Expurgated Bounds for the Asymmetric Broadcast Channel
Ran Averbuch, Nir Weinberger* and Neri Merhav
Department of Electrical Engineering,
Technion – Israel Institute of Technology,
Technion City, Haifa 3200003, Israel,
Email: nir.wein@gmail.com, {rans@campus, merhav@ee}.technion.ac.il

Abstract— This work contains two main contributions concerning the expurgation of hierarchical ensembles for the asymmetric broadcast channel. The first is an analysis of the optimal maximum likelihood (ML) decoders for the weak and strong user. Two different methods of code expurgation will be used, that will provide two competing error exponents. The second is the derivation of expurgated exponents under the generalized stochastic likelihood decoder (GLD). We prove that the GLD exponents are at least as tight as the maximum between the random coding error exponents derived in an earlier work by Averbuch and Merhav (2017) and one of our ML-based expurgated exponents. By that, we actually prove the existence of hierarchical codebooks that achieve the best of the random coding exponent and the expurgated exponent simultaneously for both users.

I. INTRODUCTION

One of the most elementary system configuration models in multi-user information theory is the broadcast channel (BC). It has been introduced more than four decades ago by Cover [1], and since then, a vast amount of papers and books, analyzing different aspects of the broadcast model, have been published. Although the characterization of the capacity region of the general BC is still an open problem, some special cases have been solved. Most notably, the broadcast channel with degraded message sets, also known as the asymmetric broadcast channel (ABC), was introduced and solved by Körner and Marton [2].

While the capacity region of the ABC has been known for many years, only little is known about its reliability functions. The earliest work on error exponents for the general ABC is of Körner and Sgarro [3]. Later, Kaspi and Merhav [4] have derived tighter lower bounds to the reliability functions of both users by analyzing random coding error exponents of their optimal decoders. Most recently [5], the exact random coding error exponents have been determined for both the strong user and the weak user, under the ensemble of fixed composition codes.

Even in the single–user case, it is known for many years that the random coding error exponent is not tight (with respect to the reliability function) for relatively low coding rates, and may be improved by expurgation [6], [10]. Specifically, improved bounds are obtained by eliminating codewords that contribute relatively highly to the error probability, and asserting that some upper bound holds for all remaining codewords.

The main objective of this paper is to study expurgation techniques for the hierarchical ensemble used over the ABC. Expurgating a code for the ABC is not a trivial extension of expurgation in the single-user case, because there might be conflicting goals from the viewpoints of the two users. Nonetheless, we were able to define expurgation procedures that guarantee no harm to the performance of either user. This has paved the way to derive tighter lower bounds on the reliability functions of the ABC.

We start by analyzing the optimal maximum likelihood (ML) decoder, and derive some expurgated bounds, that are natural generalizations of the single–user expurgated bound due to Csiszár, Körner and Marton (CKM) [6]. Although our first process of code expurgation is fairly intuitive, there is at least one specific step in our first derivation where exponential tightness might be compromised. This point gives rise to a possible room for improvement upon the results of our first theorem, and indeed, such an improvement is achieved by a second method of expurgation. Here, one starts by expurgating cloud centers, and only afterwards, single codewords. The intuition behind this technique is the following. When the exponential rate of the codewords within a cloud is too high, the weak user can still make a good estimation, merely by relying on the set of cloud centers. The expurgated bounds of our second method, however, are not always tighter than those of the first method, because of other differences in their derivations.

We then expand the scope and consider the generalized likelihood decoder (GLD), which is a more general family of stochastic likelihood decoders. For such decoders, the probability of deciding on a given message is proportional to a general exponential function of the joint empirical distribution of the cloud–center, the codeword and the received channel output vector. The random coding error exponent of the ordinary and the mismatched likelihood decoders for the single–user channel have been derived by Scarlett et al. [7]. In a more recent paper by Merhav [8], the expurgated exponent of the GLD has been derived and compared to the classical expurgated bound of [6], showing an explicit improvement at relatively high coding rates. In this paper, we consider GLD’s for both the strong and the weak users of an ABC, and derive...
expurgated exponents under these decoders. These bounds generalize the bound of [8], and prove that they are at least as tight as the maximum between the random coding error exponents of [5] and the expurgated bounds of our first theorem, which are based on the ML decoder. By that, we actually prove the existence of hierarchical codebooks that attain the best of the random coding exponent and the expurgated exponent simultaneously for both users. The main drawback of those error exponents is that they are not easy to calculate since they involve minimizations over relatively cumbersome auxiliary channels, and hence, efficient computation algorithms for the GLD bound are left for further research. From this viewpoint, the exponents of our first theorems are much more attractive.

Due to the space limitation, technical details and proofs are omitted, but can be found in the full version of this paper [11].

II. NOTATION CONVENTIONS

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Sources and channels will be subscripted by the names of the relevant random variables/vectors and their conditionings, whenever applicable, following the standard notation conventions, e.g., \( Q_X, Q_{XY} | X \), and so on. When there is no room for ambiguity, these subscripts will be omitted. For a generic joint distribution \( Q_{XY} = \{ Q_{XY}(x,y), x \in X, y \in Y \} \), which will often be abbreviated by \( Q \), information measures will be denoted in the conventional manner, but with a subscript \( Q \), that is, \( H_Q(X) \) is the marginal entropy of \( X \), \( I_Q(X;Y) \) is the mutual information between \( X \) and \( Y \), and so on. The weighted divergence between two conditional distributions (channels), say, \( Q_{Z|X} \) and \( W = \{ W(z|x), x \in X, z \in Z \} \), with weighting \( Q_X \) is defined as

\[
D(Q_{Z|X} || W|Q_X) = \sum_{x \in X} Q_X(x) \sum_{z \in Z} Q_{Z|X}(z|x) \log \frac{Q_{Z|X}(z|x)}{W(z|x)},
\]

(1)

where logarithms, here and throughout the sequel, are taken to the natural base. The probability of an event \( E \) will be denoted by \( \Pr(E) \), and the expectation operator with respect to a probability distribution \( Q \) will be denoted by \( \mathbb{E}_Q \{ \cdot \} \). The notation \( [x]_+ \) will stand for \( \max\{0, x\} \).

The type class of \( Q_U \), denoted by \( \mathcal{T}(Q_U) \), is the set of all vectors \( u \in U^n \) with \( \hat{P}_u = Q_U \), where \( \hat{P}_u \) is the empirical distribution of the sequence \( u \). Similarly, \( \mathcal{T}(Q_{XY} | Y) \) denotes the conditional type class, induced by the sequence \( y \) and the empirical conditional distribution \( Q_{X|Y} \).

III. DEFINITIONS AND PROBLEM FORMULATION

We consider a memoryless ABC with a finite input alphabet \( X \) and finite output alphabets \( Y \) and \( Z \). Let \( W_1 = \{ W_1(y|x), x \in X, y \in Y \} \) and \( W_2 = \{ W_2(z|x), x \in X, z \in Z \} \) denote the single-letter input-output transition probability matrices, associated with the strong user and the weak user, respectively. When these channels are fed by an input vector \( x \in X^n \), they produce the corresponding output vectors \( y \in Y^n \) and \( z \in Z^n \), according to \( W_1(y|x) = \prod_{t=1}^n W_1(y_t|x_t) \) and \( W_2(z|x) = \prod_{t=1}^n W_2(z_t|x_t) \). We are interested in sending one out of \( M_s \) common messages to both users, and one out of \( M_p \) private messages to the strong user, that observes \( y \). The two messages are chosen under the uniform distribution. Although our results prove the existence of a single sequence of deterministic hierarchical constant composition (HCC) codebooks, whose error probabilities are provably bounded, our proof techniques use extensively the following mechanism of random selection of an HCC code for the ABC. Let \( U \) be a finite alphabet, let \( P_U \) be a given probability distribution on \( U \), and let \( P_X|U \) be a given matrix of conditional probabilities of \( X \) given \( U \), such that the type-class \( \mathcal{T}(P_{X|U}) \) and the conditional type-class \( \mathcal{T}(P_X|U,u) \) are non-empty. We first select, independently at random, \( M_u = \lfloor e^{nR_u} \rfloor \) n-vectors (“cloud centers”), \( u_0, u_1, \ldots, u_{M_u-1} \), all under the uniform distribution over the type-class \( \mathcal{T}(P_{X|U}) \). Next, for each \( m = 0, 1, \ldots, M_u-1 \), we select conditionally independently (given \( u_m \)), \( M_z = \lfloor e^{nR_z} \rfloor \) codewords, \( x_{m,0}, x_{m,1}, \ldots, x_{m,(M_z-1)} \), under the uniform distribution across the conditional type-class \( \mathcal{T}(P_X|U,u_m) \). We denote the sub-code for each cloud by \( \mathcal{C}_m(n) = \{ x_{m,0}, x_{m,1}, \ldots, x_{m,(M_z-1)} \} \), or just \( \mathcal{C}_m \) in short. Thus, the communication rate to the weak user is \( R_z \), while the total communication rate to the strong user is \( R_s + R_y \). Once selected, the entire codebook \( \mathcal{C}(n) = \bigcup_{m=0}^{M_u-1} \mathcal{C}_m(n) \), and the collection of cloud centers, \( \{ u_0, u_1, \ldots, u_{M_u-1} \} \), are revealed to the encoder and to both decoders. We denote by \( \mathscr{C} \) a sequence of HCC codes, \( \{ \mathcal{C}(n), n = 1, 2, \ldots \} \).

For any of the following described decoding rules, denote by \( \tilde{m}(y,i(y)) \) the decoded pair of the strong user, and by \( \tilde{n}(z) \) the decoded cloud of the weak user. The ML decoder for the strong user is given by

\[
\tilde{m}(y,i(y)) = \arg \max_{0 \leq m \leq M_u-1, 0 \leq i \leq M_z-1} W_1(y|x_{m,i}),
\]

(2)

and the optimal ML decoder for the weak user is given by

\[
\tilde{n}(z) = \arg \max_{0 \leq m \leq M_u-1} \left\{ \frac{1}{M_y} \sum_{x \in \mathcal{C}_m} W_2(z|x) \right\}.
\]

(3)

The likelihood decoder is a stochastic decoder, that chooses the decoded message according to the posterior probability mass function, induced by the channel output (either \( y \) or \( z \)). For the strong user, the ordinary likelihood decoder randomly selects the estimated message \( \tilde{m}(n, i) \) according to the following posterior distribution

\[
P(m,i|y) = \frac{W_1(y|x_{m,i})}{\sum_{m'=0}^{M_u-1} \sum_{i'=0}^{M_z-1} W_1(y|x_{m',i'})}.
\]

The generalized likelihood decoder (GLD) for the strong user is defined by the conditional probability

\[
P(m,i|y) = \frac{\exp\{ ng(\hat{P}_u, x_{m,i}, y) \}}{\sum_{m'=0}^{M_u-1} \sum_{i'=0}^{M_z-1} \exp\{ ng(\hat{P}_u, x_{m',i'}, y) \}}.
\]
where \( \hat{P}_{um, x_{mi}}(y) \) is the empirical distribution of \((u_m, x_{mi}, y)\), and \( g(\cdot) \) is a given continuous, real valued functional of this empirical distribution. In the same manner, the ordinary likelihood decoder for the weak user randomly selects the estimated cloud \( \hat{m} \) according to

\[
P(\hat{m}|z) = \frac{\sum_{m=0}^{M_{-1}} W_2(z|x_{mi})}{\sum_{m'=0}^{M_{-1}} \sum_{m''=0}^{M_{-1}} W_2(z|x_{m'i})},
\]

while the GLD for the weak user is defined by

\[
P(\hat{m}|z) = \frac{\sum_{m=0}^{M_{-1}} \exp\{ng(\hat{P}_{um, x_{mi}}(y))\}}{\sum_{m'=0}^{M_{-1}} \sum_{m''=0}^{M_{-1}} \exp\{ng(\hat{P}_{um, x_{m'i}}(y))\}}.
\]

Exactly as the universal decoders derived in [5], generalized decoders may also depend on the cloud–centers which may be helpful, since all of the codewords in each sub–code are highly correlated via their cloud–center. One of the most important properties of the GLD is the following. The union bound, which is used in the first steps of the derivations for both users, actually provides an exact expression for the probability of error, unlike in the analyses of the ML decoders, where the union bound harms the exponential tightness, at least for relatively high rates. The generalized likelihood decoders cover several important special cases, such as the ordinary likelihood decoder and the mismatched likelihood decoder.

For the strong user, the deterministic ML and the maximum mutual information decoders [9] can be obtained from the GLD by limiting operations.

Let \( Y \in \mathcal{Y}^n \) and \( Z \in \mathcal{Z}^n \) be the random channel outputs resulting from the transmission of \( x_{mi} \). For a given code \( \mathcal{C}(n) \), define the error probabilities as

\[
P_{\text{in}}(\mathcal{C}(n)) = \Pr \{ \hat{m}(Y) \neq (m, i) \mid x_{mi} \text{ sent} \},
\]

and

\[
P_{\text{it}}(\mathcal{C}(n)) = \frac{1}{M_y} \sum_{i=0}^{M_y} \Pr \{ \hat{m}(Z) \neq m \mid x_{mi} \text{ sent} \},
\]

where in both definitions, \( \Pr \{ \cdot \} \) designates the probability measure associated with the randomness of the channel outputs given its input, and the randomness of the stochastic decoders. Moreover, the error probabilities are defined to be zero whenever the blocklength is such that no code can be generated.

Our main objective is to prove the existence of sequences of HCC codes \( \mathcal{C}(n) \) and obtain the tightest possible single–letter expressions that lower bound the following limits

\[
E_\in(\mathcal{C}) = \liminf_{n \to \infty} \left[ -\frac{1}{2} \log \max_{m, i} P_{\text{in}}(\mathcal{C}(n)) \right],
\]

and

\[
E_\text{it}(\mathcal{C}) = \liminf_{n \to \infty} \left[ -\frac{1}{2} \log \max_{m, i} P_{\text{it}}(\mathcal{C}(n)) \right],
\]

both for the ML decoder and the GLD.

In a recent paper [5], exact random coding error exponents have been derived for both users of the ABC. We may expect to improve these error exponents, at least when one of the coding rates is low, by code expurgation. In this paper, we derive expurgated exponents for the ABC under ML decoding in two different methods. In addition, we discuss the GLD, that enables us to achieve the best between the random coding bound and one of the ML–based expurgated bounds.

IV. MAIN RESULTS

A. Maximum Likelihood Decoding

For maximum likelihood decoding, we distinguish between two different methods of expurgation for the HCC ensemble. The first method is based on the following technique of expurgation: we randomly draw a HCC codebook, and then simultaneously expurgate both bad clouds and bad codewords within the remaining clouds. The resulting expurgated bounds are given in Theorem 1. In order to state our first theorem, we start with the following definitions. We define the sets \( S \triangleq \{ Q_{UXX'} : Q_{UX} = Q_{UX} = P_{UX} \} \) and \( D \triangleq \{ Q_{UX', X'X} : Q_{UX} = Q_{UX} = P_{UX} \} \), and the averaged Chernoff distance function by

\[
D_s(Q_{XXX'}) \triangleq -E_Q \log \left[ \sum_{y \in \mathcal{Y}} W_1^{1-s}(y) \cdot W_2^*(y) \right].
\]

For the weak user, define an error exponent function as

\[
E_{w1}(R_y, R_z) \triangleq \max_{0 \leq s \leq 1} \min_{Q_{UX}, E_{\mathcal{X}, \mathcal{X}'}, \mathcal{Y}, \mathcal{Y}', E_{\mathcal{X}'}, \mathcal{Y}, \mathcal{y}} I(Q_{UX} ; U_{X'}^2 ; X'\mathcal{Y}')
\]

\[
+ D_s(Q_{XXX'}) - R_y - R_z.
\]

Next, for the strong user we define the following error exponent functions

\[
E_{w1}(R_y, s) \triangleq \min_{Q_{UX}, X', U_{X'}^2 ; X'\mathcal{Y}', \mathcal{Y}, \mathcal{y}} I(Q_{UX} ; X'\mathcal{Y}')
\]

\[
+ D_s(Q_{XXX'}) - R_y,
\]

and

\[
E_{w1}(R_y, R_z, s) \triangleq \min_{Q_{UX}, X', U_{X'}^2 ; X'\mathcal{Y}', \mathcal{Y}, \mathcal{y}} I(Q_{UX} ; U_{X'}^2)
\]

\[
+ D_s(Q_{XXX'}) - R_y - R_z.
\]

Theorem 1. There exists a sequence \( C \) of HCC codes, with a rate pair \( (R_y, R_z) \) for which both

\[
E_\in(C) \geq E_{w1}(R_y, R_z) \] and \( E_\text{it}(C) \geq E_{w1}(R_y, R_z).\]

The second method is somewhat different, and the idea behind it is the following. At the first step, we expurgate sub-codes, merely according to their cloud–centers. Then, at the second step, we fix the set of cloud–centers of the remaining clouds from the first step, and then expurgate specific code-words, as well as clouds, according to some collective behavior of their codewords. The resulting expurgated bounds are given in Theorem 2, and as can be seen below, the expressions are more complicated than those of Theorem 1, at least for the weak user.

In order to state our second theorem, we need a few definitions. For a given marginal \( Q_{UZ} \), let \( S(Q_{UZ}) \) denote the set of conditional distributions \( Q_{UX}(x|u, z) \) such that \( \sum_z Q_{UZ}(u, z) Q_{UX}(x|u, z) = P_{UX}(u, x) \) for every
(u, x) ∈ U × X, where \( P_{UX} = P_U × P_{X|U} \). We denote \( \bar{t} = 1 - t \). For the weak user, define
\[
\hat{D}_t(R_y, QU_U) \triangleq \min_{Q: P_{UX} \in S(Q_{UX})} \min_{Q_{UX'} \in S(Q_{UX'})} \left\{ t \cdot D(Q_{Z|UX}; W_{Z|X}|Q_{UX}) + \bar{t} \cdot I_{Q}(Z; U'|U) + \right. \\
+ t \cdot [I_{Q}(X; Z|U) - R_g] + \bar{t} \cdot [I_{Q}(X'; Z|U') - R_{g'}] \}. 
\]
We define the set \( \{ Q_{UX}: Q_U = Q_U' = P_U \} \) and an error exponent function
\[
E_{\text{ML1}}(R_y, R_z) \triangleq \max_{0 \leq t \leq 1} \min_{\{ Q_{UX}: \bar{t} \cdot I_{Q}(X'; Z|U') = R_{g'} \}} \left[ I_{Q}(U; U') + \hat{D}_t(R_y, QU_U) - R_g \right]. 
\]
Next, for the strong user we define the following error exponent functions
\[
E_{\text{ML1}}(R_y, s) \triangleq \min_{\{ Q_{UX}: \bar{t} \cdot I_{Q}(X'; Z|U') \leq R_{g'} \}} \left[ I_{Q}(U; U') \right. \\
+ D_s(Q_{XX'}) - R_y, \right. \\
E_{\text{ML2}}(R_y, R_z, s) \triangleq \min_{\{ Q_{UX}: \bar{t} \cdot I_{Q}(X'; Z|U') \leq R_{g'} \}} \left[ I_{Q}(U; U') \right. \\
+ D_s(Q_{XX'}) - R_y - R_z, \\
E_{\text{ML2}}(R_y, R_z) \triangleq \max_{0 \leq s \leq 1} \{ E_{\text{ML2}}(R_y, s), E_{\text{ML2}}(R_y, R_z, s) \}. 
\]

**Theorem 2.** There exists a sequence \( C \) of HCC codes, with a rate pair \( (R_y, R_z) \) for which both
\[
E_{\text{ML}}(C) \geq E_{\text{ML1}}(R_y, R_z) \text{ and } E_{\text{ML}}(C) \geq E_{\text{ML2}}(R_y, R_z). 
\]

**Discussion:** First, all of the expressions in Theorems 1 and 2 generalize the well-known CKM expurgated bound [6]. For example, it can be easily recovered from \( E_{\text{ML1}}(R_y, R_z) \), when degenerating the hierarchical codebook by choosing \( R_y = 0 \), as well as \( P_{X|U}(x|u) = \delta(x - u) \) \((X = U)\).

Concerning the strong user, each bound is given by the minimum between two different expressions. The first expression is related to error events within the cloud of the true codeword. In fact, we have that \( E_{\text{ML1}}(R_y, s) = E_{\text{ML2}}(R_y, s) \), where the difference is given by the second components, \( E_{\text{ML1}}(R_y, R_z, s) \) and \( E_{\text{ML2}}(R_y, R_z, s) \), for which the expurgation method cause a change in the final expressions. Although the objectives in (10) and (14) are the same, the constraints are different, and are not subsets of each other.

Concerning the weak user, the situation is much more complicated, because of the structure of the optimal decoder. The derivation in the proof of Theorem 1 contains the following inequality that may harm the tightness of the bound:
\[
\sum_{z \in C} W_2(z|x)^{1-t} \cdot \sum_{x' \in C'} W_2(z|x')^t 
\leq \sum_{z \in C} \sum_{x' \in C'} W_2^t(z|x) - W_2^{1-t}(z|x'). 
\]
Because of this passage, the bound of Theorem 1 is inferior to the bound of Theorem 2, at relatively high values of \( R_y \). Specifically, the expression given in Theorem 2 reaches a plateau at high \( R_y \), while the expression of Theorem 1 reaches zero. In this regime, there is no loss in the exponent of the weak user if its decoder treats the satellites codewords as noise. In this event, the satellite-rate is immaterial, and the exponent of the weak user only depends on \( R_y \). One should note that the improvement at high rates is obtained by expressions which are more complicated to compute. However, the resulting exponent of Theorem 1 still outperforms the result of Theorem 2, at least for relatively low \( R_y \) values (see Fig. 1).

We next provide some numerical results (Fig. 1), comparing our expurgated bounds for the weak user, as given by Theorems 1 and 2. Let \( W_1 \) and \( W_2 \) be two binary symmetric channels (BSC) with crossover parameters \( p_y = 0.0005 \) and \( p_z = 0.001 \), respectively. Let \( U \) be binary as well and let \( P_U \) be uniformly distributed over \((0, 1)\). Also, let \( P_{X|U} \) be a BSC with crossover parameter \( p_{x|u} = 0.15 \).

![Fig. 1. Expurgated bounds for the weak user (Rz = 0).](image)
\[ \Omega(Q_{UUX'}, R_y, R_z) \triangleq \min_{Q_{Y'U'UX'}} \left( D(Q_{Y|UX})W_{Y|X}(Q_{UX}) + I_Q(U'X'; Y|UX) + \max \{ g(Q_{UXY}), \varphi(R_y, QUY) \} \right). \]

We define the following error exponent functions. For the weak user,
\[ E^{\text{weak}}(R_y, R_z) \triangleq \min_{Q_{UU'XX'} \in \mathcal{P}} \left[ I_Q(U; X') + \Delta \right] \]
and for the strong user
\[ E^{\text{strong}}(R_y, R_z) \triangleq \min_{Q_{UU'XX'} \in \mathcal{S}} \left[ I_Q(U; X') + \Delta \right] \]
where
\[ \Delta = \Omega(Q_{Y|UX}, R_y, R_z) - R_y - R_z. \]

Theorem 3. There exists a sequence \( \mathcal{E} \) of HCC codes, with a rate pair \((R_y, R_z)\) for which both
\[ E_{\text{w}}(\mathcal{E}) \geq E^{\text{weak}}(R_y, R_z) \quad \text{and} \quad E_{\text{s}}(\mathcal{E}) \geq E^{\text{strong}}(R_y, R_z). \]

Discussion
- An expurgated bound for the GLD in the single user regime has been derived by Merhav [8]. It should be noticed that the resulting expressions of Theorem 3, as well as some parts of its proof (in [11, Section 7]) are nontrivial generalizations of the single–user case.
- The expression of (19) has the same structure as (8), except that here the functional \( \Omega(Q, R_y, R_z) \) replaces the expected Chernoff distance, and an additional constraint \((I_Q(U; U') < R_z)\) has been added. We prove in [11, Appendix A] that at least for the choice \( g(Q) = \log W_2(Z|X) \), \( E_{\text{w}}(R_y, R_z) \) is at least as tight as \( E^{\text{weak}}(R_y, R_z) \).
- One of the main advantages of the GLD, is the fact that the derivation of its probability of error may be exponentially tighter than the derivations in the proofs of Theorems 1 or 2. As a consequence, we show in [11, Appendix B] that \( E^{\text{weak}}(R_y, R_z) \) cannot be smaller than the random coding error exponent of the weak user at any pair of rates, by examining the upper bound for the suboptimal universal metric \( g(Q) = I_Q(U; X) \). We conclude that \( E_{\text{w}}(R_y, R_z) \) is at least as tight as the maximum between \( E^{\text{weak}}(R_y, R_z) \) and the random coding error exponent, \( E^{\text{random}}(R_y, R_z) \).
- The same can be proved for the strong user, i.e., that \( E_{\text{s}}(R_y, R_z) \) is at least as tight as the maximum between \( E^{\text{strong}}(R_y, R_z) \) and the random coding error exponent, \( E^{\text{random}}(R_y, R_z) \). We conclude, that there exist a HCC codebook, for which one user works in the “expurgated region”, while the other user works in the “random coding region”. For example, it may be the case when the channel to the strong user is quite clean, while the channel to the weak user is very noisy.

We were not able to determine whether the bound of Theorem 3 is at least as tight as the maximum between the bounds of the first two theorems, although we conjecture that it is indeed the case when choosing one of the decoding metrics \( g(Q) = \beta \log W_2(Z|X) \) or \( g(Q) = \beta I_Q(U; X) \), and letting \( \beta \to \infty. \)

V. PROOF SKETCH OF THEOREM 1
We start by proving that for any \( s, t \in [0, 1] \)
\[ P_{\text{s,mi}}(c(n)) \leq \sum_{Q \in S} N^{\text{mi}}_{s}(Q, C(n)) \cdot e^{-nD(Q_{XX'})} \]
\[ + \sum_{Q \in \mathcal{P}} N^{\text{out}}_{s}(Q, C(n)) \cdot e^{-nD(Q_{XX'})}, \]
where \( N^{\text{mi}}_{s}(Q, C(n)) \), \( N^{\text{out}}_{s}(Q, C(n)) \) and \( N_{s}(Q, C(n)) \) are suitable type–class enumerators. By using the method of types and Markov’s inequality, it is proved that for every \( \varepsilon > 0 \) and all sufficiently large \( n \), there exists a code \( C(n) \) with a rate pair \((R_y, R_z)\), that satisfies, for every \((m, t) \) and every \( Q \),
\[ N^{\text{mi}}_{s}(Q, C(n)) \leq N^{\text{mi}}_{s}(Q), \]
\[ N^{\text{out}}_{s}(Q, C(n)) \leq N^{\text{out}}_{s}(Q), \]
\[ N_{s}(Q, C(n)) \leq N_{s}(Q). \]

Upon substituting these deterministic upper bounds back into (24) and (25), we conclude the bounds given in Theorem 1.

REFERENCES