Converse Bounds for Assorted Codes in the Finite Blocklength Regime

Yanina Y. Shkel
Electrical and Computer Engineering
University of Wisconsin-Madison
Email: yyshkel@wisc.edu

Vincent Y. F. Tan
Institute for Infocomm Research
A*STAR, Singapore and
ECE Dept., National University of Singapore
Email: vtan@nus.edu.sg

Stark C. Draper
Electrical and Computer Engineering
University of Wisconsin-Madison
Email: sdraper@ece.wisc.edu

***Abstract*—We study converse bounds for unequal error protection codebooks with $k > 1$ different classes of codewords. We dub these unequal error protection codes "assorted codes". We extend a finite blocklength converse bound due to Polyanskiy-Poor-Verdú to apply to assorted codes and use this extension to obtain a refined asymptotic expansion for the performance of assorted codes over a discrete memoryless channel. Our main contribution is to demonstrate that there is indeed a loss in the rates of an assorted code compared to equivalent homogeneous (classical) codes. Notably, when the number of codeword classes is polynomial in blocklength n the loss is apparent in the third order $O(\log n)$ term of the asymptotic expansion of the logarithm of the maximum number of codewords. This is in sharp contrast to the previous literature which only considers this problem within regimes where no such loss could be observed.**

I. Introduction

This paper addresses the problem of message-wise unequal error protection (UEP) over discrete memoryless channels (DMCs) in the non-asymptotic regime. We consider a codebook which contains k different classes of codewords. Codewords of the first class are reserved for more important messages and have a more stringent average error probability requirement. Codewords of the second class are reserved for less important messages and have a looser (larger) average error probability requirement, and so on. For convenience, we refer to this sort of messages-wise UEP code as an "assorted code".

This problem has been previously addressed within a framework of error exponents by Csiszár [1], and by Borade et al. [2]. These works show that if codewords in message class i are generated at rate R_i, then each class of codewords has an error exponent $E(R_i)$, where $E(R)$ is the optimal error exponent for a homogeneous (classical) codebook of rate R. Recently, Wang et al. [3] showed a similar result for the dispersion of message-wise UEP codes. They demonstrated that the dispersion of each class of codewords in an assorted code matches the dispersion of each class individually, *provided k is at most polynomial in blocklength n.* Both of these results are perhaps surprising and intriguing. They are also deeply unsatisfying from the UEP coding perspective, since they suggest that we get something for free where assorted codebooks are involved. The tradeoffs which must exist between different classes of codewords are not exposed in the regimes considered in these works.

The study of assorted codes is of great interest from the information-theoretic perspective. The main appeal of assorted codes is that they are a natural candidate for use as building blocks in larger systems. For example, Csiszár [1] and Wang et al. [3] use a message-wise UEP construction to derive the joint source-channel coding exponent and dispersion respectively. We note that both of these works derive their results as intermediate steps in analyzing joint source-channel coding. To attain their desired analysis the refined results derived herein are not necessary. Other examples include the use of *red alert codes*, which form a special subset of assorted codes, one with only *two* classes of codewords. One class contains one very well protected codeword–the red alert codeword. The second class contains exponentially many normal codewords. Red alert codes over channels with feedback have been analyzed by Kudryashov in [4], and by others in [5]–[7]. A number of other related message-wise UEP studies were also done in [8]–[10]. These applications provide an incentive for a more detailed analyses of assorted codes.

The aim of this work is to obtain sharper bounds for the non-asymptotic behavior of assorted codebooks, i.e., when blocklength n is large but finite. To this end we prove a converse bound based on the meta-converse by Polyanskiy-Poor-Verdú (PPV) [11, Theorem 27]. Our extended bound quantifies a required back off from the non-asymptotic fundamental limit of the corresponding homogeneous codes over a large class of DMCs. In particular, for k polynomial we show that the back off is in the third order $O(\log n)$ term. Our bound further applies to other regimes for the number of message classes, k. For example, for k exponential in $\sqrt{n}$ the back off is in the dispersion term and for k exponential in n the back off (not surprisingly) is in the capacity term. In addition to our converse bound we suggest a new metric for the design and analysis of assorted codes: the *expected assorted code rate*.

The rest of this paper is structured as follows. In Section I-A we formally set up the problem. In Section I-B we state some background results. In Section II we derive Theorem 4: a finite blocklength converse bound for assorted codes. In Section III we state and prove our main result: Theorem 5 which gives the normal approximation for rates of assorted codes. Finally, in Section IV we present a brief discussion of our results and we end with some concluding remarks in Section V.

A. Problem Setup

We state our initial definitions with an arbitrary channel in mind. We assume the channel has input alphabet A, output alphabet B, and a channel law $P_{Y|X} = W$. We will derive our bounds in Section II using this "one-shot" perspective. Subsequently, in Section III, we evaluate our bounds for a DMC $W^n(y^n|x^n) = \prod_{i=1}^n W(y_i|x_i)$.

Definition 1 (Assorted Code)**.** *An* assorted code *is a tuple* $(\mathcal{M}, f, g)$ *consisting of*

- k *disjoint classes of messages* $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k\}$ *which form the assorted codebook* $\mathcal{M} = \cup_{i=1}^k \mathcal{M}_i$,
- *an encoding function* $f : \mathcal{M} \to \mathsf{A}$,
- *a decoding function* $g : \mathsf{B} \to \mathcal{M}$.

For an assorted code with k *classes of codewords we define*

$$\epsilon_i = \frac{1}{|\mathcal{M}_i|} \sum_{m \in \mathcal{M}_i} \left[1 - P_{Y|X}\left(g^{-1}(m)|f(m)\right)\right] \tag{1}$$

to be the average probability of error for codewords in $\mathcal{M}_i$.

In Definition 1, the average error probability constraints differentiate the classes of messages from one another. From this perspective the "goodness" of an assorted code is then measured in terms of (a) the average probability of error vector $\boldsymbol{\epsilon} := (\epsilon_1, \ldots, \epsilon_k)$ and (b) the sizes of respective message classes $|\mathcal{M}_1|, \ldots, |\mathcal{M}_k|$. This suggests that for a fixed error vector, there might be a tradeoff in the size of different message classes in an assorted code. This, in turn, motivates us to consider the region of possible assorted codes for a given error vector.

Definition 2 (Assorted Rate Region)**.** *Fix a vector of* k *error probability constraints,* $\mathbf{e} = (e_1, \ldots, e_k)$. *Then we define the assorted rate region for an assorted code as*

$$\begin{aligned}\mathcal{R}(\mathbf{e}) = \{(R_1, \ldots, R_k) : \exists \text{ an assorted code with}\\ \log|\mathcal{M}_i| \geq R_i \text{ and } \epsilon_i < e_i, \quad 1 \leq i \leq k\}.\end{aligned} \tag{2}$$

Further, notice that the $\frac{1}{|\mathcal{M}_i|}$ factor in the statement of equation (1) implies that each messages is selected uniformly within each codeword class. This assumption allows us to talk about the rate of message class i, $R_i = \log|\mathcal{M}_i|$, in a well-defined way. However, Definition 1 does not provide a way to associate a single rate parameter with all of $\mathcal{M}$. To do so we need to have a notion of a priori likelihood of selecting a message from class i. With this observation in mind we introduce the notion of expected rate which is related to the expected transmission volume defined and analyzed in [12].

Definition 3 (Expected Code Rate)**.** *Let* γ_i *be the probability that a message from class* i *is selected for transmission, with* $\sum_{i=1}^k \gamma_i = 1$. *Then the expected assorted code rate for a code* $(\mathcal{M}, f, g)$ *is defined as*

$$R(\gamma_1, \ldots, \gamma_k) = \sum_{i=1}^k \gamma_i \log|\mathcal{M}_i|. \tag{3}$$

Once we apply our bounds to the DMC in Section III we will take $\mathsf{A} = \mathcal{A}^n$ and $\mathsf{B} = \mathcal{B}^n$ to be n-fold cartesian products of the input and output alphabet of the DMC. Then the assorted rate region in Definition 2 and the expected assorted rate in Definition 3 will be functions of blocklength n. Hence, for this application of our bounds equations (2) and (3) need to be suitably normalized by n. Namely,

$$\begin{aligned}\mathcal{R}(\mathbf{e}; n) = \{(R_1, \ldots, R_k) : \exists \text{ an assorted code with}\\ \frac{\log|\mathcal{M}_i|}{n} \geq R_i \text{ and } \epsilon_i < e_i, \quad 1 \leq i \leq k\}\end{aligned} \tag{4}$$

and

$$R(\gamma_1, \ldots, \gamma_k; n) = \sum_{i=1}^k \gamma_i \frac{\log|\mathcal{M}_i|}{n}. \tag{5}$$

B. Background

In this section we provide background theorems to put our result in context. Due to space constraints, we omit definitions of *channel capacity* C and *channel dispersion* V. For a more detailed discussion and formal definitions see any finite blocklength literature [3], [11], [13]. In addition we make the following simplifying assumptions.

- We restrict our attention to channels with a *unique* capacity-achieving input distribution.
- We assume that all average error probabilities of interest are greater than 0 and less than 1.
- We restrict our attention to channels with positive capacity ($C > 0$) and positive dispersion ($V > 0$).

The following theorem for general DMCs captures the best analyses to date of assorted codes in finite block length regime.

Theorem 1. *The assorted rate region for a DMC* W *is given by*

$$\begin{aligned}\mathcal{R}(\mathbf{e}, n) = \Big\{(R_1, \ldots, R_k) :\\ nR_i \leq nC - \sqrt{nV}Q^{-1}(e_i) + O(\log n)\Big\}\end{aligned} \tag{6}$$

where $\mathbf{e} = (e_1, \ldots, e_k)$ *is the average error probability requirement for the* k *message classes, and* k *is at most polynomial in* n *(denoted as* $k = \text{poly}(n)$*).*

For $k = 1$ this is the well-known normal approximation for the rate of a homogeneous code due to Strassen [14]. For $k > 1$, the achievability proof of this theorem is due to Wang et al. [3, Corollary 2] although they use a universal code as in Csiszár [1]. The converse follows from the $k = 1$ case. Theorem 1 suggests that in an assorted code with polynomially many classes, the number of messages of class i can equal to that of a homogeneous code with an average probability of error ϵ_i and no other messages present, *up to the dispersion term*. In other words, by considering the capacity and dispersion terms only, there is *no tradeoff* between the sizes of each message class. We are interested in quantifying these tradeoffs more precisely, and so we look at finite blocklength bounds which analyze the third-order term in more detail. It turns out that for a large class of channels, the third-order term is $\frac{1}{2}\log n + O(1)$. Indeed, we have:

Theorem 2. *Fix a DMC W. Assume that $V^r(P^*, W) := \mathrm{var}(i(X;Y)|Y) > 0$ where P^* is the capacity-achieving input distribution, $i(x;y) = \log \frac{W(y|x)}{P^*W(y)}$ is the information density and $(X,Y) \sim P^* \times W$. Let $M^*(n,\epsilon)$ be the maximum number of codewords achievable for n uses of W and attaining average error probability no larger than ϵ. Then,*

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1). \quad (7)$$

The achievability proof is provided in [15, Corollary 54]. For the class of weakly-input symmetric DMCs[1] [15, Definition 9] the converse is based on the meta-converse bound in [15, Corollary 56]. For a general DMC with $V > 0$, the converse is by Tomamichel and Tan [13]. Note that the hypothesis of the theorem, namely $V^r(P^*, W) > 0$ implies that $V > 0$ since $V \geq V^r(P^*, W)$.

In Section III, we extend Theorem 1 for k polynomial in n and we rely on Theorem 2 to show a loss in the third-order term. We restrict our attention to weakly input-symmetric DMCs. For k growing faster than polynomial in n we bound the loss in rate of assorted codes for a general DMC.

II. Finite Blocklength Bounds

The proof of our extended bound will follow closely the derivations of Theorems 28 and 29 in [15]. We outline this derivation here for completeness.

Definition 4 (Hypothesis Testing)**.** *Consider a random variable B defined on B which can take probability measure P or Q. A randomized test between those two distributions is defined by a random transformation $P_{Z|B} : \mathsf{B} \to \{0,1\}$ where 0 indicates that the test chooses Q. The best performance achievable among those randomized test is given by*

$$\beta_\alpha(P,Q) := \min_{P_{Z|B}: \int_{\mathsf{B}} P_{Z|B}(1|b)\,\mathrm{d}P(b) \geq \alpha} \int_{\mathsf{B}} P_{Z|B}(1|b)\,\mathrm{d}Q(b), \quad (8)$$

where the minimizer $P^*_{Z|B}$ is guaranteed to be attained by the Neyman-Pearson lemma.

The following is a corollary of [15, Theorem 28].

Corollary 3. *Consider two channels $(\mathsf{A}, \mathsf{B}, P_{Y|X})$ and $(\mathsf{A}, \mathsf{B}, Q_{Y|X})$. Fix an assorted code with k classes of messages, $(\mathcal{M}, f, g)$. Let $\{\epsilon_i\}_{i=1}^k$ and $\{\epsilon'_i\}_{i=1}^k$ be its probabilities of error under P and Q, respectively. Let $P_X^i = Q_X^i$ be the probability distribution on A induced by the encoder given that a codeword from $\mathcal{M}_i$ was sent. Then we have*

$$\beta_{1-\epsilon_i}(P^i_{XY}, Q^i_{XY}) \leq 1 - \epsilon'_i, \quad \forall\, 1 \leq i \leq k. \quad (9)$$

The result follows by appealing to [15, Theorem 28] separately for each class of codewords.

[1]A DMC W is *weakly input-symmetric* if there exists an $x_0 \in \mathcal{A}$ and a random transformation $T_x : \mathcal{B} \to \mathcal{B}$ for each $x \in \mathcal{A}$ such that $T_x \circ W_{x_0} = W_x$ and $T_x \circ P_{Y^*}$, where P_{Y^*} is the capacity-achieving output distribution.

We now apply Corollary 3 to show an extension of Theorem 29 in [15] to assorted codes.

Theorem 4. *Let $\mathcal{P}(\mathsf{A})$ be the space of all probability distributions on A, $\mathcal{P}(\mathsf{B})$ be the space of all probability distributions on B, and define $\mathcal{L} := \{(\lambda_1, \ldots, \lambda_k) : \sum_{i=1}^k \lambda_i = 1,\, \lambda_i \geq 0\}$. We can make the following three (equivalent) statements about assorted codes with k classes of codewords.*

1) *For some $\boldsymbol{\lambda} \in \mathcal{L}$ and any $Q_Y \in \mathcal{P}(\mathsf{B})$,*

$$\inf_{P_X^i} |\mathcal{M}_i| \beta_{1-\epsilon_i}(P^i_{XY}, P^i_X \times Q_Y) \leq \lambda_i \quad (10)$$

for all $1 \leq i \leq k$.

2) *We can restate 1) as*

$$\inf_{P_X^1 \times \ldots \times P_X^k} \sup_{Q_Y} \sum_{i=1}^k |\mathcal{M}_i| \beta_{1-\epsilon_i}(P^i_{XY}, P^i_X \times Q_Y) \leq 1 \quad (11)$$

where the inf *is over the k-fold Cartesian product of $\mathcal{P}(\mathsf{A})$ and the* sup *is over $\mathcal{P}(\mathsf{B})$.*

3) *Let $\mathbf{e} = (e_1, \ldots, e_k)$. Then the assorted rate region for a code with k types of codewords must satisfy,*

$$\mathcal{R}(\mathbf{e}) \subset \bigcup_{P_X^1 \times \cdots \times P_X^k} \bigcap_{Q_Y} \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : R_i \leq \log \lambda_i - \log \beta_{1-\epsilon_i}(P^i_{XY}, P^i_X \times Q_Y) \Big\} \quad (12)$$

where the union is over the k-fold Cartesian product of $\mathcal{P}(\mathsf{A})$ and the intersection is over $\mathcal{P}(\mathsf{B})$.

Proof: We proceed by fixing $\bar{P}^i_X = Q^i_X$ and $Q^i_{Y|X} = Q_Y$ for an arbitrary Q_Y (same for all i). Suppose that under this distribution Q_Y, the probability of decoding to a message of type i is λ_i. In this case $\epsilon'_i = 1 - \frac{\lambda_i}{|\mathcal{M}_i|}$. Then we have

$$\beta_{1-\epsilon_i}(\bar{P}^i_{XY}, \bar{P}^i_X \times Q_Y) \leq \frac{\lambda_i}{|\mathcal{M}_i|}, \quad (13)$$

where $\bar{P}^i_{XY} := \bar{P}^i_X \times P_{Y|X}$. Multiplying through by $|\mathcal{M}_i|$ yields equation (10). Now, adding all the bounds together yields

$$\sum_{i=1}^k |\mathcal{M}_i| \beta_{1-\epsilon_i}(\bar{P}^i_{XY}, \bar{P}^i_X \times Q_Y) \leq \sum_{i=1}^k \lambda_i = 1. \quad (14)$$

Since the above holds for all Q_Y we have the bound

$$\sup_{Q_Y \in \mathcal{P}(\mathsf{B})} \sum_{i=1}^k |\mathcal{M}_i| \beta_{1-\epsilon_i}(\bar{P}^i_{XY}, \bar{P}^i_X \times Q_Y) \leq 1. \quad (15)$$

And, since we have the freedom to choose any input distribution for each code word type

$$\inf_{P_X^1 \times \cdots \times P_X^k} \sup_{Q_Y \in \mathcal{P}(\mathsf{B})} \sum_{i=1}^k |\mathcal{M}_i| \beta_{1-\epsilon_i}(P^i_{XY}, P^i_X \times Q_Y) \leq 1. \quad (16)$$

This gives equation (11). To show (12) define $\bar{\mathcal{R}}(\mathbf{e})$ to be the assorted rate region restricted to only the codes whose input distribution for class i is $\bar{P}_X^i$, for some fixed element in $\mathcal{P}(\mathsf{A})^k$. Equation (14) suggests that

$$\bar{\mathcal{R}}(\mathbf{e}) \subset \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : R_i \leq \log \lambda_i - \log \beta_{1-\epsilon_i}(\bar{P}_{XY}^i, \bar{P}_X^i \times Q_Y) \Big\} \tag{17}$$

for fixed $\bar{P}_X^i$ and any Q_Y. Since this must hold for every Q_Y, the region must be in the intersection over $Q_Y \in \mathcal{P}(\mathsf{B})$,

$$\bar{\mathcal{R}}(\mathbf{e}) \subset \bigcap_{Q_Y \in \mathcal{P}(\mathsf{B})} \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : R_i \leq \log \lambda_i - \log \beta_{1-\epsilon_i}(\bar{P}_{XY}^i, \bar{P}_X^i \times Q_Y) \Big\}. \tag{18}$$

And, since $\mathcal{R}(\mathbf{e}) = \bigcup \bar{\mathcal{R}}(\mathbf{e})$ (where the union extends over all k input distributions),

$$\mathcal{R}(\mathbf{e}) \subset \bigcup_{P_X^1 \times \cdots \times P_X^k} \bigcap_{Q_Y} \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : R_i \leq \log \lambda_i - \log \beta_{1-\epsilon_i}(P_{XY}^i, P_X^i \times Q_Y) \Big\}. \tag{19}$$

■

III. Theorems for the DMC

We now state our main result, a refinement of Theorem 1.

Theorem 5. *Let $\mathcal{L}$ be as in Theorem 4. Fix blocklength n, and average error probability vector $\mathbf{e} = \{e_1, \ldots, e_k\}$.*

(i) *The assorted rate region over DMC W must satisfy*

$$\mathcal{R}(\mathbf{e}, n) \subset \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : nR_i \leq nC - \sqrt{nV} Q^{-1}(e_i) + O(\log n) + \log \lambda_i \Big\} \tag{20}$$

where k can be an arbitrary function of n.

(ii) *The assorted rate region over a weakly-input symmetric DMC W (per [15, Definition 9]) must satisfy*

$$\mathcal{R}(\mathbf{e}, n) \subset \bigcup_{\boldsymbol{\lambda} \in \mathcal{L}} \Big\{ (R_1, \ldots, R_k) : nR_i \leq nC - \sqrt{nV} Q^{-1}(e_i) + \frac{1}{2} \log n + \log \lambda_i + O(1) \Big\} \tag{21}$$

where again, k can be an arbitrary function of n.

Proof: Let $\mathcal{P}_n$ be the set of all n-types on the input alphabet $\mathcal{A}$. For part (i), we repeat the argument of [11, Theorem 48] with the caveat that we need to take care when deriving the bound on the multiple message classes. Consider an arbitrary assorted code $\mathcal{M}$ with a *maximum* probability of error vector $\mathbf{e}$. Consider only the codewords in $\mathcal{M}$ which belong only to some type $P_0 \in \mathcal{P}_n$ (across all message classes). Denote the resulting subcode by $(\mathcal{M}_{P_0}, f, g)$ and its k message classes by $\mathcal{M}_{P_0,i}, 1 \leq i \leq k$. Then by (10)

$$\log |\mathcal{M}_{P_0,i}| \leq -\log \beta_{1-e_i}(x^n, P_{Y^n}) + \log \lambda_{i,P_0} \tag{22}$$

where x^n is any element in the type class $T_{P_0}^n$ and P_{Y^n} is the output distribution induced by P_0. Note that we index $\boldsymbol{\lambda}_{P_0} = (\lambda_{1,P_0}, \ldots, \lambda_{k,P_0})$ by P_0 to emphasize that the element of $\mathcal{L}$ need not be the same in the bound for each type class $T_{P_0}^n$. Repeating the argument in [11, Theorem 48] we get

$$\log |\mathcal{M}_{P_0,i}| \leq nC - \sqrt{nV} Q^{-1}(e_i) + \frac{1}{2} \log n + \log \lambda_{i,P_0} + O(1). \tag{23}$$

By the type-counting lemma, $|\mathcal{P}_n| \leq (n+1)^{|\mathcal{A}|-1}$. Hence,

$$\log |\mathcal{M}_i| \leq nC - \sqrt{nV} Q^{-1}(e_i) + \left(|\mathcal{A}| - \frac{1}{2} \right) \log n + \log \lambda_i + O(1). \tag{24}$$

where $\lambda_i = \frac{1}{|\mathcal{P}_n|} \sum_{P_0 \in P_n} \lambda_{P_0,i}$. The result for *average* probability of error follows by a standard asymptotic expurgation argument (cf. [15, Eq. (3.260)]). The proof is then completed by taking the union over all of $\boldsymbol{\lambda} \in \mathcal{L}$.

Next we give proof outline for assertion (ii). Let $\boldsymbol{\lambda} \in \mathcal{L}$ be as in equation (10) . Bound the $-\log \beta_{1-e_i}(\bar{P}_{XY}^i, \bar{P}_X^i \times Q_Y)$ term using the argument in [15, Theorem 55], with Q_Y being the unique capacity-achieving output distribution for all i. ■

Remark 1. *The converse statement in Theorem 5 (i) is tight. The achievability part could be shown by extending the Dependence Testing bound [11, Theorems 17 and 21] to assorted codes.*

Notice that $\boldsymbol{\lambda}$ is a k-dimensional vector whose entries add up to 1 and so the value of most of its elements must depend on k. The contribution of Theorem 5 is thus two-fold.

- Statement (i) shows that for a general DMC and k growing faster than $\mathrm{poly}(n)$, there is a tradeoff in the sizes of different message classes of an assorted code. Two particularly interesting regimes are k growing exponentially in $\sqrt{n}$ and k growing exponentially in n. In these two regimes the tradeoffs are in the dispersion and capacity terms (respectively).
- Statement (ii) shows that for a weakly-input symmetric DMC and k growing as a function of n there is a tradeoff in the sizes of different message classes of an assorted code. A particular regime of interest is $k = \mathrm{poly}(n)$ where the tradeoffs become apparent in the third-order $O(\log n)$ term.

For k constant no meaningful results can be proved since the current normal approximations for homogeneous codes do not quantify the constant term.

Finally, we propose the following conjecture.

Conjecture 1. *Assertion (ii) of Theorem 5 holds for DMCs with unique capacity-achieving input distributions with $V > 0$.*

This conjecture was proved for homogeneous codes (or equivalently, the $k = 1$ case) in a paper by Tomamichel and Tan [13]. We conjecture that a similar extension for assorted codes could be proved using the techniques of [13].

IV. Discussion

A. Geometry of The Assorted Rate Region

In Figure 1, a cartoon drawing of the assorted rate region is provided. To contrast Theorem 1 to Theorem 5 we look at

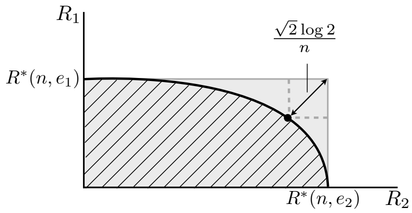

Fig. 1. A cartoon of the k-dimensional assorted rate region for $k = 2$. An outer bound due to Theorem 1 is shaded in light grey. An outer bound due to Theorem 5 is shaded with dark hash lines. We define $R^*(n, e_i) = \frac{1}{n}\log M^*(n, e_i)$. The difference in bounds provided by the two theorems is easiest to see at the "equal back off" point where $\lambda_i = \frac{1}{k}, \forall i$ and the back off per dimension is $\log k$.

the 'equal back off' point on the frontier of the rate region. This is the point where each message class loses the same (in terms of rate). That is, we consider the point $\lambda_i = \frac{1}{k}, \forall i$ and compute the Euclidean distance between the equal back off point predicted by Theorem 1 and in Theorem 5 to be

$$\left(\sum_{i=1}^{k}\left(\frac{\log\lambda_i}{n}\right)^2\right)^{\frac{1}{2}} = \frac{\sqrt{k}}{n}\log k. \tag{25}$$

B. The Expected Rate

We can further relate the assorted rate region to the expected rate. Recall from Section I-A that one measure of "goodness" proposed for assorted codes is the expected rate (see Definition 3). Let us fix the prior probabilities on k message classes $(\gamma_1, \ldots, \gamma_k)$. Then we can look for a point on $\mathcal{R}(\mathbf{e})$ which maximizes the expected rate. Ignoring the $O(1)$ term in (7),

$$\max_{\boldsymbol{\lambda}\in\mathcal{L}} R(\gamma_1,\ldots,\gamma_k;n) = \max_{\boldsymbol{\lambda}\in\mathcal{L}}\sum_{i=1}^{k}\gamma_i\frac{\log|\mathcal{M}_i|}{n} \tag{26}$$

$$\leq \max_{\boldsymbol{\lambda}\in\mathcal{L}}\sum_{i=1}^{k}\gamma_i\left(C-\sqrt{\frac{V}{n}}Q^{-1}(e_i)+\frac{\log n}{2n}+\frac{\log\lambda_i}{n}\right). \tag{27}$$

Now, the first three terms in (27) are constant since they do not involve $\boldsymbol{\lambda}$. Denote them collectively by A. Then, we have

$$\max_{\boldsymbol{\lambda}\in\mathcal{L}} R(\gamma_1,\ldots,\gamma_k;n) \leq A + \frac{1}{n}\max_{\boldsymbol{\lambda}\in\mathcal{L}}\sum_{i=1}^{k}\gamma_i\log\lambda_i \tag{28}$$

$$= A + \frac{1}{n}\sum_{i=1}^{k}\gamma_i\log\gamma_i. \tag{29}$$

Equation (29) follows from the fact that the point which maximizes the expected rate over $\mathcal{L}$ is given by proportional betting with $\lambda_i = \gamma_i$ for all $1 \leq i \leq k$ [16, Theorem 6.1.2]. In other words, our converse bound suggests that there is a one to one correspondence between the frontier of the assorted rate region and the set of prior probabilities on different message classes, $(\gamma_1, \ldots, \gamma_k)$.

V. Concluding Remarks

The assorted coding framework presented in this paper captures the non-homogeneous flavor of real world information such as unequal error protection, and unequal message probability. Specifically, we make the following contributions to the study of assorted codes.

- In Theorem 4 we prove a finite block length converse bound for the assorted rate region.
- In Theorem 5 we derive a normal approximation of the assorted rate region for the DMC and conclude that the assorted rate region has an interesting shape for some regimes of the number of classes of messages.
- Most notably, we demonstrate that in the non-asymptotic regime there are clear tradeoffs between sizes of codeword classes in assorted codes.

From a technical standpoint, future work in this area should include the proof of Conjecture 1, a corresponding achievability bound, and an extension to DMCs with cost, as well as the AWGN channel. From a conceptual viewpoint, further understanding of how assorted codes may impact the design of communication systems may be of great interest.

Acknowledgement: This work was supported in part by the National Science Foundation under Grant CAREER 0844539 and in part with funding support from A*STAR, Singapore.

References

[1] I. Csiszár, "Joint source-channel exponent," *Problems of Control and Information Theory*, vol. 9, no. 5, pp. 315–328, 1982.
[2] S. Borade, B. Nakiboglu, and Z. Lizhong, "Unequal error protection: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5511–5539, 2009.
[3] W. Da, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pp. 180–187.
[4] B. D. Kudryashov, "On message transmission over a discrete channel with noiseless feedback," *Probl. Peredachi Inf*, vol. 15, pp. 3–13, 1979.
[5] S. C. Draper and A. Sahai, "Noisy feedback improves communication reliability," in *Information Theory, 2006 IEEE International Symposium on*, pp. 69–73.
[6] Y. Y. Shkel and S. C. Draper, "Cooperative reliability for streaming multiple access," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1838–1842.
[7] Y. Y. Shkel, S. C. Draper, and B. Nazer, "On the cooperative red alert exponent for the awgn-mac with feedback," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pp. 493–500.
[8] B. Nazer, Y. Shkel, and S. C. Draper, "The awgn red alert problem," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2188–2200, 2013.
[9] A. Sahai and S. C. Draper, "The 'hallucination' bound for the bsc," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 717–721.
[10] B. Nakiboglu, S. Gorantla, L. Zheng, and T. Coleman, "Bit-wise unequal error protection for variable length block codes with feedback," *Information Theory, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2012.
[11] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.
[12] L. R. Varshney, S. K. Mitter, and V. K. Goyal, "An information-theoretic characterization of channels that die," *Information Theory, IEEE Transactions on*, vol. 58, no. 9, pp. 5711–5724, 2012.
[13] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics of discrete memoryless channels," in *2013 International Symposium on Information Theory, to appear*.
[14] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," pp. 689–723, 1962.
[15] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, 2010.
[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., 2006.