the definition in \( \[2\] \). So, suppose \( i < M \), and let

\[
\Delta = b_{i+1} - b_i \leq \frac{1}{\mu}, \tag{36}
\]

where the inequality follows from \( \[2\] \).

We now consider two cases. If \( \frac{1}{e^2} < x \leq 1 \), then \( |\eta'(x)| \leq 1 \). Thus, by the triangle inequality,

\[
|\eta(x + \delta) - \eta(x)| \leq \int_x^{x + \delta} |\eta'(\xi)| \, d\xi \leq \delta \leq \Delta = \frac{1}{\mu},
\]

where the equality follows by part (ii) of Lemma \[3\].

In the other case left to consider, \( 0 < x < \frac{1}{e^2} \). Recall that \( \mu \geq 5 \) by the assumption made in \( \[3\] \). Hence

\[
\frac{1}{\mu} < \frac{1}{e} - \frac{1}{e^2},
\]

which implies that

\[
x + \delta \leq x + \Delta \leq x + \frac{1}{\mu} < \frac{1}{e};
\]

Hence, the derivative function \( \eta' \) is positive in the range \( [x, x + \Delta] \). By the definition of \( \Delta \) in \( \[36\] \), we have that the point \( x + \Delta \) belongs to another region:

\[
b_{i+1} \leq x + \Delta < \frac{1}{e}.
\]

Thus, since \( \eta \) is strictly increasing in \( [0, \frac{1}{e}] \),

\[
|\eta(x + \delta) - \eta(x)| = \eta(x + \delta) - \eta(x)
\leq \eta(x + \Delta) - \eta(x)
= [\eta(b_{i+1}) - \eta(x)] + [\eta(x + \Delta) - \eta(b_{i+1})].
\]
Hence, by the fundamental theorem of calculus,
\[ |\eta(x + \delta) - \eta(x)| \leq \int_{x}^{x+\Delta} \eta'(\xi) \, d\xi + \int_{b_{i+1}}^{x+\Delta} \eta'(\xi) \, d\xi . \]
Since \( \eta'(x) \) is a strictly decreasing function of \( x \), the second integral can be upper-bounded by
\[ \int_{b_{i+1}}^{x+\Delta} \eta'(\xi) \, d\xi < \int_{b_{i+1} - \Delta}^{x} \eta'(\xi) \, d\xi . \]
By (36), we have that \( b_{i+1} - \Delta = b_{i} \). Thus,
\[ |\eta(x + \delta) - \eta(x)| \leq \int_{b_{i+1}}^{x} \eta'(\xi) \, d\xi + \int_{b_{i+1}}^{x} \eta'(\xi) \, d\xi = \eta(b_{i+1}) - \eta(b_{i}) \leq \frac{1}{\mu} , \]
where the last inequality follows from (37).

**Proof of Lemma 7.** Let us look at two quantization intervals \( i \) and \( j \), where \( 1 \leq i < j \leq M \). Our aim is to prove that \( \Delta_{j} \leq \Delta_{i} \). Consider first the simpler case in which \( \Delta_{j} = \frac{1}{\mu} \). Recall from (3) that \( \frac{1}{\mu} \) is an upper bound on the length of any interval, and specifically on \( \Delta_{i} \). Thus, in this case, \( \Delta_{i} \leq \Delta_{j} \).

Next, let us consider the case in which \( \Delta_{j} < \frac{1}{\mu} \). Thus, by (2), we must have that
\[ \eta(b_{j+1}) - \eta(b_{j}) = \frac{1}{\mu} . \] (37)
We will now assume to the contrary that \( \Delta_{j} < \Delta_{i} \), and show a contradiction to (37).

Since \( \Delta_{j} < \frac{1}{\mu} \), we must have by part (ii) of Lemma 3 that \( b_{j} < \frac{1}{\epsilon} \). Since every interval length is at most \( \frac{1}{\mu} \), we must have that \( \Delta_{i} < \frac{1}{\mu} \). By the above, and recalling the assumption in (3) that \( \mu \geq 5 \), we deduce that
\[ b_{j} + \Delta_{j} = b_{j} < b_{j} + \Delta_{i} \leq b_{j} + \frac{1}{\mu} < \frac{1}{\epsilon} + \frac{1}{\mu} < 1 , \]
Thus, since \( \eta'(x) \) is positive for \( x < \frac{1}{\epsilon} \),
\[ \eta(b_{j+1}) - \eta(b_{j}) = \int_{b_{j}}^{b_{j+1}} \eta'(\xi) \, d\xi < \int_{b_{j}}^{b_{j} + \Delta_{i}} \eta'(\xi) \, d\xi . \]
Now, since \( b_{i} < b_{j} \) and \( \eta'(x) \) is a strictly decreasing function of \( x \), we have that
\[ \int_{b_{j}}^{b_{j} + \Delta_{i}} \eta'(\xi) \, d\xi < \int_{b_{j}}^{b_{j} + \Delta_{i}} \eta'(\xi) \, d\xi = \eta(b_{i+1}) - \eta(b_{i}) . \]
Lastly, since \( b_{j} < \frac{1}{\epsilon} \), we have that \( b_{i+1} - \Delta_{j} < b_{j} \). Thus, by part (i) of Lemma 3 we have that
\[ \eta(b_{i+1}) - \eta(b_{j}) = \frac{1}{\mu} . \]
From the last three displayed equations, we deduce that
\[ \eta(b_{j+1}) - \eta(b_{j}) < \frac{1}{\mu} , \]
which contradicts (37).

**Proof of Lemma 12** We already know that \( \gamma(y) \leq 1 \), by (20). Thus, we now prove the lower bound on \( \gamma(y) \). To this end, we have by (2) and (14) that for all \( z \in \mathcal{Z} \) and \( y \in \mathcal{B}(z) \),
\[ \psi(u^{*}|z) = \mu_{w}(u^{*}|y) = (q - 1) \cdot \frac{1}{\mu} . \]
By (15), we can divide both sides of the above by \( \psi(u^{*}|z) \) and retain the inequality direction. The result is
\[ \varphi_{w}(u^{*}|y) = \frac{\psi(u^{*}|z)}{\psi(u^{*}|z)} \geq 1 - (q - 1) \cdot \frac{1}{\mu} , \]
where the last inequality yet again follows from (15). Thus, we have proved the lower bound on \( \gamma(y) \) as well. Since, by our assumption in (3), \( \mu \geq q - 1 \), the lower bound is indeed non-negative.

**Proof of Lemma 17** Let \( W, Q \) and \( y_{1}, \ldots, y_{r} \) be as in Lemma 17. We would like to show that \( W \) and \( Q \) satisfy both
\[ Q \preceq W \quad \text{and} \quad Q \succeq W . \]
It is obvious that \( Q \) is degraded with respect to \( W \). This is because \( Q \) is obtained from \( W \) by mapping with probability 1 one letter to another. The letters \( y_{1}, \ldots, y_{r} \) are mapped into \( z \), whereas the rest of the letters in \( \mathcal{Y} \) are mapped to themselves. We must now show that \( Q : \mathcal{X}^{t} \rightarrow \mathcal{Z} \) is upgraded with respect to \( W : \mathcal{X}^{t} \rightarrow \mathcal{Y} \). Namely, we must furnish an intermediate channel \( \mathcal{P} : \mathcal{Z} \rightarrow \mathcal{Y} \). Denote
\[ a_{i}(u) \triangleq W(y_{i}|u) = \frac{p_{w}(y_{i}|y_{i}|u)}{|_{/|_{/}}}, \]
\[ A(u) \triangleq Q(z|u) = \sum_{1 \leq i \leq r} a_{i}(u) , \]
for all \( u \in \mathcal{X}^{t} \). Note that by our running assumption on non-degenerate output letters, \( A(u) > 0 \). Given (29), we get that
\[ e_{i} \triangleq \frac{a_{i}(u)}{A(u)} \]
does not depend on \( u \). So the following is well-defined. Define for all \( y \in \mathcal{Y} \) and \( s \in \mathcal{Z} \)
\[ \mathcal{P}(y|s) = \begin{cases} e_{i} & \text{if } (y, s) = (y_{i}, z) \text{ for some } 1 \leq i \leq r \ , \\ 1 & \text{if } y = s \ , \\ 0 & \text{otherwise} . \end{cases} \]
Trivial algebra finishes the proof.

**REFERENCES**

[1] E. Arikan, “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. Arikan, “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. Shavrov, and A. Vardy, “Constructing polar codes for non-binary alphabets and MACs,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2012)*, Cambridge, Massachussets, 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, Y. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, “Nested polar codes for wiretap and relay channels,” IEEE Commun. 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,” submitted to 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] E. Şaşoğlu, “Polar codes for discrete alphabets,” in Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2012), Cambridge, Massachusetts, 2012, pp. 2137–2141.

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