On Mismatched Unequal Message Protection for Finite Block Length Joint Source-Channel Coding

Yanina Y. Shkel
Electrical and Computer Engineering
University of Wisconsin-Madison
Email: yyshkel@wisc.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*—We study the problem of lossless joint source-channel coding (JSCC) in the finite block length regime from an unequal message protection (UMP) perspective. We demonstrate that the problem of lossless JSCC can be cast in terms of UMP codes previously studied. We show that an optimal JSCC can be constructed from a matched UMP code. We further derive a finite block length bound that characterizes the performance of a JSCC constructed from a UMP code not perfectly matched to the source. This bound is evaluated for a binary memoryless source transmitted over a binary symmetric channel. Two-class schemes previously studied in literature are compared with the proposed scheme. Empirically the JSCCs based on UMP codes approach the performance of the optimal matched code quite fast in number of classes used.**

I. Introduction

We study the problem of lossless joint source-channel coding (JSCC) where a data source S is transmitted over a noisy channel W. We assume that the source S needs to be reconstructed at the channel output without distortion. However, we are interested in the finite block length regime and so some probability of incorrect transmission ϵ is tolerated. If S is a source that satisfies the asymptotic equipartition property (e.g., S is a discrete memoryless data source (DMS)) this problem has a well known solution in the asymptotic regime: the Source-Channel Separation Theorem [1, Thm. 7.13.1]. The theorem says that S can first be compressed by encoding its typical realizations, and discarding atypical ones, and then a separate channel encoder can encode the compressed source for transmission over W. Asymptotically, this allows for an optimal transmission rate with vanishing probability of error.

However, separation does not hold given more refined analysis, see for example Csiszár [2] or Kostina-Verdú [3]. An optimal finite block length code must be designed jointly for the channel and the source. Thus, the pleasing idea of decomposing the problem via separation is suboptimal at finite block lengths. Other approaches to JSCC design might perform better. For example, one distinguishing feature of JSCCs (as opposed to regular channel codes) is that some source realizations are given better error protection than others: optimal JSCCs depend on unequal message protection (UMP). Further understanding this dependence may provide insight into design of JSCCs.

UMP-based approaches to JSCC have been previously explored in the literature: Csiszár [2] was the first to study UMP codes to derive the JSCC error exponent for a DMS over a discrete memoryless channel (DMC). He analyzed a UMP code with m classes and showed that if codewords in message class i have rate R_i, then each class of codewords has an error exponent $E(R_i)$, where $E(R)$ is the reliability function of the channel. To construct the JSCC, Csiszár partitioned the source realizations into a polynomial number of type classes, and encoded each type class with a message class from the UMP code. This approach assigned different amounts of error protection to each class, as well as allowed the use of different input distributions to generate codewords in each class. Recently Tauste Campo et al. [6] showed that the same JSCC exponent is achievable using only two input distributions. Other notable examples of this approach include [7] where the same two-distribution partitioning scheme is used in the moderate deviations regime and [8] where the dispersion of JSCC is derived using UMP codes.

In this work we revisit Csiszár's approach for finite block lengths JSCC from the viewpoint of UMP codes. We rely on bounds previously studied in [4], [5] and focus on how unequal protection of messages impacts the optimality of a JSCC. Our first contribution is formally to cast lossless JSCC in terms of UMP codes: we show that any lossless JSCC can be constructed from an appropriately selected m-class UMP code, where m depends on the distribution of the source. A consequence of this theorem is that to design a good lossless JSCC it is sufficient to use a UMP code that is perfectly matched to the source. However, the theorem also demonstrates that a UMP code required to maximize performance may be a complicated one: e.g., for a DMS the number of classes required, m, scales as a polynomial in source block length. This raises the natural question of how well a JSCC based on a UMP code with fewer classes can perform. To study this we introduce the notion of a JSCC with $\tilde{m}$ UMP classes, where $\tilde{m} < m$ is the regime of particular interest. Our second contribution is to bound how well such "mismatched" codes can perform. Our final contribution is to conduct a numerical study of our finite block length bound for a binary memoryless source (BMS) and a binary symmetric channel (BSC). Empirically, the performance of mismatched codes approaches the performance of the optimal code quite fast in the number of UMP classes of the mismatched code.

II. Lossless JSCC via UMP Codes

A. Definitions

A general channel from A to B is a stochastic kernel $W(b|a)$ satisfying $\sum_{b\in\mathsf{B}} W(b|a) = 1$ for all $a \in \mathsf{A}$.

Definition 1 (UMP Code)**.** *An $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code for W is a tuple $(\{\mathcal{M}_i\}_{i=1}^m, \mathsf{f}, \mathsf{g})$ consisting of (i) 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 $i \in \{1, 2, \ldots m\}$; (ii) an encoder $\mathsf{f} : \mathcal{M} \to \mathsf{A}$ and (iii) a decoder $\mathsf{g} : \mathsf{B} \to \mathcal{M}$ such that for all $i \in \{1, 2, \ldots m\}$, the average error probabilities for each message class satisfies*

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

If the maximum probability of error for each class also satisfies

$$\max_{w\in\mathcal{M}_i} W(\mathsf{B} \setminus \mathsf{g}^{-1}(w)|\mathsf{f}(w)) \le \epsilon_i \tag{2}$$

we refer to the code as $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code (maximum probability of error).

Definition 2 (Expected Error)**.** *The expected error for an $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code induced by $(\mu_1, \ldots, \mu_m)$ is*

$$\epsilon(\boldsymbol{\mu}) = \sum_{i=1}^m \mu_i \epsilon_i. \tag{3}$$

A general source S is a finite alphabet S together with probability mass function $P_S(\cdot)$ defined on S.

Definition 3 (Lossless Joint Source-Channel Code)**.** *An ϵ-lossless JSCC for a source S over channel W is a tuple (f, g) consisting of (i) an encoding function $\mathsf{f} : \mathsf{S} \to \mathsf{A}$ and (ii) a decoding function $\mathsf{g} : \mathsf{B} \to \mathsf{S}$ such that*

$$\mathbb{P}(S \ne \mathsf{g}(Y))) \le \epsilon. \tag{4}$$

Let us define an equivalence relationship on S by

$$s \approx s' \text{ for } s, s' \in \mathsf{S} \text{ if } P_S(s) = P_S(s'). \tag{5}$$

Note that $\approx$ induces a partition of S into $m \in \{1, \ldots, |\mathsf{S}|\}$ equivalence classes.

Definition 4 (Source Profile)**.** *A profile of source S is a set of tuples $(g_1, \gamma_1), \ldots, (g_m, \gamma_m)$ such that*

$$g_i = |\mathsf{S_i}| \text{ and } \gamma_i = P_S(s), \forall s \in \mathsf{S}_i, \tag{6}$$

*where S_i is the ith set in the partition defined by the equivalence relation $\approx$. We refer to a source with profile consisting of m tuples as an m-*source.

B. Equivalence Theorem

Definition 5 (Matched Code)**.** *Consider an m-source S with profile $(g_1, \gamma_1), \ldots, (g_m, \gamma_m)$. We say that an $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code is matched to S if $M_i = g_i$ for all $i = 1, \ldots, m$.*

Theorem 1 (Equivalence)**.** *Given an m-source S and a channel W, there is an ϵ-lossless JSCC if and only if there is a matched $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code such that the expected error induced by $(g_1\gamma_1, \ldots, g_m\gamma_m)$ is $\epsilon = \sum_{i=1}^m g_i\gamma_i\epsilon_i$.*

Proof: The theorem follows by simply relabeling source realizations, calling them "messages", and then grouping realizations with the same probability into message classes. Formally, consider the partition on S induced by the equivalence relation $\approx$ defined in (5). Let $s_{i,j}$ refer to jth source symbol in S_i. Likewise, let $w_{i,j}$ refer to jth message in message class $\mathcal{M}_i$.

First, we show that an existence of a matched UMP code implies an existence of a JSCC. Suppose there exists an $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code which is matched to the m-source S, and that the expected error of this code induced by $(g_1\gamma_1, \ldots, g_m\gamma_m)$ is ϵ. Then we can construct an ϵ-lossless JSCC $(\mathsf{f}', \mathsf{g}')$ as follows,

$$\mathsf{f}'(s_{i,j}) = \mathsf{f}(w_{i,j}) \text{ and } \mathsf{g}'(Y) = s_{i,j} \text{ if } \mathsf{g}(Y) = w_{i,j}. \tag{7}$$

We can check that

$$\mathbb{P}(S \ne \mathsf{g}'(Y)) = \sum_{i=1}^m \sum_{j=1}^{g_i} \mathbb{P}(S = s_{i,j}) \mathbb{P}(S \ne \mathsf{g}'(Y)|S = s_{i,j}) \tag{8}$$

$$= \sum_{i=1}^m \sum_{j=1}^{g_i} \mathbb{P}(S = s_{i,j}) \left[1 - W({\mathsf{g}'}^{-1}(s_{i,j})|\mathsf{f}'(s_{i,j}))\right] \tag{9}$$

$$= \sum_{i=1}^m \gamma_i g_i \frac{1}{g_i} \sum_{j=1}^{g_i} \left[1 - W({\mathsf{g}'}^{-1}(s_{i,j})|\mathsf{f}'(s_{i,j}))\right] \tag{10}$$

$$= \sum_{i=1}^m \gamma_i g_i \frac{1}{M_i} \sum_{j=1}^{M_i} \left[1 - W(\mathsf{g}^{-1}(w_{i,j})|\mathsf{f}(w_{i,j}))\right] \tag{11}$$

$$= \sum_{i=1}^m \gamma_i g_i \epsilon_i = \epsilon \tag{12}$$

where (8) follows from law of total probability, (10) follows by noting that $\mathbb{P}(S = s_{i,j}) = \gamma_i$ for all j and factoring γ_i out of the inner sum, (11) follows from definition of f' and g' and the fact that the code is matched to the source, finally (12) follows from the definition of expected error.

For the other direction, let $(\mathsf{f}', \mathsf{g}')$ be an ϵ-lossless JSCC. Construct a matched UMP code as follows, $\mathcal{M}_i = \{s_{i,j} : 1 \le j \le g_i\}$, $\mathsf{f}(s_{i,j}) = \mathsf{f}'(s_{i,j})$ and $\mathsf{g}(Y) = \mathsf{g}'(Y)$. Then

$$\epsilon_i = \frac{1}{g_i} \sum_{j=1}^{g_i} \left[1 - W({\mathsf{g}'}^{-1}(s_{i,j})|\mathsf{f}'(s_{i,j}))\right] \tag{13}$$

and computing the expected error,

$$\sum_{i=1}^m g_i\gamma_i\epsilon_i = \sum_{i=1}^m \gamma_i \sum_{j=1}^{g_i} \left[1 - P_{Y|X}({\mathsf{g}'}^{-1}(s_{i,j})|\mathsf{f}'(s_{i,j}))\right] \tag{14}$$

$$= \sum_{i=1}^m \sum_{j=1}^{g_i} P(S = s_{i,j}) \left[1 - P_{Y|X}({\mathsf{g}'}^{-1}(s_{i,j})|\mathsf{f}'(s_{i,j}))\right] = \epsilon, \tag{15}$$

which concludes the proof. ∎

As Theorem 1 demonstrates, to construct an ϵ-lossless JSCC for an m-source S, a UMP code with m classes is needed.

Such a code might be difficult to design if m is large.[1] We are interested in investigating whether simpler codes perform well at finite blocklengths. This motivates the next definition.

Definition 6 (JSCC with $\tilde{m}$ UMP classes)**.** *Let (i) S be an m-source, (ii) $\left(\{\mathcal{M}_i\}_{i=1}^{\tilde{m}}, \mathsf{f}, \mathsf{g}\right)$ be an $\left((M_i)_{i=1}^{\tilde{m}}, (\epsilon_i)_{i=1}^{\tilde{m}}\right)$-UMP code with $\sum_{i=1}^{\tilde{m}} M_i = |\mathsf{S}|$, (iii) and $\mathsf{h} : \mathsf{S} \to \mathcal{M}$ be a one-to-one mapping with an associated equivalence relation $\sim$ on S defined by*

$$s \sim s' \text{ for } s, s' \in \mathsf{S} \text{ iff } \mathsf{h}(s), \mathsf{h}(s') \in \mathcal{M}_i. \tag{16}$$

If it is true that

$$\sum_{i=1}^{\tilde{m}} \sum_{s \in \mathsf{S}_i} P_S(s)\epsilon_i \le \epsilon \tag{17}$$

we refer to $(\mathsf{f} \circ \mathsf{h}, \mathsf{h}^{-1} \circ \mathsf{g})$ as an ϵ-lossless JSCC with $\tilde{m}$ UMP classes.

To relate Definition 3 to Definition 6 observe that (i) if we use $\left((M_i)_{i=1}^{\tilde{m}}, (\epsilon_i)_{i=1}^{\tilde{m}}\right)$-UMP code (maximum probability of error) then the ϵ-lossless JSCC with $\tilde{m}$ UMP classes is also an ϵ-lossless JSCC in the sense of Definition 3.[2] (ii) if S is an $\tilde{m}$-source and the partition induced by $\sim$ is the same as the one induced by $\approx$ (cf. (5)) then the ϵ-lossless JSCC with $\tilde{m}$ UMP classes is also an ϵ-lossless JSCC in the sense of Definition 3. (iii) Otherwise, the mapping h could be chosen badly and associate more likely elements in S_i with worse codewords in $\mathcal{M}_i$. This difficulty could be circumvented by considering a randomized map h.

III. Finite Block Length Bounds

To state our bounds we define the *information density* of joint distribution $P_X \times P_{Y|X}$ at (x, y) as

$$i_{X;Y}(x; y) := \log \frac{P_{Y|X=x}}{P_Y}(y) \tag{18}$$

and the *information* of distribution S at outcome s as

$$i_S(s) := \log \frac{1}{P_S(s)}. \tag{19}$$

Finally, define $\mathcal{L}_m = \{\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}_+^m : \sum_{i=1}^m \lambda_i = 1\}$ to be the m-simplex. We first state the following theorem without proof.

Theorem 2 (Achievability for UMP Codes)**.** *Let $\mathcal{M} = \bigcup_{i=1}^m \mathcal{M}_i$ be a message set with m disjoint message classes and $|\mathcal{M}_i| = M_i$. Fix a $\boldsymbol{\lambda} \in \mathcal{L}_m$, and let P_X be a distribution on A. Assume the CDF of $\mathbb{P}[i(x, Y) \le \alpha]$ does not depend on x for any α when Y is distributed according to P_Y.[3] Then there exists a $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code (maximum probability of error) with*

$$\epsilon_i \le \mathbb{E}\left[\exp\left\{-[i_{X;Y}(X; Y) - \log M_i + \log \lambda_i]^+\right\}\right]. \tag{20}$$

If we drop the assumption on the CDF of $\mathbb{P}[i(x, Y) \le \alpha]$ and let $\boldsymbol{\mu} \in \mathcal{L}_m$, we can say the following: For some error vector $(\epsilon_i)_{i=1}^m$ there exists a $((M_i)_{i=1}^m, (\epsilon_i)_{i=1}^m)$-UMP code with expected error induced by $\boldsymbol{\mu}$ not exceeding

$$\epsilon(\boldsymbol{\mu}) \le \sum_{i=1}^m \mu_i \mathbb{E}\left[\exp\left\{-\left[i_{X;Y}(X,Y) - \log \frac{M_i}{\lambda_i}\right]^+\right\}\right]. \tag{21}$$

In both, (20) and (21), $(X, Y) \sim P_X P_{Y|X}$.

The next bound lets us evaluate the performance of ϵ-lossless JSCCs with $\tilde{m}$ UMP classes. The dummy random variable $\tilde{S}$ can be through of as capturing the partition induced by equivalence relation $\sim$ (cf. (16)), or equivalently map h, in Definition 6.

Theorem 3 (Almost Lossless JSCC)**.** *Let S be an m-source defined on S and $\tilde{S}$ be some other $\tilde{m}$-source also supported on S. Then, there exists an ϵ-JSCC with $\tilde{m}$ UMP classes for source S over channel W satisfying*

$$\epsilon \le \mathbb{E}\left[\exp\left(-|i_{X;Y}(X; Y) - i_{\tilde{S}}(S)|^+\right)\right] \tag{22}$$

$$= \mathbb{E}\left[\exp\left(-\left|i_{X;Y}(X; Y) - i_S(S) - \log \frac{P_S(S)}{P_{\tilde{S}}(S)}\right|^+\right)\right] \tag{23}$$

where the expectation is with respect to $P_S P_X P_{Y|X}$.

Proof: Let $(\mathsf{S}_i)_{i=1}^{\tilde{m}}$ be the partition of S induced by $\approx$ with respect to the random variable $\tilde{S}$ and define,

$$M_i = |\mathsf{S}_i|, \text{ for } i = 1, \dots, \tilde{m}. \tag{24}$$

By Theorem 2 there exists a $\left((M_i)_{i=1}^{\tilde{m}}, (\epsilon_i)_{i=1}^{\tilde{m}}\right)$-UMP code which satisfies (20) for any $\boldsymbol{\lambda} \in \mathcal{L}_m$ if the condition on the CDF of $\mathbb{P}[i(x, Y) \le \alpha]$ is satisfied. Regardless, there exists a $\left((M_i)_{i=1}^{\tilde{m}}, (\epsilon_i)_{i=1}^{\tilde{m}}\right)$-UMP code satisfying (21) for any $\boldsymbol{\mu}, \boldsymbol{\lambda} \in \mathcal{L}_m$. By construction $\sum_{i=1}^{\tilde{m}} M_i = |\mathsf{S}|$ and we pick any one-to-one map h that assigns source realizations from S_i to messages in $\mathcal{M}_i$. Next observe that,

$$\epsilon(\boldsymbol{\mu}) \le \sum_{i=1}^{\tilde{m}} \mu_i \mathbb{E}\left[\exp\left\{-\left[i_{X;Y}(X; Y) - \log \frac{M_i}{\lambda_i}\right]^+\right\}\right] \tag{25}$$

$$= \sum_{i=1}^{\tilde{m}} \sum_{s \in \mathsf{S}_i} P_S(s) \mathbb{E}\left[\exp\left\{-\left[i_{X;Y}(X; Y) - \log \frac{1}{P_{\tilde{S}}(s)}\right]^+\right\}\right] \tag{26}$$

where (25) is the expected error bound for the constituent UMP code and (26) follows by setting

$$\mu_i = \sum_{s \in \mathsf{S}_i} P_S(s), \quad \lambda_i = \sum_{s \in \mathsf{S}_i} P_{\tilde{S}}(s). \tag{27}$$

Thus, combining (3) and (17) with (26) we see that the $\left((M_i)_{i=1}^{\tilde{m}}, (\epsilon_i)_{i=1}^{\tilde{m}}\right)$-UMP code and the map h gives the required ϵ-lossless JSCC with $\tilde{m}$ UMP classes. ■

For $\tilde{S} = S$, Theorem 3 recovers [9] and is dispersion-tight [3]. If $\tilde{m} < m$, we are only allowed $\tilde{m}$ message classes to construct a JSCC for S. In this case, we construct an optimal JSCC for the $\tilde{m}$-source $\tilde{S}$ instead. We use this code for S and the performance degradation due to transmitting the wrong source is captured by the $\log \frac{P_S(S)}{P_{\tilde{S}}(S)}$ term, whose expectation is exactly the redundancy in source coding [1, Theorem 5.4.3]).

[1] See [5, Theorem 8] for motivation for why UMP codes with more classes may be more complex.

[2] This is the case for our bounds in Section IV.

[3] BSC and BEC with P_X being the equiprobable input distribution satisfy this assumption.

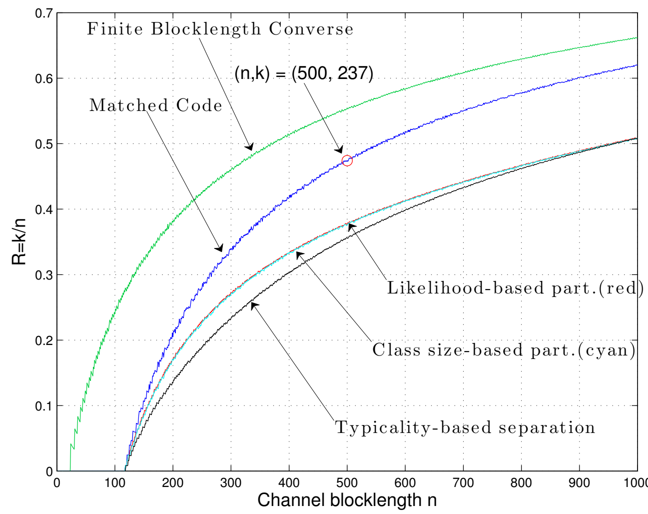

Fig. 1. Rate-block length tradeoff for $\epsilon = 10^{-6}$, $\delta = p = 0.11$ for two-class codes compared to the perfectly matched code, and a finite block length converse [3, Thm. 12]. The rate of the class size-based partition is at or below the rate of the likelihood-based partition. The partitions used to construct the JSCC at the point $(n, k) = (500, 237)$ are plotted in detail in Figure 2.

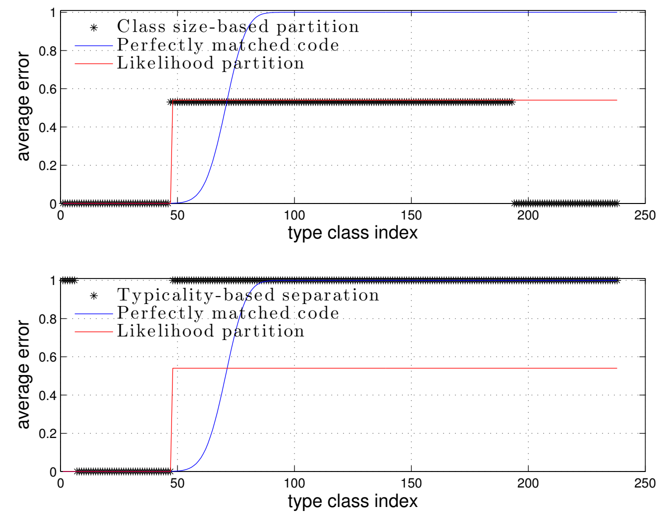

Fig. 2. Average probability of error for each type class of $\{0,1\}^k$ (type classes are indexed by the # of ones they contain). The likelihood-based partition and the perfectly matched code are compared to the rate-based partition (top) and typicality-based separation (bottom) for $n = 500$, $k = 237$ and $\delta = p = 0.11$. The probability of error for the perfectly matched code is 8.64×10^{-7}; for the likelihood-based partition it is 1.17×10^{-4}; for the class size-based partition it is 1.23×10^{-4}; for the typicality-based separation it is 1.25×10^{-4}.

IV. Numerical Evaluation - BMS over BSC

We present numerical evaluations of Theorem 3 for a BMS with bias $\delta < 0.5$ over a BSC with crossover probability $p < 0.5$. If S is distributed as Bernoulli(δ) we refer to our BMS of interest as $S^k \sim \prod_{i=1}^k P_S(s_i)$. We use $\tilde{S}$ to refer to an arbitrary random variable on $\mathsf{S} = \{0,1\}^k$.

Fix some $\tilde{S}$ with profile $(g_1, \gamma_1), \ldots, (g_{\tilde{m}}, \gamma_{\tilde{m}})$ and let $(\mathsf{S}_i)_{i=1}^{\tilde{m}}$ be the partition of S induced by $\approx$ with respect to $\tilde{S}$. To evaluate our bound we select an UMP code with $M_i = g_i$ and $\lambda_i = g_i\gamma_i$ for all i. Evaluating (20) for the BSC we obtain the probability of error for the ith message class,

$$\epsilon_i \leq \sum_{t=0}^{n} \binom{n}{t} p^t(1-p)^{n-t} \min\left[1, \frac{1}{\gamma_i} 2^{-n} p^{-t} (1-p)^{t-n}\right]. \quad (28)$$

Thus, by Theorem 3, there exists an ϵ-lossless JSCC with $\tilde{m}$ UMP classes such that $\epsilon \leq \sum_{i=1}^{\tilde{m}} \sum_{s^k \in \mathsf{S}_i} P_{S^k}(s^k)\epsilon_i$.

Next we address the issue of selecting a good $\tilde{S}$. We will construct $\tilde{S}$ by first defining the partition $(\mathsf{S}_i)_{i=1}^{\tilde{m}}$ and defining all elements in each S_i to be equiprobable. We propose the following likelihood-based partition scheme. For some $(P_j)_{j=0}^{\tilde{m}}$ such that $P_0 = 1, P_{\tilde{m}} = 0$ and $\delta^k \leq P_{\tilde{m}-1}, \leq \cdots \leq P_1 \leq (1-\delta)^k$ the *likelihood-based* $\tilde{m}$ partition of S is

$$\mathsf{S}_i = \left\{ s^k : P_i < P_{S^k}(s^k) \leq P_{i-1} \right\}, \quad 1 \leq i \leq \tilde{m}. \quad (29)$$

We define $\tilde{S}$ to have the following distribution

$$P_{\tilde{S}}(s^k) = \gamma_i, \text{ where } \gamma_i = \frac{P_{S^k}(\mathsf{S}_i)}{|\mathsf{S}_i|} \text{ and } s^k \in \mathsf{S}_i. \quad (30)$$

The intuition behind the likelihood-based partition is that it constructs $\tilde{S}$ which approximates S^k well by combining realizations with similar likelihoods.

A. Two-class Codes

In Figures 1 and 2 we compare our proposed likelihood-based partition to two other two-class partitions previously studied in literature. We define the *class size-based* partition to be

$$\mathsf{S}_1 = \left\{ s^k : |\mathcal{T}(s^k)| \geq 2^{kR_0} \right\}, \quad \mathsf{S}_2 = \mathsf{S} \setminus \mathsf{S}_1 \quad (31)$$

where $\mathcal{T}(s^k)$ denotes the type-class corresponding to s^k. This is the partition studied in [6] and we use it to construct $\tilde{S}$.

Likewise, for some $\epsilon_0 > 0$ the *typicality-based* partition is

$$\mathsf{S}_1 = \left\{ s^k : 2^{-k(H(S)+\epsilon_0)} \leq P_{S^k}(s^k) \leq 2^{-k(H(S)-\epsilon_0)} \right\}, \quad (32)$$
$$\mathsf{S}_2 = \mathsf{S} \setminus \mathsf{S}_1, \text{ where } H(S) = \mathbb{E}_S\left[i_S(S) \right].$$

The typicality-based partition is the classical near-lossless source coding approach of encoding typical realizations used in e.g. [1, Thm. 7.13.1]. The error of the resulting JSCC is thus

$$\epsilon \leq \epsilon^*(|\mathsf{S}_1|, n) + P_{S^k}(\mathsf{S}_2) \quad (33)$$

where $\epsilon^*(|\mathsf{S}_1|, n)$ is the error of a classical channel code of block length n and $|\mathsf{S}_1|$ codewords as given by the Dependence Testing bound [10].

We optimize over P_1, R_0 and ϵ_0 and plot the resulting bounds in Figure 1. The typicality-based separation approach has the worst performance of the three schemes, with the difference being more dramatic when the tolerated probability of error ϵ gets smaller. The class-based and likelihood-based partitions perform similarly, but on closer examination thelikelihood-based partition is always as good or better than the class-based partition. After considering Figure 2 this

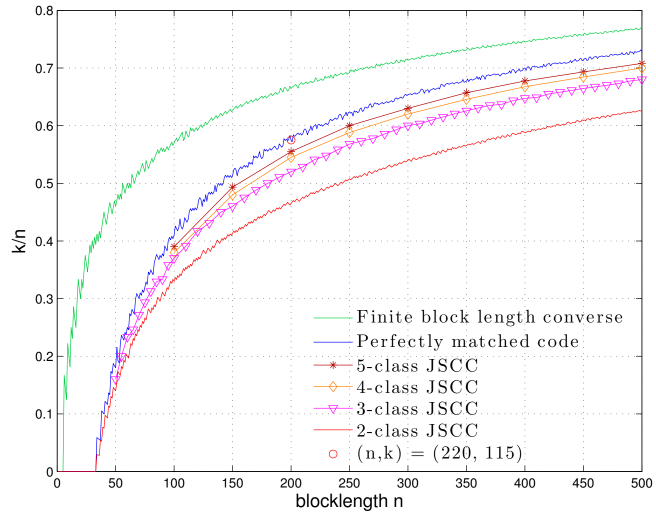

Fig. 3. Rate-block length tradeoff for $\epsilon = 10^{-2}$ and $\delta = p = 0.11$ for 2-class, 3-class, 4-class, and 5-class codes compared to the perfectly matched code, and a finite block length converse [3, Thm. 12]. The partitions used to construct the JSCC at the point $(n, k) = (200, 115)$ are plotted in Figure 4.

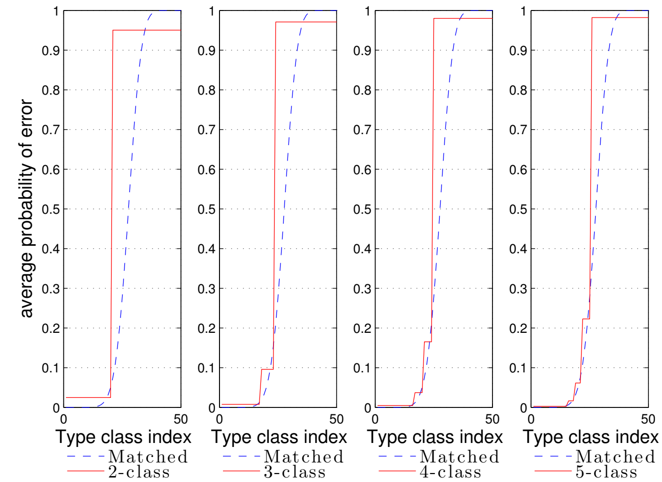

Fig. 4. Average probability of error for each type class of $\{0, 1\}^k$ (type classes are indexed by the # of ones they contain). Partitions for two, three, four, and five (from left to right) UMP class codes compared to the perfectly matched code for $(n, k) = (200, 115)$ and $\delta = p = 0.11$ are plotted. The probability of error for the perfectly matched code is 0.0085; 5-class code it is 0.0129; 4-class code it is 0.0156; 3-class code it is 0.022, and 2-class code it is 0.049.

difference is easy to explain. The class-based partition mimics the likelihood-based partition with the exception that it places twice as many messages in the "good" class.

B. Codes with $\tilde{m}$ UMP classes

We compare our proposed likelihood-based partition for $\tilde{m} \in \{2, 3, 4, 5\}$ in Figures 3 and 4. Figure 3 demonstrates that the rate-block length tradeoff of an ϵ-lossless JSCC with $\tilde{m}$ UMP classes approaches that of the perfectly matched code. The code with four UMP classes already performs well compared with to the perfectly matched code. Note that the number of classes of the perfectly matched code grows polynomially in block length. In Figure 4 we can see how the code with $\tilde{m}$ UMP classes approximates the error profile for the perfectly matched code. The plots do not reflect that most of the probability mass of this source is centered around $k = 22$, which is where the optimal $2-4$ class partitions try best to approximate the curve.

V. Concluding Remarks

We make some comments regarding the normal approximation of Theorem 3. In particular, following the approach of [3] we can apply the Berry-Esseen theorem to show that if S and $\tilde{S}$ are DMSs and W is a DMC with positive dispersion, then there exists an ϵ-joint source channel code for S^k over W^n if

$$nC - kH(S) \geq kD(S\|\tilde{S}) + \sqrt{nV + k\mathcal{V}(S\|\tilde{S})}Q^{-1}(\epsilon) + \theta_n$$

where $\mathcal{V}(S\|\tilde{S}) = \mathsf{Var}_S(i_{\tilde{S}}(S))$ and $\theta_n \leq \frac{1}{2}\log n + O(1)$.

By letting $\tilde{S} = S$ the bound reduces exactly to [3, Theorem 10(e)]. Taking $\tilde{S}$ to be a 1-source we recover the normal approximation for channel coding (we only use one class for JSCC, and we must treat every source realization equally). Taking $\tilde{S}$ as some arbitrary DMS gives us a mismatched coding result. That is, we get a bound on the performance of a JSCC designed for $\tilde{S}$ when the actual source is S. Of course our goal is to obtain a similar statement about ϵ-lossless joint source channel codes with $\tilde{m}$ UMP classes for any $\tilde{m}$. Since in general an $\tilde{m}$-source $\tilde{S}$ will not have a product measure we can not apply the Berry-Esseen theorem to obtain the required result. Some other limit theorem is needed, and this is left for future work.

Acknowledgements: This work was supported in part by the NSF under Grant CAREER 0844539, in part by an NSERC Discovery Research Grant, and in part by NUS startup grant WBS R-263-000-A98-750 (FoE).

References

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

[2] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inform. Theory*, vol. 9, pp. 315–328, 1980.

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

[4] Y. Y. Shkel, V. Y. F. Tan, and S. C. Draper, "Converse bounds for assorted codes in the finite blocklength regime," in *IEEE Intl. Symp. on Inform. Theory*, pp. 1720–1724.

[5] Y. Shkel, V. Y. F. Tan, and S. C. Draper, "Achievability bounds for unequal message protection at finite block lengths," *ISIT 2014*, 2014.

[6] A. T. Campo, G. Vazquez-Vilar, A. G. i Fàbregas, T. Koch, and A. Martinez, "Random-coding bounds that attain the joint source-channel exponent," in *46th Annual Conference on Information Sciences and Systems*, Princeton, NJ, 2012.

[7] V. Y. F. Tan, S. Watanabe, and M. Hayashi, "Moderate deviations for joint source-channel coding of systems with markovian memory," in *ISIT 2014*, 2014.

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

[9] A. Tauste Campo, G. Vazquez-Vilar, A. Guillen i Fabregas, and A. Martinez, "Random-coding joint source-channel bounds," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2011, pp. 899–902.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307–2359, May 2010.