Second-Order Coding Rate for m-class Source-Channel Codes

Yanina Y. Shkel
Electrical Engineering
Princeton University
Email: yshkel@princeton.edu

Vincent Y. F. Tan
Electrical and Computer Engineering
National University of Singapore
Email: vtan@nus.edu.sg

Stark C. Draper
Electrical and Computer Engineering
University of Toronto
Email: stark.draper@utoronto.ca

***Abstract*—The second-order coding rate for source-channel codes with m levels of unequal message protection (UMP) is derived. The second-order coding rate takes the form of an optimization problem which reduces to separate source-channel coding rate for $m = 2$. A procedure for solving the given optimization problem is proposed. The proposed procedure exploits the structure of the problem to reduce the optimization search space to one dimension. Numerical results for the BSC are obtained and it is shown, empirically, that the second-order coding rate of source-channel codes with m levels of UMP approaches the optimal joint source-channel coding rate as m grows.**

I. Introduction

When it comes to transmitting an information source over a noisy communication channel, traditional information theory proposes two solutions: *source-channel separation* and *joint source-channel coding* (JSCC). In the separation approach, first proposed by Shannon [1], the information source is compressed with a source encoder, and a separate channel encoder is used for transmission of the compressed source over the channel. For almost lossless transmission of a discrete source (the setting studied in this work) this could be accomplished, for example, by encoding all typical source realizations and discarding the atypical ones. If the information source satisfies asymptotic equipartition property, the separation approach is asymptotically optimal [2, Thm. 7.13.1] in the first-order sense. However, the optimality of separation fails to hold given more refined analysis, see for example [3]. To achieve the best possible tradeoff between rate and reliability of transmission, a code needs to be designed jointly for the given source and channel pair.

Separation and JSCC can be interpreted as unequal error protection schemes. On the one hand, separation uses two levels of a kind of unequal error protection called *unequal message protection* (UMP). A subset of source realizations, the ones discarded during compression, receive no protection from channel noise. All of the remaining source realizations are protected equally by the channel code. On the other hand, optimal JSCCs implicitly sets up many more levels of UMP which gives JSCC schemes an advantage over separation schemes. Often, such superior performance of JSCC is accomplished through the use of maximum a posteriori (MAP) decoding which is intractable in practice.

When viewing separation and JSSC strategies through the lens of UMP a natural question arises which is what sorts of schemes bridge these two approaches. This work addresses exactly this question by characterizing the second-order coding rate of source-channel codes restricted to only m UMP classes, for some integer $m \geq 2$. The rest of this paper is structured as follows. Section I-A overviews second-order coding rates for optimal JSCCs, source-channel separation codes, and m-class source-channel codes. Section I-B surveys relevant work. Section II formally sets up the problem and restates relevant background results. Section III contains the main result, Theorem 1. In Section IV a method for solving the optimization problem presented in Section III is proposed and numerical results for the BSC are presented. Section V concludes the paper with discussion, additional problem motivation, and remarks on future work.

A. Second-Order Rate for Source-Channel Codes

This work studies lossless transmission of k realizations of a discrete memoryless source (DMS) over n uses of a discrete memoryless channel (DMC), under the fidelity constraint that the probability of incorrect source recovery cannot exceed ϵ. It is shown in [4], [5] that the relationship between k, n, and ϵ for optimal JSCC admits the following asymptotic expansion

$$nC - kH(S) = \sqrt{nV + k\mathcal{V}}Q^{-1}(\epsilon) + O(\log n) \tag{1}$$

where C is channel capacity, V is channel dispersion, $H(S)$ is the entropy of the source, and $\mathcal{V}$ is the varentropy of the source. By contrast, the asymptotic expansion for optimal separate source-channel coding is obtained through the following optimization problem,

$$nC - kH(S) = \min_{\mathcal{E}_2(\epsilon)} \left\{ \sqrt{k\mathcal{V}}Q^{-1}(\gamma_1) + \sqrt{nV}Q^{-1}(\epsilon_1) \right\} + O(\log n) \tag{2}$$

where

$$\mathcal{E}_2(\epsilon) = \{(\gamma_1, \epsilon_1) : \epsilon_1 + \gamma_1 - \gamma_1\epsilon_1 \leq \epsilon\}. \tag{3}$$

It is easy to check that, in general, the RHS of (2) is greater than (1). The main result of this work is that optimal m-class source-channel codes admit the following asymptotic expansion,

$$nC - kH(S) = \min_{\mathcal{E}_m(\epsilon)} \max_{1 \leq i < m} \left\{ \sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) + \sqrt{nV}Q^{-1}(\epsilon_i) \right\}$$

$$+ O(\log n) \quad (4)$$

where $\mathcal{E}_m(\epsilon)$ is a generalization of (3); it will be suitably defined for $m \geq 2$, and its operational significance made clear, in Section III. For $m = 2$ equation (4) recovers (2), while it is shown, empirically, in Section IV that (4) approaches (1) for large m.

B. Prior Work

Csiszár [3], [6] was the first to identify UMP as a key component in construction of source-channel codes. His achievability result for optimal lossless JSCC exponent [3] used a UMP construction in which all source realizations in the same type class were encoded with their own UMP class (that is, given the same amount of error protection). Thus, for a DMS, the number of classes used by Csiszár's construction is polynomial in k.

The second-order coding rate (or dispersion) characterization of optimal (lossless and lossy) JSCCs was done concurrently by Kostina-Verdú [4] and Wang-Ingber-Kochman [5]. The separation asymptotic expansion (2), which can be viewed as a direct corollary of [7]–[10], is also stated in [4], [5]. Second-order coding rate results for the problem of UMP were derived in [5] as a step towards the JSCC dispersion. In [11] the problem of UMP was analyzed via single-shot bounds, as well as dispersion and moderate deviation asymptotic analysis. Additional work on UMP was previously done by Borade-Nakiboğlu-Zheng [12], Wang-Ingber-Kochman [5], and Nazer-Shkel-Draper [13]. There have been a number of recent works on source-channel coding with several classes inspired by the approach of Csiszár. Single-shot bounds and numerical results for source-channel codes with m UMP classes were presented in [14]. In a series of concurrent works Bocharova et al. studied the problem of source-channel coding with multiple classes in [15], [16]. The focus of [15], [16] is error exponent analysis rather than second-order coding rate as is presented here.

II. Preliminaries

Let S be an information source with probability density function $P_S(s)$ defined on discrete alphabet $\mathcal{S}$. Let W be a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. This work studies the transmission of DMS S^k over a DMC W^n. We refer the reader to previous works, see for example [4], [5], for definition of channel capacity C, channel dispersion V, source entropy $H(S)$, and source varentropy $\mathcal{V}$. For simplicity of exposition we assume that W and S are such that W has a unique capacity achieving distribution and that $V, \mathcal{V} > 0$. The Q function, the tail probability of a standard normal distribution, is defined to be

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-\frac{u^2}{2}\right\} \mathrm{d}u. \quad (5)$$

A. Channel Coding and UMP

Unequal message protection is a generalization of the traditional channel coding paradigm. In a UMP setup the message set is partitioned into several message classes with each class having its own error protection requirement, and some messages receiving better protection from noise than others. A formal definition follows.

Definition 1 (UMP code). *An $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code for W^n is a tuple $(\{\mathcal{M}_i\}_{i=1}^m, \mathsf{f}, \mathsf{g})$ consisting of*

1) *m disjoint message classes $\{\mathcal{M}_1, \ldots, \mathcal{M}_m\}$ forming the message set $\mathcal{M} := \cup_{i=1}^m \mathcal{M}_i$ and satisfying $|\mathcal{M}_i| = M_i$ for each $i \in \{1, 2, \ldots m\}$*
2) *An encoder $\mathsf{f} : \mathcal{M} \to \mathcal{X}^n$*
3) *A decoder $\mathsf{g} : \mathcal{Y}^n \to \mathcal{M}$*

such that for all $i \in \{1, 2, \ldots m\}$, the average error probabilities for each message class satisfy

$$\frac{1}{M_i} \sum_{w \in \mathcal{M}_i} W^n(\mathsf{B} \setminus \mathsf{g}^{-1}(w) | \mathsf{f}(w)) \leq \epsilon_i. \quad (6)$$

For $m = 1$ Definition 1 becomes an (n, M, ϵ)-channel code studied in [9].

It was essentially shown in [11] that a sequence of optimal $((M_{n,i})_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP codes admits the following asymptotic expansion,

$$\log M_{n,i} = nC - \sqrt{nV} Q^{-1}(\epsilon_i) + O(\log n) - \log \frac{1}{\lambda_{n,i}} \quad (7)$$

for some $\lambda \in \mathcal{L}_m$ where

$$\mathcal{L}_m = \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) : \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0 \quad \forall i\}. \quad (8)$$

Remark 1. *In fact, [11] shows a stronger result for m scaling in n. In the setting of this paper m is fixed. The $\log \frac{1}{\lambda_{n,i}}$ will turn out to be negligible asymptotically. The result stated in [5] would have been sufficient.*

In the case $m = 1$ equation (7) reduces to classical channel coding studied in [7]–[9]

$$\log M_n = nC - \sqrt{nV} Q^{-1}(\epsilon) + O(\log n). \quad (9)$$

B. Source Coding and Source Partitioning

Subsequent results for lossless source-channel coding rely on the following lossless source coding definitions and results.

Definition 2 (Almost-lossless Source Code). *A lossless source code for S^k is a tuple (f, g) consisting of*

1) *an encoding function $\mathsf{f} : \mathcal{S}^k \to \mathcal{M}$,*
2) *and a decoding function $\mathsf{g} : \mathcal{M} \to \mathcal{S}^k$.*

A lossless source code (f, g) is a (k, M, γ)-source code if

$$\mathbb{P}(S^k \neq \mathsf{g}(\mathsf{f}(S^k)))) \leq \gamma. \quad (10)$$

As discussed in [17] an optimal lossless source code for S^k is always known. That is, suppose S^k is to be encoded with M codewords. To minimize the probability of error an optimal code encodes the first most likely M realizations of S^k, breaking ties arbitrarily. The remaining source realizations are discarded and their total probability is the smallest γ such that there exists a (k, M, γ)-source code for S^k.

The asymptotic expansion for source coding was studied in [7], [17]

$$\log M_k = kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma) + O(\log k). \quad (11)$$

The next definition addresses partitioning of the source realizations. It can be thought of as a generalization of Definition 2 in the same way as Definition 1 is a generalization of channel coding.

Definition 3 (Source Partition Code)**.** *An $((M_i)_{i=1}^m, (\gamma_i)_{i=1}^m)$-source partition code for S^k is a tuple $(\{\mathcal{M}_i\}_{i=1}^m, \mathsf{f}, \mathsf{g})$ consisting of*

1) *m disjoint message classes $\{\mathcal{M}_1, \ldots, \mathcal{M}_m\}$ forming the message set $\mathcal{M} := \cup_{i=1}^m \mathcal{M}_i$ and satisfying $|\mathcal{M}_i| = M_i$ for each $i \in \{1, 2, \ldots m\}$*
2) *An encoder $\mathsf{f} : \mathcal{S}^k \to \mathcal{M}$*
3) *A decoder $\mathsf{g} : \mathcal{M} \to \mathcal{S}^k$*

such that for all $i \in \{1, 2, \ldots m\}$,

$$\mathbb{P}\left(\{\mathsf{f}(S^k) \notin \cup_{j=1}^i \mathcal{M}_j\} \cup \{\mathsf{f}(S^k) \neq \mathsf{g}(\mathsf{f}(S^k))\}\right) \leq \gamma_i. \quad (12)$$

It follows from equation (12) that $\gamma_m \leq \gamma_{m-1} \leq \cdots \leq \gamma_1$ for any $((M_i)_{i=1}^m, (\gamma_i)_{i=1}^m)$-source partition code. Furthermore, the observation in [17] regarding optimal source codes also extends to $((M_i)_{i=1}^m, (\gamma_i)_{i=1}^m)$-source partition codes. To minimize all γ_i's simultaneously an optimal $((M_i)_{i=1}^m, (\gamma_i)_{i=1}^m)$-partition code encodes M_1 most likely realizations of S^k with messages from $\mathcal{M}_1$, the next M_2 most likely realizations of S^k with messages from $\mathcal{M}_2$ and so on. It follows from (11) that $((M_{k,i})_{i=1}^m, (\gamma_i)_{i=1}^m)$-source partition code for S^k admits the following asymptotic expansion for $i \in \{1, \ldots, m\}$

$$\log \sum_{j=1}^{i} M_{k,j} = kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) + O(\log k). \quad (13)$$

C. Source-Channel Coding

Definition 4 (Lossless Joint Source-Channel Code)**.** *A source-channel code for source S^k over channel W^n is a tuple (f, g) consisting of*

1) *an encoding function $\mathsf{f} : \mathcal{S}^k \to \mathcal{X}^n$,*
2) *and a decoding function $\mathsf{g} : \mathcal{Y}^n \to \mathcal{S}^k$.*

The source-channel code (f, g) is an (n, k, ϵ)-source-channel code if

$$\mathbb{P}(S^k \neq \mathsf{g}(Y^n))) \leq \epsilon. \quad (14)$$

We refer to all codes allowable by Definition 4 as *joint source-channel codes*. The asymptotic expansion for joint source-channel codes is given by (1).

We say that (n, k, ϵ)-source-channel code (f, g) is a *separation strategy* if it can be expressed as a concatenation of a source code and a channel code. That is,

$$\mathsf{f}(s^k) = \mathsf{f}_{\mathrm{c}}(\mathsf{f}_{\mathrm{s}}(s^k)), \quad \mathsf{g}(y^n) = \mathsf{g}_{\mathrm{s}}(\mathsf{g}_{\mathrm{c}}(y^n)), \quad (15)$$

for some (k, M, γ_1)-source code $(\mathsf{f}_{\mathrm{s}}, \mathsf{g}_{\mathrm{s}})$ and some (n, M, ϵ_1)-channel code $(\mathsf{f}_{\mathrm{c}}, \mathsf{g}_{\mathrm{c}})$. Note that this construction implies,

$$\epsilon = \gamma_1 + (1 - \gamma_1)\epsilon_1. \quad (16)$$

The asymptotic expansion for separate source-channel coding is given by (2).

We say that (n, k, ϵ)-source-channel code (f, g) is an *m-class source-channel code* if it can be expressed as a concatenation of an $(m-1)$-class partition code together with an $(m-1)$-class UMP code. That is,

$$\mathsf{f}(s^k) = \mathsf{f}_{\mathrm{ump}}(\mathsf{f}_{\mathrm{sp}}(s^k)), \quad \mathsf{g}(y^n) = \mathsf{g}_{\mathrm{sp}}(\mathsf{g}_{\mathrm{ump}}(y^n)), \quad (17)$$

for some $\left((M_i)_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code $(\mathsf{f}_{\mathrm{sp}}, \mathsf{g}_{\mathrm{sp}})$ and some $\left((M_i)_{i=1}^{m-1}, (\epsilon_i)_{i=1}^{m-1}\right)$-UMP code $(\mathsf{f}_{\mathrm{ump}}, \mathsf{g}_{\mathrm{ump}})$ both of which utilize the same message set $\mathcal{M} := \cup_{i=1}^{m-1} \mathcal{M}_i$. It follows from this construction that

$$\epsilon = \sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i \quad (18)$$

with $\gamma_0 = 1$ and $\epsilon_m = 1$. We say that (n, k, ϵ)-source-channel code is an optimal JSCC if for any other $(n, \tilde{k}, \epsilon)$-source-channel code $\tilde{k} \leq k$. Likewise, m-class (separation) (n, k, ϵ)-source-channel code is optimal if for any other m-class (separation) $(n, \tilde{k}, \epsilon)$-source-channel code $\tilde{k} \leq k$.

Remark 2. *We assume, with a small loss in generality, that $\gamma_{m-1} < 1$; that is, one class of source symbols receives no error protection at all. Thus, an m-class source-channel code only uses $(m-1)$ level UMP code. The remaining level of UMP is attributed to the source symbols which are discarded by the source encoder. Although for full generality we should allow for the possibility of all source realizations to receive some error protection, this restricted set up aligns well with the separation strategy and leads to nicer overall exposition.*

Remark 3. *As observed in [5, Section V], there is a possibility in separate source-channel coding of a fortuitous assignment of likely source realization to better protected codewords. This may lead to better source-channel coding performance than noted in equation (16). Such theoretical caviat could be circumvented by randomising over the common message set $\mathcal{M}$. The same considerations apply to m-class source-channel codes presenting in this work.*

III. Main Result

Theorem 1. *A sequence of optimal m-class (n, k, ϵ)-source-channel codes must satisfy*

$$nC - kH(S) = \min_{\mathcal{E}_m(\epsilon)} \max_{1 \leq i < m} \left\{\sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) + \sqrt{nV}Q^{-1}(\epsilon_i)\right\} + O(\log n) \quad (19)$$

where

$$\mathcal{E}_m(\epsilon) = \{(\epsilon_1, \gamma_1), \ldots, (\epsilon_{m-1}, \gamma_{m-1}) : \sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i \leq \epsilon, \; 1 = \gamma_0 > \gamma_1 > \ldots > \gamma_{m-1} > \gamma_m = 0, \; 0 = \epsilon_0 < \epsilon_1 < \ldots < \epsilon_{m-1} < \epsilon_m = 1\}. \quad (20)$$

The statement of Theorem 1 can be dissected in the following way. Fix ϵ and a finite channel block length n. Let k be the largest source block length for which an m-class (n, k, ϵ)-source-channel code exists. Then, n and k must be such that

$$nC - kH(S) \simeq \min_{\mathcal{E}_m(\epsilon)} \max_{1 \le i < m} \left\{ \sqrt{k\mathcal{V}} Q^{-1}(\gamma_i) + \sqrt{nV} Q^{-1}(\epsilon_i) \right\} \tag{21}$$

where $\simeq$ denotes that the relationship holds approximately, up to some correction term $\theta_{n,k}$. Theorem 1 says that $\theta_{n,k} = O(\log n)$ and we refer to (21) as *normal approximation* for m-class source-channel codes.

It is also useful to discuss the construction of an optimal m-class source-channel code for fixed source and channel block lengths k and n. Per Section II, every m-class source-channel code can be represented as a concatenation of an $\left((M_{k,i})_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code and an $\left((M_{n,i})_{i=1}^{m-1}, (\epsilon_i)_{i=1}^{m-1}\right)$-UMP code. Let the UMP code be fixed and assume, without loss of generality,

$$\epsilon_1 < \ldots < \epsilon_{m-1}. \tag{22}$$

Consider selection of the best $\left((M_{k,i})_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code given the UMP code. It must be the case that $M_{n,i} = M_{k,i}$ by definition (in the case $M_{n,i} > M_{k,i}$ it is still possible to define a concatenation in a meaningful way, but the resulting source-channel code clearly would not be optimal). Next, we argue that for the given $(M_{k,i})_{i=1}^{m-1}$ no $\left((M_{k,i})_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code leads to smaller source-coding error ϵ than an optimal one (recall discussion below Definition 3).

Indeed, suppose the source-partition code $\left(\{\mathcal{M}_{k,i}\}_{i=1}^{m-1}, \mathsf{f}_{\text{sp}}, \mathsf{g}_{\text{sp}}\right)$ were not optimal. Then, there must be at least two source realizations $s_1^k, s_2^k \in \mathcal{S}^k$ with $P_{S^k}(s_1^k) > P_{S^k}(s_2^k)$ and with

$$\begin{aligned} &\mathsf{f}_{\text{sp}}(s_1^k) \in \mathcal{M}_i, \quad s_1^k = \mathsf{g}_{\text{sp}}(\mathsf{f}_{\text{sp}}(s_1^k)) \\ &\mathsf{f}_{\text{sp}}(s_2^k) \in \mathcal{M}_j, \quad s_2^k = \mathsf{g}_{\text{sp}}(\mathsf{f}_{\text{sp}}(s_2^k)), \quad j < i. \end{aligned}$$

But, then s_2^k has better error protection than s_1^k and swapping the encoding of the two source realizations would only lead to a smaller source-channel coding error.

Finally, $2(m-1)$ parameters determine the source-channel coding error:

- the probability of all recoverable source symbols mapped to message class $\mathcal{M}_i$ is given by $(\gamma_{i-1} - \gamma_i)$, with a convention $\gamma_0 = 1$,
- the error protection of message class $\mathcal{M}_i$ is ϵ_i, with a convention $\epsilon_m = 1$.

From this, it is easy to observe that

$$\sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i = \epsilon. \tag{23}$$

Recalling from Definition 3 that

$$\gamma_1 > \ldots > \gamma_{m-1} \tag{24}$$

holds for any $\left((M_i)_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code together with (23) and (22) gives an interpretation of $\mathcal{E}_m(\epsilon)$. See Appendix A for the remainder of proof outline.

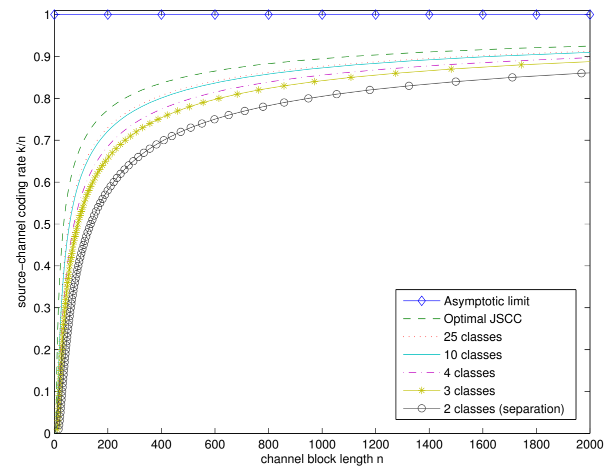

Fig. 1. Normal approximation of rate-block length tradeoff for BMS with bias $\delta = 0.11$ over a BSC with cross-over probability $p = 0.11$. This normal approximation is for $\epsilon = 0.1$.

Remark 4. *Just like $\mathcal{E}_2(\epsilon)$ in (2), the set $\mathcal{E}_m(\epsilon)$ is symmetric in γ_i's and ϵ_i's. That is,*

$$\sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i = \sum_{j=1}^{m} \tilde{\epsilon}_j (\tilde{\gamma}_{j-1} - \tilde{\gamma}_j). \tag{25}$$

with $\gamma_i = \tilde{\epsilon}_j$, $\epsilon_i = \tilde{\gamma}_j$ and $i + j = m$. See Appendix B for proof.

IV. Numerical Results

This section compares normal approximation (21) with normal approximations for JSCC

$$nC - kH(S) \simeq \sqrt{nV + k\mathcal{V}} Q^{-1}(\epsilon), \tag{26}$$

obtained from (1), and normal approximation for separation

$$nC - kH(S) \simeq \min_{\mathcal{E}_2(\epsilon)} \left\{ \sqrt{k\mathcal{V}} Q^{-1}(\gamma_1) + \sqrt{nV} Q^{-1}(\epsilon_1) \right\}, \tag{27}$$

obtained from (2). A procedure for solving the optimization problem (21) is proposed and implemented.[1] Numerical plots for the Binary Symmetric Source (BMS) transmitted over the Binary Symmetric Channel (BSC) are given in Figures 1 and 2.

A. Plots

Figure 1 plots the channel block length vs. source-channel coding rate tradeoff. Going from a 2-class (separation) code to a 3-class code leads to marked improvement in performance. Figure 1 also shows that there is diminishing returns as m grows. The performance of 25-class code is not much better than that of 10-class code. This is to be expected since the number of levels of UMP in the optimal JSCC actually grows as a polynomial with block length. In a sense, m-class codes

[1] When we say 'solving' we mean, 'obtaining a reasonably good solution'. We make not claims of optimality.

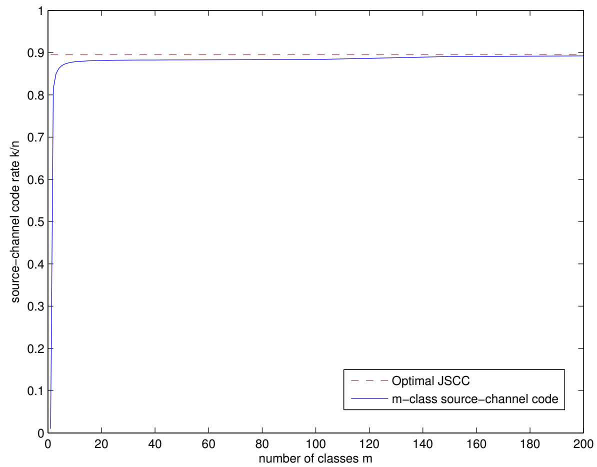

Fig. 2. Normal approximation of optimal JSCC compared to m-class source-channel rate as a function of m for $n = 1000$. The plot is for BMS with bias $\delta = 0.11$ over a BSC with cross-over probability $p = 0.11$. This normal approximation is for $\epsilon = 0.1$.

are trying to catch up with a moving target. Nevertheless, Figure 1 provides a visual demonstration of how m-class source-channel codes may be viewed as a bridge between separation strategy and optimal joint source-channel coding.

Figure 2 plots the number of classes, m, vs. source-channel coding rate for fixed channel block length $n = 1000$. As the number of classes increases, the performance of an m-class source-channel code (as given by normal approximation (21)) approaches that of optimal JSCC (as given by normal approximation (26)). We remark that the statement of Theorem 1 holds for fixed m. As such, it may not be meaningful to consider the setting where m is on the same order as n and so in Figure 2 we only plot the source-channel coding rate for up to $m = 200$. Finally, it would be interesting to numerically compare normal approximation with finite block length bounds. Due to space constraints, we leave this to future work.

B. Proposed Procedure for Solving (21)

Define source-channel coding rate $\rho = \frac{k}{n}$ and divide both sides of (26) by n to obtain,

$$C - \rho H(S) \approx \sqrt{\frac{V + \rho \mathcal{V}}{n}} Q^{-1}(\epsilon). \quad (28)$$

The term on the RHS of (28) is the gap between channel capacity and source entropy needed at finite block lengths for a JSCC with rate ρ and average error ϵ. It goes to zero as $\sqrt{n}$. Call this gap $\frac{G_{opt}(\rho)}{\sqrt{n}}$, that is

$$G_{opt}(\rho) = \sqrt{V + \rho \mathcal{V}} Q^{-1}(\epsilon). \quad (29)$$

Likewise, call the optimal gap for an m-class source-channel code

$$G_m(\rho) = \min_{\mathcal{E}_m(\epsilon)} \max_{1 \le i < m} \left\{ \sqrt{\mathcal{V}} Q^{-1}(\gamma_i) + \sqrt{\rho V} Q^{-1}(\epsilon_i) \right\} \quad (30)$$

where $G_2(\rho)$ gives the gap for separation. Thus, the relationship between ϵ and $G_m(\rho)$ is given by the following optimization problem:

$$\text{minimize} \sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i = \epsilon \quad (31)$$
$$\text{subject to } \sqrt{V} Q^{-1}(\epsilon_i) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_i) \le G_m(\rho),$$
$$i \in \{1, \ldots, m-1\}.$$

We set up a Lagrange multiplier problem

$$g(\vec{\gamma}, \vec{\epsilon}, \vec{\mu}) = \sum_{i=1}^{m} (\gamma_{i-1} - \gamma_i)\epsilon_i$$
$$+ \sum_{i=1}^{m-1} \mu_i \left(\sqrt{V} Q^{-1}(\epsilon_i) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_i) - G_m(\rho) \right) \quad (32)$$

where $\vec{\gamma} = (\gamma_1, \ldots, \gamma_{m-1})$, $\vec{\epsilon} = (\epsilon_1, \ldots, \epsilon_{m-1})$, and $\vec{\mu} = (\mu_1, \ldots, \mu_{m-1})$. Taking the derivatives for $i \in \{1, \ldots, m-1\}$ with respect to $\gamma_i, \epsilon_i, \mu_i$ we obtain,

$$\frac{d}{d\gamma_i} g(\vec{\gamma}, \vec{\epsilon}, \vec{\mu}) = \epsilon_{i+1} - \epsilon_i + \mu_i \sqrt{\rho \mathcal{V}} (Q^{-1})'(\gamma_i) \quad (33)$$
$$\frac{d}{d\epsilon_i} g(\vec{\gamma}, \vec{\epsilon}, \vec{\mu}) = \gamma_{i-1} - \gamma_i + \mu_i \sqrt{V} (Q^{-1})'(\epsilon_i) \quad (34)$$
$$\frac{d}{d\mu_i} g(\vec{\gamma}, \vec{\epsilon}, \vec{\mu}) = \sqrt{V} Q^{-1}(\epsilon_i) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_i) - G_m(\rho). \quad (35)$$

Setting (33), (34), and (35) to zero and solving for μ_i gives a system of equations

$$(\epsilon_{i+1} - \epsilon_i)\sqrt{V} \left(Q^{-1}\right)'(\epsilon_i) = (\gamma_{i-1} - \gamma_i)\sqrt{\rho \mathcal{V}} \left(Q^{-1}\right)'(\gamma_i) \quad (36)$$
$$\sqrt{V} Q^{-1}(\epsilon_i) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_i) = G_m(\rho). \quad (37)$$

Finally, we can compute,

$$\left(Q^{-1}\right)'(x) = -\sqrt{2\pi} e^{\frac{(Q^{-1}(x))^2}{2}} \quad (38)$$

using the chain rule and $Q(x) = 1 - \Phi(x)$ where $\Phi(x)$ is Guassian CDF.

We solve the system of equations given by (36) and (37) using a simple one dimensional binary search which we arrive at via the following line of reasoning. Observe that for $i = 1$ (36) and (37) become,

$$(\epsilon_2 - \epsilon_1)\sqrt{V} \left(Q^{-1}\right)'(\epsilon_1) = (\gamma_0 - \gamma_1)\sqrt{\rho \mathcal{V}} \left(Q^{-1}\right)'(\gamma_1) \quad (39)$$
$$\sqrt{V} Q^{-1}(\epsilon_1) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_1) = G_m(\rho). \quad (40)$$

However, we already know the 'boundary condition' $\gamma_0 = 1$ and thus we have two equations and three unknowns (γ_1, ϵ_1, and ϵ_2). If an oracle gave us the value of ϵ_1 we could compute γ_1 and ϵ_2. Moreover, the $i = 2$ equation would become

$$(\epsilon_3 - \epsilon_2)\sqrt{V} \left(Q^{-1}\right)'(\epsilon_2) = (\gamma_1 - \gamma_2)\sqrt{\rho \mathcal{V}} \left(Q^{-1}\right)'(\gamma_2) \quad (41)$$
$$\sqrt{V} Q^{-1}(\epsilon_2) + \sqrt{\rho \mathcal{V}} Q^{-1}(\gamma_2) = G_m(\rho) \quad (42)$$

which given the knowledge of γ_1 and ϵ_2 from $i = 1$ equations turns into a system of two equations and two unknowns. Propagating through till $i = m - 2$ gives us all the unknown ϵ_i's and γ_i's. Moreover, if the oracle gave us correct value for ϵ_1 the $i = m - 1$ equation

$$(\epsilon_m - \epsilon_{m-1})\sqrt{V}\left(Q^{-1}\right)'(\epsilon_{m-1}) \\ = (\gamma_{m-2} - \gamma_{m-1})\sqrt{\rho\mathcal{V}}\left(Q^{-1}\right)'(\gamma_{m-1}) \tag{43}$$

should hold with the 'boundary condition' $\epsilon_m = 1$ satisfied. Thus, the problem becomes to pick the initial value ϵ_1 in such a way that the boundary condition in the last equation is met. We do so via binary search over possible values of ϵ_1.

One last caveat is that solving (36) and (37) numerically is hard, and the solution is very unstable and sensitive to initial conditions and search precision. To get a more stable solution we approximate (36) with

$$\frac{(\epsilon_{i+1} - \epsilon_i)\sqrt{V}}{\epsilon_i} = \frac{(\gamma_{i-1} - \gamma_i)\sqrt{\rho\mathcal{V}}}{\gamma_i}. \tag{44}$$

The approximation follows from using Chernoff Bound

$$Q(x) \leq e^{\frac{-x^2}{2}} \tag{45}$$

on (38).

V. Discussion

From a practical perspective, separate source-channel coding is an appealing strategy since it provides a way to decompose the problem of source-channel coding into two easier problems: source coding and channel coding. Joint source-channel coding, on the other hand, is an appealing strategy because it leads to the optimal tradeoff between rate and reliability. But, as pointed out in Section I, this performance boost comes with increased decoding complexity making joint source-channel codes intractable. Ideally, we would like to design source-channel codes witch are both tractable and have close-to-optimal rate vs. reliability tradeoffs.

The main result of this work, Theorem 1, characterizes second-order coding rate for m-class source-channel codes. As seen in Figure 1, m-class source-channel codes bridge the source-channel separation and JSCC approaches. In fact, Figure 1 shows that there is a marked improvement in the rate of m-class codes over separation with just a few additional classes. The m-class approach also decomposes the problem of source-channel coding into two problems: source partitioning and UMP coding. As shown in [11], UMP codes can be constructed from linear codes, and their decoding complexity scales in m. In other words, the problem of m-class source-channel coding is closely tied to UMP coding, and there is a path towards practical implementation of both.

For the remainder of this section we address additional open questions about m-class source-channel codes.

A. Lossy Source-Channel Coding

We conjecture that Theorem 1 generalizes to lossy source-channel coding. That is, a sequence of optimal m-class (n, k, ϵ, d)-lossy source-channel codes should satisfy

$$nC - kR(d) \simeq \min_{\mathcal{E}_m(\epsilon)} \max_{1\leq i<m} \left\{\sqrt{k\mathcal{V}(d)}Q^{-1}(\gamma_i) + \sqrt{nV}Q^{-1}(\epsilon_i)\right\}. \tag{46}$$

By generalizing Definition 3 to lossy source coding it should be possible to apply the construction in this work and obtain (46). Unlike for lossless source coding, in the lossy setup we do not expect to find one source-partition code which is optimal for every γ_i. The key step is to show that a source-partition code which is sufficiently good for all γ_i exists.

B. Optimization of normal approximation (21)

In Section IV we propose a procedure for obtaining a good solution to the optimization problem (21). Although this problem is not convex, it does have a lot of symmetry and structure. We leverage this structure to reduced the numerical optimization of (21) to a search in one dimension, with each search step having complexity of $O(m)$. Thus, it appears that the complexity of our current approach scales linearly in m. However, it is not clear how the precision of the one-dimensional search behaves, in general. It too may scale with m leading to a less desirable complexity of this approach. It would be of interest to gain better analytical understanding of the proposed procedure, or even derive alternative approaches to the optimization of the normal approximation (21).

C. Refined Asymptotics

Other questions of interest include obtaining more refined version of Theorem 1. The constant for the $O(\log n)$ (and $O(\log k)$) third order terms are generally known for both channel and source coding. As such, there is hope for characterizing the third order term of m-class source-channel codes as well. Another asymptotic question of interest is the behavior of m-class source-channel codes in the regime were m is allowed to scale in n. In that set up the more refined results of [11] may prove to be helpful. Finally, much of the normal approximation analysis cited here has been extended to Markov sources and Markov channels [8], [18]. The m-class source-channel coding results of Theorem 1 should also hold for a wider class of sources and channels than those studied here.

Appendix

A. Proof Outline of Theorem 1

We begin by proving Theorem 1 for $m = 2$ (separation) case. Fix ϵ_1 and γ_1 such that

$$(1 - \gamma_1)\epsilon_1 + \gamma_1\epsilon_1 \leq \epsilon. \tag{47}$$

Consider an m-class (n, k, ϵ)-source channel code constructed by concatenating a (k, M_k, γ_1)-source code and an (n, M_n, ϵ)-channel code with $M_k = M_n$. From (9) and (11) we obtained

$$kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_1) + O(\log k) = nC - \sqrt{nV}Q^{-1}(\epsilon_1)$$

$$+ O(\log n). \quad (48)$$

We know that $\frac{k}{n} \to \frac{C}{H(S)}$ for a sequence of optimal codes and $O(\log k) = O(\log n)$. Rearranging we obtain

$$nC - kH(S) = \sqrt{k\mathcal{V}}Q^{-1}(\gamma_1) + \sqrt{nV}Q^{-1}(\epsilon_1) + O(\log n).$$

Optimizing over the two parameters γ_1 and ϵ_1 we recover performance of the separation scheme

$$nC - kH(S) = \min_{\mathcal{E}_2(\epsilon)} \left\{ \sqrt{k\mathcal{V}}Q^{-1}(\gamma_1) + \sqrt{nV}Q^{-1}(\epsilon_1) \right\} + O(\log n). \quad (49)$$

Finally, we need to be careful about the $O(\log k)$ and $O(\log n)$ terms since those depend on γ_1 and ϵ_1 and can be potentially unbounded for γ_1 and ϵ_1 approaching zero or one. However, we can argue that for sufficiently large n and k the optimizing γ_1 and ϵ_1 would be each bounded away from zero and one. Hence, the worst case correction terms (c.f. equation (21)) still scale as $O(\log n)$.

Now, consider an m-class source-channel code for some arbitrary fixed number $m \geq 2$. Each m-class (n, k, ϵ)-source code can be expressed as a concatenation of an $\left((M_{k,i})_{i=1}^{m-1}, (\gamma_i)_{i=1}^{m-1}\right)$-source partition code and an $\left((M_{n,i})_{i=1}^{m-1}, (\epsilon_i)_{i=1}^{m-1}\right)$-UMP code for γ_i's and ϵ_i's belonging to $\mathcal{E}_m(\epsilon)$.

Just as above, $M_{k,i} = M_{n,i}$ since each UMP class should have enough codewords to accommodate each source partition. We obtain the first equation from (13) and (7) ,

$$kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_1) + O(\log k) = nC - \sqrt{nV}Q^{-1}(\epsilon_1) - \log \frac{1}{\lambda_1} + O(\log n). \quad (50)$$

Next, we calculate the number of source symbols in class $i > 1$ by applying (13) twice,

$$M_{k,i} = \exp\left\{ kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) \right\} - \exp\left\{ kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_{i-1}) \right\} \quad (51)$$

$$= \exp\left\{ kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) \right\} \left(1 - \exp\left\{ \sqrt{k\mathcal{V}}(Q^{-1}(\gamma_{i-1}) - Q^{-1}(\gamma_i)) \right\}\right). \quad (52)$$

To accommodate all the elements in the ith partition with the ith UMP class it must hold that,

$$kH(S) + \sqrt{k\mathcal{V}}Q^{-1}(\gamma_i) + \theta_k = nC - \sqrt{nV}Q^{-1}(\epsilon_i) - \log \frac{1}{\lambda_i}$$

where

$$\theta_k = \log\left(1 - \exp\left\{ \sqrt{k\mathcal{V}}\left(Q^{-1}(\gamma_{i-1}) - Q^{-1}(\gamma_i)\right) \right\}\right).$$

Taking $\lambda_i = \frac{1}{m}$, noting that $\theta_k = o(\log k)$ (follows from Taylor expansion of $\log(1-x)$ an the fact that $Q^{-1}(\cdot)$ is a decreasing function), and optimizing over $\mathcal{E}_m(\epsilon)$ gives the desired result. Similar to the $m = 2$ case, we can argue that the dependence of third order terms on γ_i's and ϵ_i's is not a problem since for sufficiently large n and k all γ_i's and ϵ_i's will be bounded away from zero and one.

B. Symmetry of the error expression

To argue that constraint set $\mathcal{E}_m(\epsilon)$ is symmetric observe,

$$\sum_{i=1}^{m}(\gamma_{i-1} - \gamma_i)\epsilon_i = \sum_{i=1}^{m}\gamma_{i-1}\epsilon_i - \sum_{i=1}^{m}\gamma_i\epsilon_i \quad (53)$$

$$= \sum_{i=1}^{m-1}\gamma_i\epsilon_{i+1} - \sum_{i=1}^{m-1}\gamma_i\epsilon_i + \epsilon_1 = \sum_{i=1}^{m-1}\gamma_i(\epsilon_{i+1} - \epsilon_i) + \gamma_0(\epsilon_1 - \epsilon_0)$$

$$= \sum_{i=0}^{m-1}\gamma_i(\epsilon_{i+1} - \epsilon_i) = \sum_{j=1}^{m}\tilde{\epsilon}_j(\tilde{\gamma}_{j-1} - \tilde{\gamma}_j). \quad (54)$$

The last line comes from relabeling $\gamma_i = \tilde{\epsilon}_j$ and $\epsilon_i = \tilde{\gamma}_j$ with $i + j = m$.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal,*, vol. 27, pp. 379–423 and 623–656, July/October 1948.

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

[3] I. Csiszár, "Joint source-channel error exponent," *Prob. Contr. & Info. Theory*, vol. 9, no. 5, pp. 315–328, 1980.

[4] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[5] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Allerton Conference*, 2011, arXiv:1109.6310.

[6] I. Csiszar, "On the error exponent of source-channel transmission with a distortion threshold," *Information Theory, IEEE Transactions on*, vol. 28, no. 6, pp. 823–828, Nov 1982.

[7] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

[8] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4947–4966, Nov 2009.

[9] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[10] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[11] Y. Shkel, V. Tan, and S. Draper, "Unequal message protection: Asymptotic and non-asymptotic tradeoffs," *Information Theory, IEEE Transactions on*, vol. 61, no. 10, pp. 5396–5416, Oct 2015.

[12] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5511–5539, Dec 2009.

[13] B. Nazer, Y. Shkel, and S. Draper, "The AWGN red alert problem," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2188–2200, April 2013.

[14] Y. Shkel, V. Tan, and S. Draper, "On mismatched unequal message protection for finite block length joint source-channel coding," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 1692–1696.

[15] I. Bocharova, A. Guillén i Fàbregas, B. Kudryashov, A. Martinez, A. Tauste Campo, and G. Vazquez-Vilar, "Source-channel coding with multiple classes," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 1514–1518.

[16] I. E. Bocharova, A. G. i Fàbregas, B. D. Kudryashov, A. Martinez, A. T. Campo, and G. Vazquez-Vilar, "Multi-class source-channel coding," *CoRR*, vol. abs/1410.8714, 2014. [Online]. Available: http://arxiv.org/abs/1410.8714

[17] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *Information Theory, IEEE Transactions on*, vol. 60, no. 2, pp. 777–795, Feb 2014.

[18] M. Tomamichel and V. Tan, "Second-order coding rates for channels with state," *Information Theory, IEEE Transactions on*, vol. 60, no. 8, pp. 4427–4448, Aug 2014.